Resilient Sensor Placement for Fault Localization in Water Distribution Networks

Waseem Abbas
Vanderbilt University
Nashville, TN 37212
waseem.abbas@vanderbilt.edu

Lina Sela Perelman
University of Texas Austin
Austin, TX 78712
linasela@utexas.edu

Saurabh Amin
Massachusetts Institute of Technology
Cambrdige, MA 02139
amins@mit.edu

Xenofon Koutsoukos
Vanderbilt University
Nashville, TN 37212
xenofon.koutsoukos@vanderbilt.edu

ABSTRACT

In this paper, we study the sensor placement problem in urban water networks that maximizes the localization of pipe failures given that some sensors give incorrect outputs. False output of a sensor might be the result of degradation in sensor's hardware, software fault, or might be due to a cyber attack on the sensor. Incorrect outputs from such sensors can have any possible values which could lead to an inaccurate localization of a failure event. We formulate the optimal sensor placement problem with erroneous sensors as a set multicover problem, which is NP-hard, and then discuss a polynomial time heuristic to obtain efficient solutions. In this direction, we first examine the physical model of the disturbance propagating in the network as a result of a failure event, and outline the multi-level sensing model that captures several event features. Second, using a combinatorial approach, we solve the problem of sensor placement that maximizes the localization of pipe failures by selecting m sensors out of which at most e give incorrect outputs. We propose various localization performance metrics, and numerically evaluate our approach on a benchmark and a real water distribution network. Finally, using computational experiments, we study relationships between design parameters such as the total number of sensors, the number of sensors with errors, and extracted signal features.

CCS CONCEPTS

•**Networks** →*Cyber-physical networks; Sensor networks; Network performance evaluation;*

KEYWORDS

Resilient sensor placement, fault localization, water distribution networks, minimum set cover

ICCPS 2017, Pittsburgh, PA USA

978-1-4503-4965-9/17/04…$15.00
DOI: http://dx.doi.org/10.1145/3055004.3055020

ACM Reference format:
Waseem Abbas, Lina Sela Perelman, Saurabh Amin, and Xenofon Koutsoukos. 2017. Resilient Sensor Placement for Fault Localization in Water Distribution Networks. In *Proceedings of the 8th ACM/IEEE International Conference on Cyber-Physical Systems, Pittsburgh, PA USA, April 2017 (ICCPS 2017),* 10 pages.
DOI: http://dx.doi.org/10.1145/3055004.3055020

1 INTRODUCTION

Water distribution systems (WDS) are critical infrastructure networks that play a momentous role towards the societal well-being. The complexity of such systems, comprising of water supply sources, treatment plants, and pipe networks, is manifested both at the structural and operational levels. The expansive nature of WDS make them susceptible to disruptions, faults, and failures. For instance, pipe bursts and leakages are inescapable in WDS operations, and if not timely detected, can cause significant loss of water, result in service interruptions, damage surrounding property, and can become a source of introducing contaminants in water distribution system. The ability of the water utility to identify and repair failures in the minimal amount of time is crucial to mitigate the impacts of pipe failures on water supply.

In this direction, real-time monitoring of the hydraulics, such as pressure within pipes, through low-cost and high-rate online sensors enable the timely detection and localization of pipe failures. Some examples of such real-time monitoring of water pipes are WaterWise platform in Singapore [29] and PIPENET in Boston, US [27]. In designing efficient monitoring systems to localize pipe bursts and failures, one of the primary issues is to determine the most effective locations to deploy sensors within the network. In practice, a limited number of sensors are available, and due to the enormous scale of the networks, measurements can only be performed at a limited number of locations. To exacerbate the situation, sensors are error prone, and can give incorrect outputs due to degradations in sensor hardware or software, or due to cyber attacks, which could lead to a false decision regarding the detection and localization of pipe failures.

In this paper, our goal is to *design a sensor placement scheme that maximizes the localization of pipe bursts under a limited budget and error prone sensors.* In our previous works [1, 23], we presented efficient sensor placement designs to maximize localization in WDS

without considering any sensor errors. Here, we examine the consequences of incorrect sensor observations on the localization of pipe failures in detail, and present a sensor placement algorithm that also considers possible errors in sensors' outputs.

First, we discuss the transient model that characterizes the system response to pipe failures that is observable by the sensors. We also present a multi-level sensing model, in which multiple features are extracted from the pressure signal and represented as a boolean string with σ bits. In the case of erroneous sensors, these bits can be flipped from their actual values. Second, we formulate the problem of selecting optimal locations for a given number of sensors, out of which a certain number of sensors can be erroneous, as a set multicover (SMC) problem. SMC is a well-known combinatorial optimization problem, and is known to be NP-hard [9, 28]. We suggest a greedy heuristic to solve the SMC problem and to find the sensor locations. We state the conditions under which a pipe failure can always be localized correctly even in the presence of erroneous sensors. Third, we compare different sensor configurations and study the dependencies between the localization performances and design parameters. Application to case studies using a benchmark and a real water distribution network demonstrate the value of our approach.

Sensor placement problem for fault detection and localization appears in the context of many different networked systems, such as, power and transportation.Various formulations and solution approaches have been proposed to solve the sensor placement problem, including integer and mixed integer programming based methods [3], combinatorial optimization techniques [16, 23], evolutionary algorithms [8], and data-driven approaches [15]. We give a brief overview of the work most relevant to ours in Section 6. We note here that our approach is general and can also be applied to other networks. The rest of the paper is organized as follows: In Section 2, we present the fault, sensing, and error models, and formulate the localization problem. In Section 3, we propose our solution to the sensor placement problem, and present a number of metrics to measure the localization performance in Section 4. We evaluate our approach on two water distribution networks in Section 5, and give an overview of the related work in Section 6. Finally, we conclude the paper in Section 7.

2 SYSTEM MODEL AND PROBLEM

A water distribution network has broadly three main components, water sources, treatment plants, and distribution network consisting of pipes, valves, and pumps etc. The pipe network is often represented by a graph model, in which links represent the pipes and nodes represent pipe junctions, waypoints on curved pipes, or sensor locations (e.g., see [29]). In this section, first, we present the transient model for pipe failures, sensing model, and sensor error model. Then, we state the sensor placement problem that maximizes the localization of failures. A list of symbols used throughout the paper is given in Table 1.

2.1 Transient Model for Pipe Failures in Water Distribution Systems

Physical failures of the infrastructure, such as pipe bursts, cause a disturbance in the flow, which moves through the system as a pressure wave with very high velocity ($500-1400[\frac{m}{s}]$) [29], known as *water hammer*. The transient system state can be described by mass and momentum partial differential equations formulated as [30]:

$$\frac{\partial h}{\partial t} + \frac{a^2}{gA}\frac{\partial q}{\partial x} = 0 \tag{1}$$

$$\frac{1}{gA}\frac{\partial q}{\partial t} + \frac{\partial h}{\partial x} + \frac{cq|q|}{2gDA^2} = 0 \tag{2}$$

where h is the hydraulic head $[m]$, q is the volumetric flow rate $[\frac{m^3}{sec}]$, g is the gravitational acceleration $[\frac{m}{sec^2}]$, x is distance along the pipe $[m]$, t is the time $[sec]$, a is the wave speed in the conduit $[\frac{m}{sec}]$, c is a friction factor, D is the pipe diameter $[m]$, and A is the pipe cross sectional area $[m^2]$.

The effect of a pipe burst at location i can be translated into boundary conditions using the orifice head-flow relation [30]. Before the burst occurs, the cross-section area of the orifice is equal to zero and it increases during a burst, as a result we can expect a sudden change in the hydraulic pressure head. Consequently, the disturbance caused by a pipe burst can be detected by sensing the hydraulic pressure.

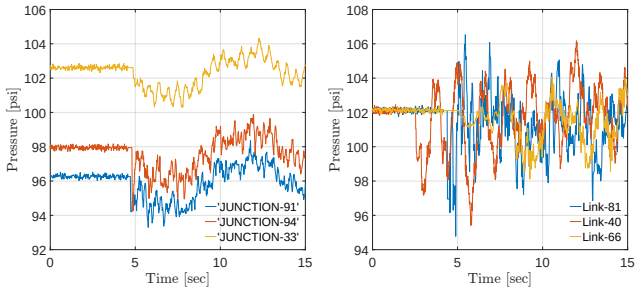
We use a benchmark network [21] to simulate the pipe failures. The system consists of 126 nodes, 168 pipes, one pump, one reservoir, and two storage tanks and its layout is shown in Figure 4(a). The network has a total pipe length of $37.5 \times 10^3[m]$, and supplies a daily demand of $5.15 \times 103[m^3/\text{day}]$. Full details of the network can be found in [21]. Figure 1(a) shows the pressure signals at three different locations in a network resulting from a simulated pipe burst. As the pressure wave arrives at a each location a rapid ($< sec$) drop in the pressure occurs followed by a gradual return to previous operating state. Furthermore, we can observe different arrival times, magnitude, and shape characterizing the pressure wave at different locations in the network. Figure 1(b) shows the pressure signals at a single location in a network in response to simulated bursts at three different pipes in the network. We can again observe, that each pipe burst produces unique pressure signal.

Table 1: A list of Symbols.

Symbol	Description
m	number of sensors
n	number of events
σ	number of bits in a sensor output
ℓ_j	failure event at the pipe j
e	max. number of sensors that can give errors
S_i	output of sensor i
S	array of sensor outputs $S = [S_1 \ \cdots \ S_m]$
$\tilde{S}$	array of sensor outputs with errors
$H(x, y)$	Hamming distance between strings x and y.

2.2 Multi-level Sensing Model

The pressure signal generated as a result of a pipe burst has various characteristic features including the time of arrival, rate of pressure drop, and rate of pressure recovery. These features can be extracted



(a) (b)

Figure 1: (a) Pressure signals at three selected locations in the network from a simulated pipe burst. (b) Pressure signals at selected location in the network from three simulated pipe bursts.

from the signal and can be analyzed to detect and locate pipe bursts. For instance, if the rate of pressure drop is greater than a certain threshold value, then the event is detected. By considering multiple features and thresholds for the signal, the location of the event can also be identified. For instance, the rate of pressure drop in the received signal can classified into slow, gradual and rapid depending on the range within which the actual pressure drop lies.

For the purpose of sensor placement, we consider a discrete representation of the raw pressure signal. The pressure signal received at the sensor within a certain time window is reduced to a σ-bit boolean string representing a single sensor output. We first extract η significant features from the pressure signal and transform these features into a boolean string as follows:

Let $\mathcal{Y} = \{1, \cdots, \eta\}$ be the set of extracted features in the transient signal, and $\mathcal{L} = \{\ell_1, \cdots, \ell_n\}$ be the set of events to be localized. We represent the value of feature $y \in \mathcal{Y}$ in the signal generated as a result of event ℓ_j by $f_y(\ell_j)$, i.e.,

$$f_y : \mathcal{L} \longrightarrow \mathbb{R} \tag{3}$$

The range of f_y can be divided into intervals, and a unique σ_y-bit boolean string can be associated with each interval. Thus, whenever ℓ_j occurs, a unique σ_y-bit string is generated, denoted by $s_y(\ell_j)$, that represents the discretized value of the feature y. More precisely,

$$s_y(\ell_j) = \begin{cases} b_1 & \text{if } \beta_0 \leq f_y(\ell_j) \leq \beta_1 \\ b_2 & \text{if } \beta_1 < f_y(\ell_j) \leq \beta_2 \\ \vdots & \vdots \\ b_\upsilon & \text{if } \beta_{\upsilon-1} < f_y(\ell_j) \leq \beta_\upsilon \end{cases} \tag{4}$$

Here, $\forall i \in \{1, \cdots, \upsilon\}$, b_i is a boolean string with σ_y bits, and β_i's $\forall i \in \{0, 1 \cdots, \upsilon\}$ are the threshold values of the intervals of f_y. Note that σ_υ is at least $\lceil \log_2 \upsilon \rceil$.

The output of sensor i as a result of event ℓ_j, denoted by $S_i(\ell_j)$ is simply the concatenation of $s_y(\ell_j)$'s for all $y \in \{1, 2, \cdots, \eta\}$, i.e.,

$$S_i(\ell_j) = \begin{bmatrix} s_1(\ell_j) & s_2(\ell_j) & \cdots & s_\eta(\ell_j) \end{bmatrix} \tag{5}$$

The array consisting of outputs from m individual sensors in response to ℓ_j is referred to as the *signature of event* ℓ_j, and is denoted by

$$S(\ell_j) = \begin{bmatrix} S_1(\ell_j) & S_2(\ell_j) & \cdots & S_m(\ell_j) \end{bmatrix} \tag{6}$$

Example. As an example, consider the network in Figure 2 with ten links of the same length (1000[m]) and eight possible sensors. The failure events are the pipe bursts in the middle of pipes. The sensor extracts the time of arrival from the signal generated as a result of an event. Since the pressure transient decays with time as it travels distance, we assume that a sensor either detects the event within 1.5[sec] of its occurrence or does not detect at all (assuming that the velocity of propagation is 700[$\frac{m}{s}$]). Moreover, in the case of detection, the time of arrival belongs to one of two intervals, [0 .75][sec] and (.75 1.5][sec]. A sensor output consists of two bits and has three possible outcomes, [0 0] in the case of no detection, [1 0] if the time of arrival is in [0 .75][sec], and [1 0] if the time of arrival lies within the interval (.75 1.5][sec]. The outputs of sensors for events $\{\ell_1, \ell_2, \ell_3, \ell_4\}$ are shown below.

	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

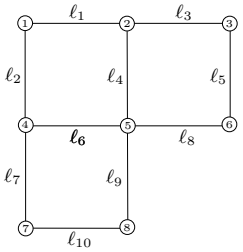


Figure 2: An example network.

2.3 Sensor Errors

We assume that the sensors are not perfect and can give errors, which might lead to an incorrect decision regarding the localization of a pipe burst. By a sensor error, we mean that a single or multiple bits in the sensor output are flipped. Thus, as a result of an error, the σ-bit output of a sensor can have any of the 2^σ possible values. Note that in reality such effects can be introduced due to sensor degradation, especially since the sensor assembly is attached to a physical infrastructure component that is subject to corrosion, loose connections, etc. At the same time, sensor errors can be introduced due to cyber attacks, in which an attacker corrupts the actual output of a sensor after compromising the sensor. A sensor with an error in its output, either due to a faulty hardware or software, or as a result of a cyber attack, is referred to as an *erroneous sensor*.

In our model, we consider an upper bound on the number of sensors that can be erroneous, that is, given m sensors, at most e of them can be erroneous, and the outputs of erroneous sensors can be altered arbitrarily. The output of erroneous sensor i is denoted $\tilde{S}_i$, and the array of all sensors outputs' containing some erroneous sensors is denoted by $\tilde{S}$. The proposed error model can be used to

model a class of attacks in which an attacker takes control of at most e sensors and changes their output in any possible way.

2.4 Problem Description

A primary objective of placing sensors within a water network is to uniquely detect and localize the source of pressure transient associated with a pipe burst. The ability to localize pipe bursts accurately depends on the uniqueness of signatures corresponding to the link failure events. In the best scenario, sensors are placed such that the signatures corresponding to all possible events are unique, and the output of sensors, as a result of some event, always matches the right signature. Thus, in the case of n events, there are n unique signatures, and the array of sensors' outputs due to some event is always the signature of the event. However, in practice, it is not always possible owing to a number of reasons. For instance, a limited number of sensors are available, thus, pressure transients can only be measured at a limited number of locations within a network. At the same time, sensors might be erroneous, which may lead to an incorrect decision regarding the location of event. For instance, in Figure 2, consider that sensors are placed at nodes 1,2,3,6, and 7. In the case of event ℓ_3, if sensor at node 3 gives an incorrect output $\tilde{S}_3(\ell_3) = [0\ 1]$ instead of $S_3(\ell_3) = [1\ 0]$, then the pipe burst is incorrectly localized at ℓ_4. Thus, our first problem is to maximize the localization of events with a limited number of sensors, some of which might give incorrect outputs. More precisely, we aim to study,

How to place m sensors, each with a σ-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?

In our setup, the design parameters that affect the localization performance of the sensor placement are the number of sensors to be deployed m, the maximum number of erroneous sensor e, and the number of output bits σ in a sensor. An interesting consideration here is to study their dependencies on the localization performance of the sensor placement. For instance, to achieve a desired localization performance with σ-bit sensors, how m changes with e? Similarly, fixing the number of erroneous sensors and the number of output bits in a sensor, how does the localization of events improve by increasing the number of deployed sensors? More generally, we aim to investigate the following:

What is the trade-off between m, σ, e, 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?

We study above problems in the next sections. First, using a combinatorial setting, we reduce the sensor placement problem with erroneous sensors to a well known combinatorial optimization problem known as the set multicover problem. Then, we present heuristics to solve the problem and evaluate our approach.

3 LOCALIZATION OF FAULTS IN THE PRESENCE OF SENSOR ERRORS

In this section, we present a sensor placement algorithm to localize pipe bursts in water distribution networks. The algorithm is resilient to a fixed number of sensor errors. First, we overview the sensor placement in the case of no erroneous sensors.

3.1 Localization with No Sensor Errors

To localize event ℓ_i through $S = \begin{bmatrix} S_1 & S_2 & \cdots & S_m \end{bmatrix}$, it is necessary and sufficient that for every $\ell_{j\neq i}$, there always exists a sensor output S_k that is different for ℓ_i and ℓ_j. If for a pair of events ℓ_i and ℓ_j, there exists a sensor that gives different outputs, that is $S_k(\ell_i) \neq S_k(\ell_j)$, then we say that the *pair-wise event, denoted by $\ell_{i,j}$ is detectable.* Consequently, event ℓ_i can be uniquely detected and localized if pair-wise events $\ell_{i,j}\ \forall j \neq i$ are detectable. As an example, consider ℓ_1 and ℓ_2 in the above example. Since $S_2(\ell_1) \neq S_2(\ell_2)$, we can always distinguish between ℓ_1 and ℓ_2 by the output of sensor 2. In other words, the pair-wise event $\ell_{1,2}$ is detectable by the sensor 2.

The localization problem can thus be formulated as a detection problem with the event space consisting of all pair-wise events $\ell_{i,j}$. Moreover, we define *identification score as the fraction of pair-wise events that are detectable by the sensor outputs.* Note that in the case of no errors, the array of sensors' outputs is always a signature corresponding to the event occurred. The sensor selection problem to maximize the identification score, and hence to achieve the maximum localization, can be solved using a well known *maximum coverage* problem (e.g., [23]).

Definition 3.1. (Maximum Coverage Problem (MCP)) Given a set of elements $\mathcal{U}$, a collection C of subsets of $\mathcal{U}$, and a positive integer m. The maximum coverage problem is to select the sub-collection $C_s \subseteq C$ containing m subsets, such that the union of subsets in C is maximized.

For the localization purpose, $\mathcal{U}$ is the set of all pair-wise events $\ell_{i,j}, \forall i, j \in \{1, 2, \cdots, n\}$ and $i \neq j$; $C = \{C_1, \cdots, C_r\}$ is the collection of r subsets of $\mathcal{U}$, each of which corresponds to a particular sensor. C_i contains all pair-wise events that are detectable by the sensor i. The sensor selection problem to maximize the identification score is to select m subsets (sensors) in C whose union is of maximum cardinaltiy, and thus, maximizes the number of detectable pair-wise events. We studied this problem for σ-bit sensors in [1, 23], wherein we presented algorithms to efficiently select σ-bit sensors to maximize the identification score.

3.2 Localization with Sensor Errors

A σ-bit boolean string, representing a single sensor output, can be considered as one of the possible 2^σ symbols. The outputs of m sensors will then be a string of m such symbols. The number of locations at which the two strings of m symbols are different from each other is referred to as the *Hamming distance* between the strings, denoted by H(string 1, string 2). For instance, consider $\sigma = 2$, then there are four possible symbols, a = [0 0], b = [0 1], c = [1 0], and d = [1 1]. Let $m = 3$, then the hamming distance between two strings, [a b c] and [a c d] is two.

Next, we consider a scenario in which a set of m sensors, each having a σ-bit output, is deployed. Let $S(\ell_i)$ and $S(\ell_j)$ be the signatures corresponding to events ℓ_i and ℓ_j respectively. Both $S(\ell_i)$ and $S(\ell_j)$ consist of boolean strings of length σm (or a string of m symbols, where each symbol represents a σ-bit output). Moreover, we assume that at most e of the m sensors can be erroneous. In

the case of event ℓ_i or ℓ_j, an array of sensors' outputs $\tilde{S}(\ell_i)$ or $\tilde{S}(\ell_j)$, is generated in which at most e sensors give incorrect outputs. Let $\mathcal{S}(\ell_i)$ and $\mathcal{S}(\ell_j)$ be the set of all possible $\tilde{S}(\ell_i)$ and $\tilde{S}(\ell_j)$ respectively. Now, the question is that given $\tilde{S} \in \mathcal{S}(\ell_i) \cup \mathcal{S}(\ell_j)$, under what conditions can we distinguish between events ℓ_i and ℓ_j correctly through $\tilde{S}$? Or, in other words, when can we correctly map $\tilde{S}$ to the correct output which is either $S(\ell_i)$ or $S(\ell_j)$? To map (or decode) $\tilde{S}$ to the correct signature, we use the minimum distance decoding (MDD) principle, in which $\tilde{S}$ is mapped to the signature that is at the minimum hamming distance from $\tilde{S}$. Unlike the no error case, in which $\ell_{i,j}$ is either detected correctly or not detected at all, there is another possibility of incorrectly detecting $\ell_{i,j}$ here. For instance, $\tilde{S}(\ell_i)$ generated as a result of event ℓ_i is incorrectly mapped to $S(\ell_j)$ if

$$H(\tilde{S}(\ell_i), S(\ell_i)) > H(\tilde{S}(\ell_i), S(\ell_j)). \tag{7}$$

The following condition ensures that $\ell_{i,j}$ is always detected correctly.

Proposition 3.2. *A pair-wise event $\ell_{i,j}$ is always detected correctly in the presence of e erroneous sensors if and only if the Hamming distance between the signatures of ℓ_i and ℓ_j is at least $2e+1$.*

Proof. $H(S(\ell_i), S(\ell_j)) \geq 2e+1$ implies that $\mathcal{S}(\ell_i) \cap \mathcal{S}(\ell_j) = \emptyset$. Thus, $H(\tilde{S}(\ell_i), S(\ell_i)) < H(\tilde{S}(\ell_i), S(\ell_j))$, and therefore, $\tilde{S}(\ell_i)$ will be correctly mapped to $S(\ell_i)$. Similar is true for $\tilde{S}(\ell_j)$ due to ℓ_j. On the contrary, if $H(S(\ell_i), S(\ell_j)) < 2e+1$, there always exists an output $\tilde{S}(\ell_i)$ that satisfies $H(\tilde{S}(\ell_i), S(\ell_i)) \geq H(\tilde{S}(\ell_i), S(\ell_j))$, and therefore, is mapped incorrectly.

□

For illustration, consider events $\{\ell_1, \ell_2, \ell_3\}$, and outputs of sensors 2, 3, and 4, that is $S = [S_2\ S_3\ S_4]$ in the example in Section 2.2. We define the following symbols corresponding to 2-bit outputs: a = [0 0], b = [0 1], c = [1 0], and d = [1 1]. The signatures corresponding to ℓ_1, ℓ_2, and ℓ_3 are $S(\ell_1)$ = [cbb], $S(\ell_2)$ = [bac], and $S(\ell_3)$ = [cca]. Considering $e = 1$, the set of all possible outputs corresponding to ℓ_1, ℓ_2, and ℓ_3 are shown in Figure 3. Since $H(S(\ell_1), S(\ell_2)) = 3$, we have $\mathcal{S}(\ell_1) \cap \mathcal{S}(\ell_2) = \emptyset$. As a result, the pair-wise event $\ell_{1,2}$ is always correctly detected. On the other hand, $H(S(\ell_1), S(\ell_3)) = 2$, and $\mathcal{S}(\ell_1) \cap \mathcal{S}(\ell_3)$ = {[ccb], [cba]}, which means that the pair-wise event $\ell_{1,3}$ cannot be detected in the case of sensor outputs [ccb] or [cba].

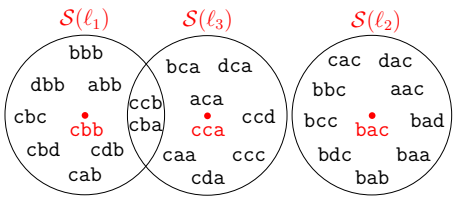


Figure 3: An example for detecting pair-wise events.

From Proposition 3.2, we know that a pair-wise event $\ell_{i,j}$ is always detected correctly if $H(S(\ell_i), S(\ell_j)) \geq (2e+1)$. However, even if $0 < H(S(\ell_i), S(\ell_j)) < (2e+1)$, then still there exist sensor outputs corresponding to ℓ_i and ℓ_j for which $\ell_{i,j}$ is detected correctly. In fact, greater the hamming distance between $S(\ell_i)$ and $S(\ell_j)$, higher will be the number of such outputs resulting in accurate detection of $\ell_{i,j}$. Thus, from a given set of sensors, our objective is to select a subset of m sensors, say $\mathcal{A}$ such that the Hamming distance between the the signatures of events ℓ_i and ℓ_j, $\forall i, j$, is maximized, under the condition that a subset of at most e sensors can be erroneous. More precisely, if we define

$$f(\ell_{i,j}) = \begin{cases} 1 & \text{if } H(S(\ell_i), S(\ell_j)) \geq 2e+1 \\ \frac{H(S(\ell_i), S(\ell_j))}{2e+1} & \text{otherwise.} \end{cases}, \tag{8}$$

then, our sensor placement problem can be written as

$$\underset{\mathcal{A}}{\text{argmax}} \left(\frac{\sum_{\ell_{i,j}} f(\ell_{i,j})}{\text{Total number of pair-wise links}} \right) \quad \text{subject to } |\mathcal{A}| \leq m. \tag{9}$$

In other words, the goal is to select m sensors such that for an arbitrary pair of link failures ℓ_i and ℓ_j, the number of sensors that have different outputs for ℓ_i and ℓ_j are as close to $(2e+1)$ as possible. For the ease of notation, we call the quantity below as the *identification score* of the sensor placement, and denote it by $\mathcal{I}_g$.

$$\mathcal{I}_g = \frac{\sum_{\ell_{i,j}} f(\ell_{i,j})}{\text{Total number of pair-wise links}} \tag{10}$$

For $e = 0$, $\mathcal{I}_g$ is exactly same as the identification score defined for the no sensor error case in [23], and the problem in (9) is equivalent to solving the maximum coverage problem on the pair-wise events. For $e \geq 1$, the setup remains exactly the same. However, instead of simply covering a pair-wise link failure only once, we need to cover it at least $2e+1$ times through the selection of sensors.

3.2.1 Sensor Placement Algorithm for Localization with Sensor Errors. Here, we discuss that the optimal sensor placement problem to maximize the localization of faults with a given upper bound on the number of erroneous sensors can be formulated as a well-studied *set multicover problem*. First, we define the problem and state the known results.

Definition 3.3. (Set Multicover Problem (SMP)) Given a set of elements $\mathcal{U}$, a collection $\mathcal{C} = \{C_1, C_2, \cdots, C_r\}$ of subsets $\mathcal{U}$, and a positive integer k. The SMP is to select a minimum sub-collection $\mathcal{C}_k \subseteq \mathcal{C}$ such that for every $x \in \mathcal{U}$, we get $|C_j \in \mathcal{C}_s : x \in C_j| \geq k$.

For $k = 1$, the problem is a well known *set cover problem*, which is NP-hard and cannot be approximated in polynomial time to within a factor of $(1-\epsilon)\ln|\mathcal{U}|$ for any constant $0 < \epsilon < 1$ (unless P=NP) [9]. On the other hand, for any k, SMP can be approximated to within the factor $(1 + \ln d)$ using a simple greedy approach [28]. Here, d is the size of the largest subset in $\mathcal{C}$. Greedy approach is the one in which at every step, a subset from $\mathcal{C}$ is selected that covers the maximum number of elements that has not been covered k times in the previous steps. In [2], a randomized approximation algorithm with a slight improved performance is presented with an expected approximation ratio of $(1 + o(1))\ln\frac{d}{k}$ when d/k is atleast 7.39, and $1 + 2\sqrt{d/k}$ for smaller d/k.

For our sensor placement problem, let $X_{i,j} \subseteq \mathcal{L}$ be the set of link failures that can be detected by the j^{th} output bit of a σ-bit sensor

when placed at the location (node) i, and $X_i = \cup_{j=1}^{\sigma} X_{i,j}$. Given $X_{i,j}$ for all possible sensor locations $i \in \{1, 2, \cdots, r\}$, $j \in \{1, 2, \cdots, \sigma\}$, and the maximum number of erroneous sensors e, the objective is to select m sensor locations from a set of r possible locations so that the number of pair-wise link failures that can be detected correctly in the presence of at most e erroneous sensors is maximized.

A greedy heuristic that approximately solves this problem using the set multicover formulation is as follows:

(1) For each (sensor) location i, compute $C_{i,j}$, which is the set of pair-wise events detected by the j^{th} output bit of the sensor placed at the location i. Note that the pair-wise event $\ell_{x,y} \in C_{i,j}$ if and only if $|\{\ell_x, \ell_y\} \cap X_{i,j}| = 1$, that is either $\ell_x \in X_{i,j}$, or $\ell_y \in X_{i,j}$. Next for each i, compute $C_i = \cup_{j=1}^{\sigma} C_{i,j}$. Define $\mathcal{C} = \{C_1, C_2, \cdots, C_r\}$.
(2) Iteratively select $C_i \in \mathcal{C}$ that contains 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, and select sensors at locations corresponding to the selected C_i's.

For a total of n link failures, the time complexity of the above greedy heuristic is $O(nmk)$.

In a given set, the maximum number of sensors that can give errors is a reflection of the reliability of sensors in the set. Based on the fraction of sensors with errors, the sensors in a given set can be attributed to a certain type that can be characterized by the ratio e/m. An interesting consideration here is the placement of sensors with different types. In this direction, consider a sensor placement in which two groups of sensors are placed. The first group has a total of m_1 sensors, out of which at most e_1 can be erroneous, and the second group has a total of m_2 sensors with at most e_2 erroneous ones. Note that the ratios e_1/m_1 and e_2/m_2 might be different. Assuming that each sensor is a σ-bit sensor, the signature corresponding to an event ℓ_i consists of $\sigma(m_1 + m_2)$ bits and can be divided into two parts $S(\ell_i) = [S^a(\ell_i)\ S^b(\ell_i)]$, where $S^a(\ell_i)$ and $S^b(\ell_i)$ are the parts corresponding to the outputs of sensors outputs in the first and second groups respectively. Similarly, the actual output of sensors as a result of ℓ_i has two parts $\tilde{S}(\ell_i) = [\tilde{S}^a(\ell_i)\ \tilde{S}^b(\ell_i)]$. Under this setup, we get the following result.

Proposition 3.4. *Consider two groups of sensors, denoted by a and b, where group a contains m_1 sensors out of which at most e_1 can be erroneous, and group b contains m_2 sensors out of which at most e_2 can be erroneous. A pair-wise event $\ell_{i,j}$ is always detected correctly if and only if at least one of the following is true*

$$H(S^a(\ell_i), S^a(\ell_j)) \geq 2e_1 + 1, \tag{11}$$

$$H(S^b(\ell_i), S^b(\ell_j)) \geq 2e_2 + 1. \tag{12}$$

Proof. For a given $\ell_{i,j}$, $S(\ell_i) = [S^a(\ell_i)\ S^b(\ell_i)]$, and $S(\ell_j) = [S^a(\ell_j)\ S^b(\ell_j)]$, first, we observe that $\mathcal{S}(\ell_i) \cap \mathcal{S}(\ell_j) = \emptyset$ if and only if at least one of the above two conditions is true. Then, using the same argument as in Proposition 3.2, the claim follows directly. □

Given two such groups containing m_1 and m_2 sensors respectively with e_1 and e_2 specified. A simple and effective sensor placement strategy to maximize the number of pair-wise events that can be detected correctly can be obtained by running the previous sensor placement algorithm twice. First, place m_1 sensors from the first group that maximize the number of pair-wise events that are covered by at least $2e_1 + 1$ sensors. Then, consider only the pair-wise events that remained uncovered by at least $2e_1 + 1$ sensors in the first step. Next, again using the sensor placement algorithm, place m_2 sensors from the other group to maximize the pair-wise events that are covered by at least $2e_2 + 1$ sensors. Note that the step-wise sensor placement strategy can be extended to sensors divided into any number of groups.

4 LOCALIZATION PERFORMANCE METRICS

In this section, we propose different metrics along with the *generalized identification score* in (10) to measure the localization performance of the sensor placement. In particular, we define the notions of *good pair-wise events* and the *localization set size* below.

4.1 Detection of Good Pair-wise Events

For a given sensor placement and a fixed number of erroneous sensors e, the output $\tilde{S}(\ell_i)$ generated as a result of event ℓ_i is at most e hamming distance away from the signature $S(\ell_i)$. Consequently, one of (13), (14), or (15) is always true for a pair of events ℓ_i and ℓ_j.

$$H(\tilde{S}(\ell_i), S(\ell_i)) > H(\tilde{S}(\ell_i), S(\ell_j)) \tag{13}$$

$$H(\tilde{S}(\ell_i), S(\ell_i)) < H(\tilde{S}(\ell_i), S(\ell_j)) \tag{14}$$

$$H(\tilde{S}(\ell_i), S(\ell_i)) = H(\tilde{S}(\ell_i), S(\ell_j)) \tag{15}$$

For a fixed ℓ_i, there are $(n-1)$ pair-wise events $\ell_{i,j}$, which can be partitioned into three categories based on the outputs corresponding to ℓ_i.

(1) *Bad pair-wise events* – These are the pair-wise events for which there exist some $\tilde{S}(\ell_i)$ satisfying (13). In other words, there exists an output corresponding to ℓ_i that indicates the occurrence of ℓ_j in the case of event ℓ_i, thus detecting the pair-wise event incorrectly. We denote the fraction of bad pair-wise events corresponding to ℓ_i by $\mathcal{B}(\ell_i)$.
(2) *Good pair-wise events* – These are the pair-wise events which are *always* detected correctly. As a result, for any $\tilde{S}(\ell_i)$, (14) is always satisfied. We denote the fraction of good pair-wise events corresponding to ℓ_i by $\mathcal{G}(\ell_i)$.
(3) *Neutral pair-wise events* – These are the pair-wise events that are neither good, nor bad at the same time. In other words, there exists an $\tilde{S}(\ell_i)$ that satisfies (15). We represent the fraction of such pair-wise events by $\mathcal{N}(\ell_i)$.

If all the events and corresponding outputs are equally likely, then the probabilities that an arbitrary pair-wise event is bad, good, or neutral are given by (16), (17), and (18) respectively.

$$\mathcal{B} = \sum_{\ell_i} \mathcal{B}(\ell_i)/n \tag{16}$$

$$\mathcal{G} = \sum_{\ell_i} \mathcal{G}(\ell_i)/n \tag{17}$$

$$\mathcal{N} = \sum_{\ell_i} \mathcal{N}(\ell_i)/n \tag{18}$$

Along with the generalized identification score, the values of $\mathcal{B}$, $\mathcal{G}$, and $\mathcal{N}$ also measure the quality of sensor placement for localization. In the best possible case, all pair-wise events are good ones and $\mathcal{G} = 1$. As a result, one of the goals of sensor placement is to maximize $\mathcal{G}$ and minimize $\mathcal{B}$ values.

4.2 Localization Sets and Uncorrectable Outputs

Under MDD, $\tilde{S}(\ell_i)$ generated as a result of event ℓ_i is mapped to some signature $S(\ell_j)$ that is at a minimum Hamming distance from $\tilde{S}(\ell_i)$. If $h = \min_{S(\ell_j)} H(\tilde{S}(\ell_i), S(\ell_j))$, then $\tilde{S}(\ell_i)$ can be mapped to any signature in

$$Q_{\tilde{S}(\ell_i)} = \{S(\ell_j) : H(\tilde{S}(\ell_i), S(\ell_j)\} = h\}. \quad (19)$$

If $S(\ell_i) \in Q_{\tilde{S}(\ell_i)}$, then we say that $Q_{\tilde{S}(\ell_i)}$ *is the localization set of* $\tilde{S}(\ell_i)$. If $S(\ell_i) \notin Q_{\tilde{S}(\ell_i)}$, then $\tilde{S}(\ell_i)$ *is the uncorrectable output.*

The foremost consideration in designing sensor placement with erroneous sensors is to minimize the number of uncorrectable outputs. At the same time, it is desired to minimize the cardinality of localization sets corresponding to sensor outputs. For instance, consider the case of 1-bit sensors, in which for any ℓ_i, there are $\sum_{j=0}^{e} \binom{m}{j}$ possibilities of $\tilde{S}(\ell_i)$. Assuming that all outputs are equally likely, the probability that $\tilde{S}(\ell_i)$ is uncorrecatble is given by

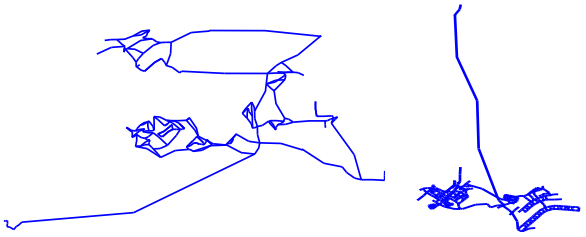
$$\mathcal{E}_{\tilde{S}(\ell_i)} \triangleq \frac{\text{\# of uncorrectable outputs corresponding to } \ell_i}{\sum_{j=0}^{e} \binom{m}{j}}. \quad (20)$$

Assuming that all events and all outputs are equally likely, the probability of an output to be uncorrectable is given by (21). Sensor placements resulting in smaller values of $\mathcal{E}$ are desirable as compared to the ones resulting in higher values of $\mathcal{E}$.

$$\mathcal{E} = \frac{\sum_{\forall \ell_i} \mathcal{E}_{\tilde{S}(\ell_i)}}{\text{\# of events}} \quad (21)$$

5 NUMERICAL EVALUATION

Here, we evaluate our approach on two water distribution networks. Water network 1(WN-1) [21] is a benchmark network previously introduced in Section 2.1, and Water network 2 (WN-2) [12] is a grid system in Kentucky with 366 pipes, 270 nodes, three tanks, and five pumps, and its layout is shown in Figure 4(b). We consider that the pressure sensors are placed at the nodes to detect the pressure signals generated as a result of pipe bursts. For all simulations, a failure event is a pipe burst occurring at the center of each pipe. The detection of transient pressure signal by a sensor is approximated by a distance threshold model [6, 23], in which the j^{th} output bit of a σ-bit sensor is 1 if the distance between the location of failure event and the sensor lies within the interval $[\epsilon_{j-1}\ \epsilon_j]$. In the case of 1-bit sensors, $\epsilon_1 = 1000[m]$, and for the 2-bit sensors the distance thresholds are $\epsilon_1 = 500[m]$, $\epsilon_2 = 1000[m]$. For both networks $\epsilon_0 = 0[m]$. Assuming that σ-bit sensors can be placed at any of the nodes within the network, we evaluate our approach using metrics defined in the previous section.



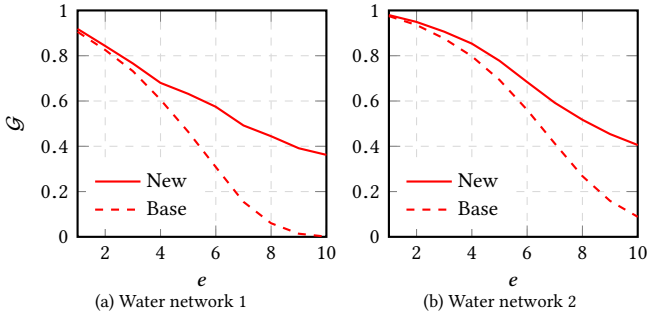
(a) Water network 1 (b) Water network 2

Figure 4: Layout of the water-distribution networks.

Table 2: Comparison of sensor placements for $\mathcal{B}$, $\mathcal{G}$, and $\mathcal{N}$ in water network 1 ($m = 30$).

	$\mathcal{B}$		$\mathcal{G}$		$\mathcal{N}$	
	New	Base	New	Base	New	Base
$e = 2$	.0781	.0973	.843	.8254	.078	.0773
$e = 3$	.1603	.201	.7659	.7320	.0738	.067
$e = 4$	.23	.3214	.6783	.606	0.958	.0726

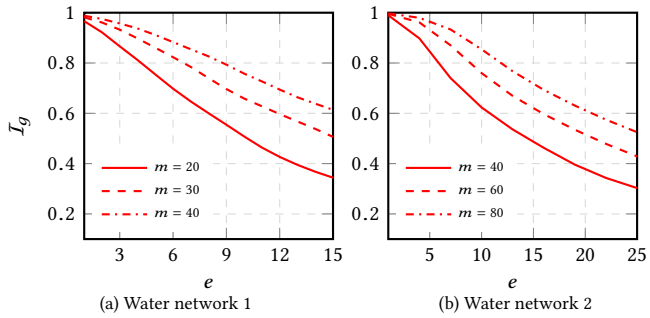
First, by computing $\mathcal{G}$, $\mathcal{B}$, and $\mathcal{N}$, we compare sensor placement using the proposed approach with the one in which sensor errors are not considered. We call the approach with $e = 0$ as the *base case* strategy. For the case of water network 1, $m = 30$, and $e = \{2, 3, 4\}$, the comparison of $\mathcal{B}$, $\mathcal{G}$ and $\mathcal{N}$ in Table 2 shows that the proposed approach clearly outperforms the base case, both in terms of minimizing $\mathcal{B}$ and maximizing $\mathcal{G}$. In fact, as e increases, improvements in terms of reducing the number of bad pair-wise events and increasing good pair-wise events due to the new approach become significant. In Figure 5, a comparison of new and base case approach is shown by plotting $\mathcal{G}$ as a function of e. Again, as e increases, the new approach significantly outperforms the base case approach.



(a) Water network 1 (b) Water network 2

Figure 5: $\mathcal{G}$ as a function of e for $m = 30$ and $m = 60$ in water networks 1 and 2 respectively.

Second, for the sensor placement using the proposed algorithm, we plot the generalized identification score $\mathcal{I}_g$ as a function of e for water networks 1 and 2 in Figures 6 (a) and (b) respectively. As expected, for fixed m, $\mathcal{I}_g$ decreases with increasing values of e. At the same time, for a fixed e, sensor placements involving more sensors result in higher values of $\mathcal{I}_g$.



(a) Water network 1 (b) Water network 2

Figure 6: $\mathcal{I}_g$ as a function of number of sensor attacks e for various m using the proposed approach.

Table 3: Comparison of uncorrectable outputs ($\mathcal{E}$) in water network 1.

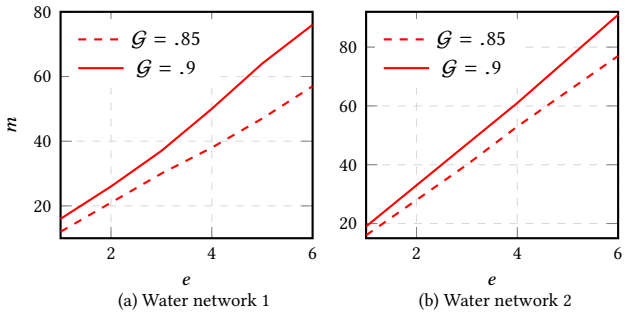
m	$\mathcal{E}$ ($e = 2$)		$\mathcal{E}$ ($e = 3$)	
	New	Base	New	Base
20	0.19	0.253	0.299	0.38
30	0.11	0.16	0.175	0.24
40	0.057	0.11	0.084	0.181

Third, for sensor placements using the proposed and the base case strategy, we compare the values of $\mathcal{E}$, as defined in (21). We compute $\mathcal{E}$ for various m and $e \in \{2, 3\}$ in the case of water network 1. The values are shown in Table 3. For a fixed m, $\mathcal{E}$ obtained using the new approach is always lesser than the one obtained using the base case approach. Hence, the new approach clearly outperforms the base case approach in all the cases. Moreover, for the same m, the difference between $\mathcal{E}$ values obtained using the new approach and the base case approach increases with the higher values of e, thus making the new approach particularly suitable for the situations involving higher number of sensor errors.

In Figure 7, for both water networks and various values of m, we illustrate the fraction of outputs having a localization set of a particular size while considering $e = 2$. To compute them, we pick an event ℓ_i, and assuming that all corresponding outputs $\tilde{S}(\ell_i)$ are equally likely, we compute the fraction of outputs corresponding to ℓ_i that have a localization set of a particular size, say z. We do this for all events ℓ_i, and then take the average, which basically gives the probability of an output to have a localization set of size z. We repeat this for all possible values of localization sets' sizes. These values are plotted in Figures 7(a) and (b) for the localization sets of sizes between 1 and 8. We observe that as the ratio m/e increases, the number of outputs that have localization sets of smaller sizes also increase. For instance, in WN-2, the percentage of outputs that have localization sets of size at most 5 is about 90%, 80%, and 58% for $m = 80$, 60, and 40 respectively. Similarly, in the case of WN-1, these percentages are approximately 60%, 52%, and 34% for $m = 40$, 30, and 20 respectively. Finally, the average localization set sizes in WN-2 are 2.2 for $m = 80$, 2.8 for $m = 60$, 5.1 for $m = 40$. Similalr, in WN-1, the average localization set sizes are 4.5 for $m = 40$, 5.2 for $m = 30$, and 8.1 for $m = 20$.

Comparison of Sensor Configurations. Next, we discuss the effect of varying the number of sensors m and the maximum number of erroneous sensors e on the localization performance as measured by the fraction of good pair-wise events $\mathcal{G}$. We also discuss the placement of different types of sensors to maximize the localization of pipe failure events.

First, we plot m as a function of e using 2-bit sensors while fixing the localization performance $\mathcal{G}$ in Figure 8. For both networks, we observe that m increases (almost) linearly with e, and the ratio e/m approximately remains the same. For instance, in the case of water network 2, e/m is about 0.08 and 0.065 for $\mathcal{G}$ of 0.85 and 0.9 respectively. Here, e/m, which is the fraction of sensors that can be erroneous, also reflects the reliability of a given set of sensors. A particular value of e/m can be associated with every $\mathcal{G}$. Thus, we can express a reliability specification in terms of e/m, which if satisfied, ensures that a particular value of $\mathcal{G}$ is achieved. We note here that the similar trend – constant e/m for a specific $\mathcal{G}$ – is observed with sensors having different σ's.



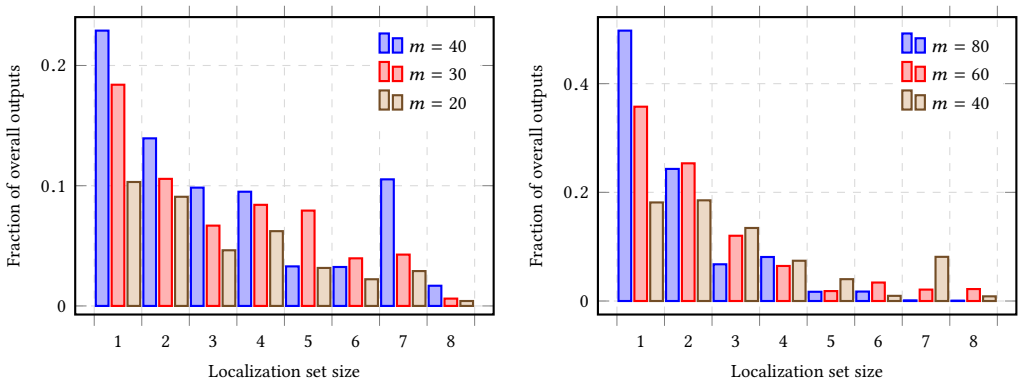
(a) Water network 1 (b) Water network 2

Figure 8: m as a function of e for fixed $\mathcal{G}$.

Next, in Figure 9, we illustrate variation in $\mathcal{G}$ with m for a fixed e. As m increases, e/m decreases, and we expect an increase in $\mathcal{G}$, which is indeed the case. In Figure 10, we illustrate the performance of sensor placement in terms of $\mathcal{G}$ when two different groups of sensors are used. Here, we fix e_1, m_1, e_2, and plot $\mathcal{G}$ as a function of m_2. As an example, consider water network 1 and $\mathcal{G} = 0.8$. From the plot in Figure 9(a), we see that 26 (2-bit) sensors, out of which at most 3 can be erroneous, can be placed to achieve the desired $\mathcal{G}$. At the same time, sensors from two groups; first containing 30 sensors with at most 4 erroneous ones, and second containing 12 in which a maximum of 3 can be erroneous, can be placed to achieve the same value of $\mathcal{G}$.

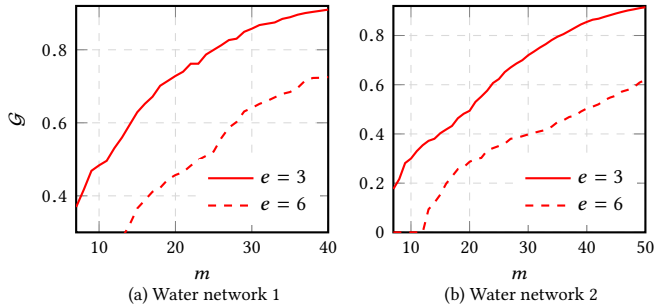
6 RELATED WORK

To make network operations resilient, the problem of placing sensing devices to efficiently detect and localize events, including faults and failures, arise in the context of many different networks including energy distribution networks, transportation systems, and water distribution systems. In the urban water sector, majority of previous works focused on the sensor placement for detecting potential contaminants in the water distribution systems. Using deterministic, stochastic, and combinatorial optimization techniques, sensor placement algorithms were obtained to optimize one or more objectives such as affected population, detection likelihood, expected



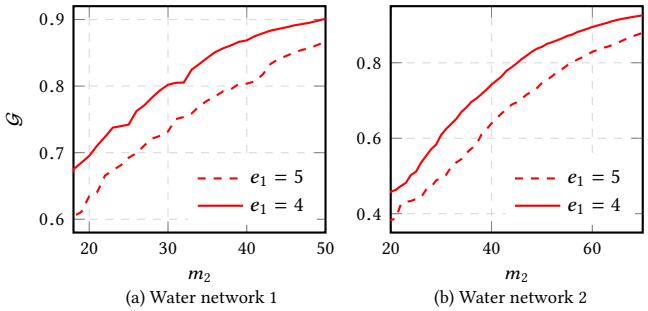
(a) Water network 1 (b) Water network 2

Figure 7: Fraction of overall outputs as a function of sizes of their localization sets



(a) Water network 1 (b) Water network 2

Figure 9: $\mathcal{G}$ as a function of m for fixed e.



(a) Water network 1 (b) Water network 2

Figure 10: $\mathcal{G}$ as a function of m_2 for a fixed $m_1 = 12$ and $e_1 = 3$.

contaminated water volume, and overall design cost [3, 21, 22]. For leakage detection purposes, model based techniques have been developed that are mostly employed on the operational side to effectively utilize available measurements along with the available system model to determine faults within a system (e.g., [4, 8, 24, 26]). In contrast, there has been a little work on the side of developing online monitoring systems equipped with sensing devices deployed within the network to enable remote detection and localization of pipe failure events [13, 29]. Our approach could be used towards an online decision support tool that remotely detects and localizes pipe failure events.

In a general setup, fault localization problems using measurements from sensing devices are closely related to the *group testing problems*, which have been studied widely. In group testing, the objective is to determine a subset of elements that are 'defective' in some way from a set of given elements by asking queries such as if a *group* of elements contains any defective element? By collecting yes/no responses to such queries, the task is to determine the subset of defective elements through a minimum number of queries or tests. The strategy in which all tests are designed a priori before the start of experiment is known as the *non-adaptive group testing (NAGT)*, whereas, the one in which each new test is designed by considering the outcomes of previous tests is known as the *adaptive group testing (AGT)* (e.g., see [5, 11]). If n is the total number of elements out of which at most d are defective, then using the non-adaptive strategy, at least $m = O(\frac{d^2}{\log d} \log n)$ tests (or queries) are required [7]. Non-adaptive group testing schemes that determine the defective elements in $m = O(d^2 \log n)$ are known (e.g., [11, 25]). In a variant of the problem, a certain number of test outcomes are allowed to be erroneous, and the problem is referred to as the *group testing with errors*, or *group testing with unreliable tests*, which has been studied under various error models (e.g., [14, 17–20]).

The problem in this paper is related to NAGT with errors. Here, links are the set of elements, failed link is the defective element that needs to be localized, and each sensor output is a test output since a sensor output notifies if the failed link belongs to a certain subset of links – links that are at a certain distance from the sensor. However, there is a major difference between NAGT with errors and link failure localization through sensors with possible errors. In NAGT, any subset of elements can be grouped together to design the most appropriate tests. However, in link failure localization through sensors, the links at which failures can be detected by the sensors cannot be arbitrarily chosen. In fact, the dynamics of the physical process define the set of links at which failures can be detected by a particular sensor. Thus, unlike general NAGT problem, tests cannot be arbitrarily designed in terms of grouping elements into tests.

In [10], for the diagnosis of optical link failures, a relevant notion of *combinatorial group testing on graphs* was defined in which certain constraints were posed to select the elements into tests. As with our setup in this paper, elements in the test were the links in a

graph. However, the specific conditions to include links in a particular test were different. In their work, only the links in a sub-tree that could be traversed at most once in each direction constituted a test. In contrast to that, in our work, the links included in a test (sensor) do not have to form a sub-tree. In fact, all links that are at a certain distance from the node at which the sensor is placed are included in the corresponding test.

7 CONCLUSIONS

In this paper, we proposed a sensor placement scheme that maximized the localization of pipe failure events in water networks. Instead of assuming that all sensors were perfect and always gave correct outputs, we considered that a subset of sensors could give errors. These errors could be the result of attacks on sensors, in which the attacker corrupted sensors' outputs after compromising them, or errors could be the result of degradation in sensor hardware or software. Using combinatorial setting, we posed the sensor placement problem as a set multicover problem, and presented a greedy heuristic to solve the problem. We compared our sensor placement solution to the one that did not consider any sensor errors, and observed significant improvements in the localization performance. Further, we explored trade-offs between the total number of sensors, number of erroneous sensors, and the number of output bits in a sensor on the localization performance by performing numerical experiments on two water distribution networks. Our approach could be used to design a decision support tool for water utilities to detect and localize faults in an online manner.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (award numbers: CNS-1453126, CNS-1238959, CNS-1238962, CNS-1239054, CNS-1239166), Air Force Research Laboratory (contract ID: FA8750-14-2-0180, SUB 2784-018400), and National Institute of Standards and Technology (70NANB15H263).

REFERENCES

[1] Waseem Abbas, Lina Sela Perelman, Saurabh Amin, and Xenofon Koutsoukos. 2015. An Efficient Approach to Fault Identification in Urban Water Networks Using Multi-Level Sensing. In *Proc. of the 2nd ACM Intl. Conf. on Embedded Systems for Energy-Efficient Built Environments (BuildSys)*. 147–156.

[2] Piotr Berman, Bhaskar DasGupta, and Eduardo Sontag. 2007. Randomized approximation algorithms for set multicover problems with applications to reverse engineering of protein and gene networks. *Discrete Applied Mathematics* 155, 6 (2007), 733–749.

[3] Jonathan Berry, William E Hart, Cynthia A Phillips, James G Uber, and Jean-Paul Watson. 2006. Sensor placement in municipal water networks with temporal integer programming models. *Journal of Water Resources Planning and Management* 132, 4 (2006), 218–224.

[4] Myrna V Casillas, Vicen Puig, Luis E Garza-Castanón, and Albert Rosich. 2013. Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* 13, 11 (2013), 14984–15005.

[5] Chun Lam Chan, Sidharth Jaggi, Venkatesh Saligrama, and Samar Agnihotri. 2014. Non-adaptive group testing: Explicit bounds and novel algorithms. *IEEE Transactions on Information Theory* 60, 5 (2014), 3019–3035.

[6] Ajay Deshpande, Sanjay E Sarma, Kamal Youcef-Toumi, and Samir Mekid. 2013. Optimal coverage of an infrastructure network using sensors with distance-decaying sensing quality. *Automatica* 49, 11 (2013), 3351–3358.

[7] AG Dyachkov and VV Rykov. 1983. A survey of superimposed code theory. *Problems of Control and Information Theory* 12, 4 (1983), 1–13.

[8] Demetrios G Eliades and Marios M Polycarpou. 2010. A fault diagnosis and security framework for water systems. *IEEE Transactions on Control Systems Technology* 18, 6 (2010), 1254–1265.

[9] Uriel Feige. 1998. A threshold of ln n for approximating set cover. *J. ACM* 45, 4 (1998), 634–652.

[10] Nicholas JA Harvey, Mihai Patrascu, Yonggang Wen, Sergey Yekhanin, and Vincent WS Chan. 2007. Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs. In *26th IEEE International Conference on Computer Communications (INFOCOM)*.

[11] Frank K Hwang and Du Ding-Zhu. 2000. *Combinatorial Group Testing and its Applications*. World Scientific.

[12] Matthew D Jolly, Amanda D Lothes, L Sebastian Bryson, and Lindell Ormsbee. 2013. Research database of water distribution system models. *Journal of Water Resources Planning and Management* 140, 4 (2013), 410–416.

[13] Sokratis Kartakis, Edo Abraham, and Julie A McCann. 2015. Waterbox: A testbed for monitoring and controlling smart water networks. In *Proc. of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks*.

[14] Emanuel Knill, William J Bruno, and David C Torney. 1998. Non-adaptive group testing in the presence of errors. *Discrete Applied Mathematics* 88 (1998), 261–290.

[15] Andreas Krause, Carlos Guestrin, Anupam Gupta, and Jon Kleinberg. 2011. Robust sensor placements at informative and communication-efficient locations. *ACM Transactions on Sensor Networks* 7, 4 (2011).

[16] Andreas Krause, Jure Leskovec, Carlos Guestrin, Jeanne VanBriesen, and Christos Faloutsos. 2008. Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management* 134, 6 (2008), 516–526.

[17] Anthony J Macula. 1997. Error-correcting nonadaptive group testing with d^e-disjunct matrices. *Discrete Applied Mathematics* 80, 2 (1997), 217–222.

[18] Arya Mazumdar. 2012. On almost disjunct matrices for group testing. In *International Symposium on Algorithms and Computation (ISAAC)*. Springer, 649–658.

[19] Arya Mazumdar and Soheil Mohajer. 2014. Group Testing with Unreliable Elements. In *52nd Annual Allerton Conf. on Communication, Control, and Computing*.

[20] Hung Q Ngo and Ding-Zhu Du. 2002. New constructions of non-adaptive and error-tolerance pooling designs. *Discrete Mathematics* 243, 1 (2002), 161–170.

[21] Avi Ostfeld, James G Uber, Elad Salomons, Jonathan W Berry, William E Hart, Cindy A Phillips, Jean-Paul Watson, Gianluca Dorini, Philip Jonkergouw, Zoran Kapelan, and others. 2008. The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms. *Journal of Water Resources Planning and Management* 134, 6 (2008), 556–568.

[22] Lina Perelman, Jonathan Arad, Mashor Housh, and Avi Ostfeld. 2012. Event detection in water distribution systems from multivariate water quality time series. *Environmental Science & Technology* 46, 15 (2012), 8212–8219.

[23] Lina Sela Perelman, Waseem Abbas, Xenofon Koutsoukos, and Saurabh Amin. 2016. Sensor placement for fault location identification in water networks: a minimum test cover approach. *Automatica* 72 (2016), 166–176.

[24] Ramon Perez, Gerard Sanz, Vicenc Puig, Joseba Quevedo, Miquel Angel Cuguero Escofet, Fatiha Nejjari, Jordi Meseguer, Gabriela Cembrano, Josep M Mirats Tur, and Ramon Sarrate. 2014. Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in Barcelona. *IEEE Control Systems* 34, 4 (2014), 24–36.

[25] Ely Porat and Amir Rothschild. 2011. Explicit nonadaptive combinatorial group testing schemes. *IEEE Transactions on Information Theory* 57, 12 (2011).

[26] R Puust, Z Kapelan, DA Savic, and T Koppel. 2010. A review of methods for leakage management in pipe networks. *Urban Water Journal* 7, 1 (2010), 25–45.

[27] Ivan Stoianov, Lama Nachman, Sam Madden, and Timur Tokmouline. 2007. PIPENET: A wireless sensor network for pipeline monitoring. In *2007 6th International Symposium on Information Processing in Sensor Networks (IPSN)*. 264–273.

[28] Vijay V Vazirani. 2001. *Approximation Algorithms*. Springer Verlag, Berlin.

[29] AJ Whittle, M Allen, A Preis, and M Iqbal. 2013. Sensor networks for monitoring and control of water distribution systems. In *Proc. of the 6th Intl. Conf. on Structural Health Monitoring of Intelligent Infrastructure (SHMII)*.

[30] E. B. Wylie, V. L. Streeter, and L. Suo. 1993. *Fluid Transients in Systems*. Prentice Hall.