OPEN ACCESS

Conference Proceedings Paper–Sensors and Applications

Effectively Positioning Water Loss Event in Smart Water Networks

Weiren Yu* and Julie A. McCann

Department of Computing, Imperial College London, South Kensington Campus, London, SW7 2AZ;
E-Mails: {weiren.yu | j.mccann}@imperial.ac.uk

***** Author to whom correspondence should be addressed; E-Mail: weiren.yu@imperial.ac.uk (W.Y.)

Version November 10, 2015 submitted to Sensors. *Typeset by LaTeX using class file mdpi.cls*

Abstract: With the eye-catching advances in sensing technologies, smart water networks have been attracting immense research interest in recent years. One of the most overarching tasks in smart water network management is the reduction of water loss (such as leaks and bursts in a pipe network). In this paper, we propose an efficient scheme to position water loss event based on water network topology. The state-of-the-art approach to this problem, however, utilizes the limited topology information of the water network, that is, only one single shortest path between two sensor locations. Consequently, the accuracy of positioning water loss events is still less desirable. To resolve this problem, our scheme consists of two key ingredients: First, we design a novel graph topology-based measure, which can recursively quantify the "average distances" for all pairs of senor locations simultaneously in a water network. This measure will substantially improve the accuracy of our positioning strategy, by capturing the entire water network topology information between every two sensor locations, yet without any sacrifice of computational efficiency. Then, we devise an efficient search algorithm that combines the "average distances" with the difference in the arrival times of the pressure variations detected at sensor locations. The viable experimental evaluations on real-world test bed (WaterWiSe@SG) demonstrate that our proposed positioning scheme can identify water loss event more accurately than the best-known competitor.

Keywords: water loss event; graph topology; smart water network; sensing technologies

---

1. Introduction

The advent of sensing technologies in water supply systems has led to an increasing need for the development of smart data technologies in water resource management. Today, water loss has become a serious problem for almost all urban areas around the world [1], and it can be even worse in areas with scarcity of water. As a statistical example, the water industry in England and Wales loses 3.36 billion liters of water a day in leaks [2]. If those leaking locations were found as early as possible, sufficient water resources could be saved to supply 22.4 million people. However, it is often difficult to position such water loss events accurately as (a) the supply pipe is usually buried at least 3 feet below the ground surface, and (b) there are typically many paths connected by pipe sections between two pipe junctions. Therefore, it is imperative for us to devise an efficient model that can position water loss event automatically and accurately in a real water supply system.

1.1. Prior Work

Over the last decade, there have been several pioneering approaches proposed for water leak or burst localization, such as gradient intersection methods [3,4], wave propagation analysis [5], spectral clustering [6], and multiple hypotheses testing [7] (see [8] for a survey). Nonetheless, only a paucity of methods have been proposed in the context of a water network structure that explores graph topology.

One excellent piece of work is due to Misiunas *et al.* [9] who leveraged a search-based technique to localize a burst point. Its main idea consists of two phases: in the first phase, the search is performed globally over all nodes in the network; in the second phase, if the burst is inferred to have occurred along the pipe, extra nodes are placed along each of the pipes and the global search is repeated. However, both steps of this method require to perform a global search over all sensor locations. Hence, its computational efficiency is cost-inhibitive especially when a water network has high density of nodes.

Recently, Srirangarajan *et al.* [10] proposed an interesting technique that utilizes wave-based multiscale analysis of the pressure signal to detect burst transients. To identify the location of water burst events, they also exploited the Dijkstra's algorithm [11] for calculating the shortest distance between every two sensor locations. Nevertheless, we observe that, when a burst occurs, its wave may travel in all the possible directions of the paths (rather than only the paths with the shortest distance) from the burst location to the measurement points. Thus, in order to accurately position water loss events, it seems not appropriate to rely on only the shortest travel time between every two sensor locations.

1.2. Our Contributions

To resolve the above limitations, in this paper, we propose an efficient scheme that can position water loss event more accurately based on water network topology. Our main contributions can be summarized as follows:

- We first devise a novel graph topology-based measure, which can recursively quantify the "average distance" between every two senor locations simultaneously in a water network. This measure can significantly improve the accuracy of positioning water loss events, in that it can capture the

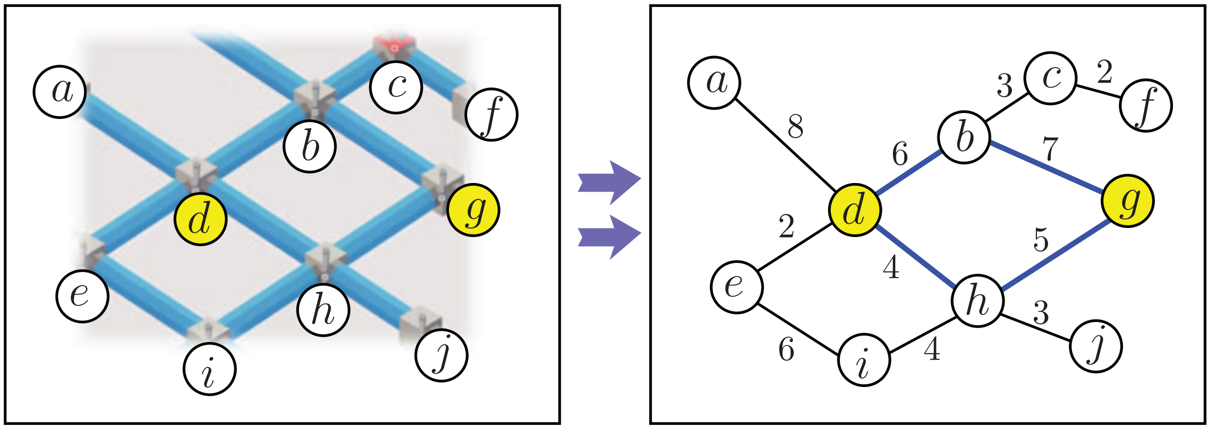

Figure 1. Modelling a water network (left) as a weighted graph (right) based on topology

multi-faceted relationships among senor locations in a global manner, yet without any sacrifice of computational efficiency. (Section 2.1)

- We next propose a fast and accurate search algorithm to efficiently position water loss events, which utilizes our "average distance" measure to determine the difference in the arrival times of the pressure variations detected at sensor locations. (Section 2.2)

The viable experimental evaluations on a real-world test bed demonstrate that our proposed scheme can identify water loss event more accurately than the state-of-the-art competitor. (Section 3)

2. The Proposed Model for Positioning Water Loss Event

We first devise a novel graph topology-based measure that can effectively quantify the "average distance" between senor locations, and then propose our search algorithm to position water loss events.

2.1. A Graph Topology-Based Measure

A water network can be modelled as a graph. Let $G = (V_J \cup V_S, E, A)$ be an attributed water network, where V_J is a vertex set of pipe junctions, V_S is a vertex set of deployed sensor locations, E denotes an edge set of pipe sections connecting two vertices, and A carries the length of each pipe section.

To evaluate the average distance between every two vertices over graph G, we first introduce the notions of *the distance matrix* $\mathbf{D}$ and *the adjacency matrix* $\mathbf{A}$.

Definition 1. *Given a water network* $G = (V_J \cup V_S, E, A)$ *with* $|V| = |V_J| + |V_S|$ *vertices and* $|E|$ *edges, its distance matrix* $\mathbf{D}$ *is a* $|V| \times |V|$ *matrix, whose element* $\mathbf{D}_{u,v}$ *is defined as*

$$\mathbf{D}_{u,v} = \begin{cases} \text{the length of pipe section } (u,v), & \text{if } u \neq v \text{ and } \exists \text{ pipe section } (u,v) \in E; \\ 0, & \text{otherwise}. \end{cases}$$

The adjacency matrix of G*, denoted as* $\mathbf{A}$*, is defined by*

$$\mathbf{A}_{u,v} = \begin{cases} 1, & \text{if } u \neq v \text{ and } \exists \text{ pipe section } (u,v) \in E; \\ 0, & \text{otherwise}. \end{cases}$$

Example 1. *Consider the water network G in Figure 1, whose edge weights carry the length of each pipe section. By Definition 1, its distance matrix $\mathbf{D}$ and adjacency matrix $\mathbf{A}$ are as follows:*

$$\mathbf{D} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \end{array} \begin{array}{c} \begin{array}{cccccccccc} a & b & c & d & e & f & g & h & i & j \end{array} \\ \left[\begin{array}{cccccccccc} 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 6 & 0 & 0 & 7 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 8 & 6 & 0 & 0 & 2 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 5 & 0 & 4 & 3 \\ 0 & 0 & 0 & 0 & 6 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \end{array}\right] \end{array} \qquad \mathbf{A} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \end{array} \begin{array}{c} \begin{array}{cccccccccc} a & b & c & d & e & f & g & h & i & j \end{array} \\ \left[\begin{array}{cccccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \end{array}$$

Based on Definition 1, we notice that $\mathbf{D}$ and $\mathbf{A}$ are both symmetric matrices. Leveraging $\mathbf{D}$ and $\mathbf{A}$, we are now ready to determine the "average distance" between every two sensor locations on graph G.

Let us first introduce a $|V| \times |V|$ matrix, $\mathbf{W}^{(d)}$, whose element $[\mathbf{W}^{(d)}]_{(u,v)}$ denotes the "average distance" of all paths with d hops between vertices u and v. Then, $[\mathbf{W}^{(d)}]_{(u,v)}$ can be represented as

$$[\mathbf{W}^{(d)}]_{(u,v)} = \frac{\text{the sum of the pipe section lengths over all paths with } d \text{ hops between vertices } u \text{ and } v}{\text{the number of the paths with } d \text{ hops between vertices } u \text{ and } v}. \tag{1}$$

To obtain the denominator of Equation (1), we can directly use an elegant property in graph theory about the power of an adjacency matrix: the (u, v)-th element of the d-th power of $\mathbf{A}$, that is, $[\mathbf{A}^d]_{(u,v)}$, counts the number of the paths with d hops between vertices u and v.

However, it is not easy to evaluate the nominator of Equation (1) as the power of a distance matrix can only evaluate the *product* (instead of *sum*) of the pipe section lengths over all paths. As an example, in Figure 1, to determine the *sum* of the pipe section lengths over all paths with 2 hops between vertices d and g, the result of $[\mathbf{D}^2]_{(d,g)}$ would produce the *product* of the pipe section lengths as follows:

$$\begin{aligned} [\mathbf{D}^2]_{(d,g)} &= (\text{the } d\text{-th row of } \mathbf{D}) \times (\text{the } g\text{-th column of } \mathbf{D}) \\ &= \left[\begin{array}{cccccccccc} \overset{a}{8} & \overset{b}{6} & \overset{c}{0} & \overset{d}{0} & \overset{e}{2} & \overset{f}{0} & \overset{g}{0} & \overset{h}{4} & \overset{i}{0} & \overset{j}{0} \end{array}\right] \cdot \left[\begin{array}{cccccccccc} \overset{a}{0} & \overset{b}{7} & \overset{c}{0} & \overset{d}{0} & \overset{e}{0} & \overset{f}{0} & \overset{g}{0} & \overset{h}{5} & \overset{i}{0} & \overset{j}{0} \end{array}\right]^T \\ &= 6 \times 7 + 5 \times 4 \neq \underbrace{(6+7)}_{d \to b \to g} + \underbrace{(5+4)}_{d \to h \to g} \end{aligned} \tag{2}$$

We notice that, if the "$\times$" sign in Equation (2) were changed into "$+$" sign, the result would desirably turn into the *sum* of the pipe section lengths over all paths ($d \to b \to g$ and $d \to h \to g$) with 2 hops between vertices d and g. To obtain the correct "$+$"-based results, can we still take good advantage of the power of a distance matrix while changing its "$\times$" sign (in Equation (2)) into "$+$" sign ?

To address this question, our technique is to introduce an element-wise operator $\exp(*)$. We construct the element-wise *exponential distance matrix*, denoted as $\exp(t\mathbf{D})$, as follows:

$$[\exp(t\mathbf{D})]_{u,v} = \begin{cases} \exp(t\mathbf{D}_{u,v}), & \text{if } \mathbf{D}_{u,v} \neq 0; \\ 0, & \text{if } \mathbf{D}_{u,v} = 0. \end{cases} \quad \text{where } t \in \mathbb{R} \text{ denotes an arbitrary scalar.}$$

Intuitively, the matrix $\exp(t\mathbf{D})$ is formed by replacing every nonzero element in $\mathbf{D}$, say x, with e^x, and keeping the zero elements of $\mathbf{D}$ unchanged.

Then, to assess the *sum* of the pipe section lengths over all paths with 2 hops between vertices d and g, we compute the (d,g)-th element of $(\exp(t\mathbf{D}))^2$, that is,

$$\begin{aligned}[(\exp(t\mathbf{D}))^2]_{(d,g)} &= (\text{the } d\text{-th row of } \exp(t\mathbf{D})) \times (\text{the } g\text{-th column of } \exp(t\mathbf{D})) \\ &= \left[\begin{array}{cccccccccc} \scriptstyle a & \scriptstyle b & \scriptstyle c & \scriptstyle d & \scriptstyle e & \scriptstyle f & \scriptstyle g & \scriptstyle h & \scriptstyle i & \scriptstyle j \\ e^{8t} & e^{6t} & 0 & 0 & e^{2t} & 0 & 0 & e^{4t} & 0 & 0 \end{array}\right] \cdot \left[\begin{array}{cccccccccc} \scriptstyle a & \scriptstyle b & \scriptstyle c & \scriptstyle d & \scriptstyle e & \scriptstyle f & \scriptstyle g & \scriptstyle h & \scriptstyle i & \scriptstyle j \\ 0 & e^{7t} & 0 & 0 & 0 & 0 & 0 & e^{5t} & 0 & 0 \end{array}\right]^T \\ &= e^{6t} \times e^{7t} + e^{5t} \times e^{4t} = e^{(6+7)t} + e^{(5+4)t} \end{aligned} \tag{3}$$

In contrast to Equation (2), we can see that, by utilizing the operator $\exp(*)$, Equation (3) converts all "$\times$" signs into "$+$" signs. In light of Equation (3), our next step is to find out an "inverse" operator that can map $e^{(6+7)t} + e^{(5+4)t}$ back into $(6+7)+(5+4)$. Our key observation is that

$$\lim_{t\to 0} \frac{2}{t} \log\left(\frac{1}{2}\left(e^{xt} + e^{yt}\right)\right) = x + y \tag{4}$$

Thus, applying the "inverse" operator $\lim_{t\to 0} \frac{2}{t} \log\left(\frac{1}{2}(*)\right)$ (in Equation (4)) into Equation (3) produces

$$\lim_{t\to 0} \frac{2}{t} \log\left(\frac{1}{2}\left([(\exp(t\mathbf{D}))^2]_{(d,g)}\right)\right) = \lim_{t\to 0} \frac{2}{t} \log\left(\frac{1}{2}\left(e^{(6+7)t} + e^{(5+4)t}\right)\right) = (6+7)+(5+4), \tag{5}$$

whose result gives the *sum* of the pipe section lengths over all paths ($d \to b \to g$ and $d \to h \to g$) with 2 hops between vertices d and g.

Equations (3)–(5) provide an effective technique to obtain the nominator of $[\mathbf{W}^{(d)}]_{(u,v)}$ in Equation (1). To generalize our above result for any arbitrary element of $(\exp(t\mathbf{D}))^2$, we need to extend the "inverse" operator in Equation (4) as follows.

Theorem 1. *For any positive integer* $N = 1, 2, \cdots$, *the following equation holds:*

$$\lim_{t\to 0} \frac{N}{t} \log\left(\frac{1}{N}\left(e^{x_1 t} + e^{x_2 t} + \cdots + e^{x_N t}\right)\right) = x_1 + x_2 + \cdots + x_N. \tag{6}$$

As a special case when $N = 2$, Theorem 1 reduces to the result in Equation (4). Theorem 1 is used for generalizing the result of Equation (3) for any *arbitrary* element of $(\exp(t\mathbf{D}))^k$. More specifically, in our aforementioned example, we choose Equation (4) (that is, $N = 2$ in Equation (6)) to "inverse" $[(\exp(t\mathbf{D}))^2]_{(d,g)}$ because there are *two* summands ($e^{(6+7)t}$ and $e^{(5+4)t}$) in Equation (3). In general case, we observe that the number of summands for *arbitrary* element (u,v) of $(\exp(t\mathbf{D}))^k$ in Equation (3) should be consistent with (a) the choice of N in Equation (6) and (b) the number of the paths with d hops between vertices u and v (that is, $[\mathbf{A}^d]_{(u,v)}$).

Example 2. *Consider the water network in Figure 1. To compute the sum of the pipe section lengths over all paths with* $d = 3$ *hops between vertices* b *and* i*, we first obtain its distance matrix* $\mathbf{D}$ *and adjacency matrix* $\mathbf{A}$*, as illustrated in Example 1. Next, we evaluate*

$$[\mathbf{A}^3]_{(b,i)} = 3 \quad \textit{and} \quad [(\exp(t\mathbf{D}))^3]_{(b,i)} = \underbrace{e^{(6+2+6)t}}_{b\to d\to e\to i} + \underbrace{e^{(6+4+4)t}}_{b\to d\to h\to i} + \underbrace{e^{(7+5+4)t}}_{b\to g\to h\to i}$$

Finally, choosing $N = 3$ in Theorem 1, we can "inverse" $[(\exp(t\mathbf{D}))^3]_{(b,i)}$ as follows:

$$\begin{aligned}\lim_{t\to 0}\frac{3}{t}\log\left(\frac{1}{3}\left([(\exp(t\mathbf{D}))^3]_{(b,i)}\right)\right) &= \lim_{t\to 0}\frac{3}{t}\log\left(\frac{e^{(6+2+6)t}+e^{(6+4+4)t}+e^{(7+5+4)t}}{3}\right)\\ &= (6+2+6)+(6+4+4)+(7+5+4) = 44.\end{aligned}$$

Hence, the sum of the pipe section lengths over all paths with 3 hops between vertices b and i is 44. □

After the nominator of Equation (1) is obtained, the "average distance" $[\mathbf{W}^{(d)}]_{(u,v)}$ follows directly:

Theorem 2. *The "average distance" of all paths with d hops between every two vertices u and v, $[\mathbf{W}^{(d)}]_{(u,v)}$, can be quantified as*

$$[\mathbf{W}^{(d)}]_{(u,v)} = \begin{cases} \lim_{t\to 0}\frac{1}{t}\log\left(\frac{[(\exp(t\mathbf{D}))^d]_{(u,v)}}{[\mathbf{A}^d]_{(u,v)}}\right), & \textit{if } [\mathbf{A}^d]_{(u,v)} \neq 0;\\ 0, & \textit{if } [\mathbf{A}^d]_{(u,v)} = 0;\end{cases}$$

As a special case, $\mathbf{W}^{(1)} = \mathbf{D}$. This is because, when $d = 1$ and $u \neq v$, $[\mathbf{A}^d]_{(u,v)} = 1$. Then,

$$[\mathbf{W}^{(1)}]_{(u,v)} = \lim_{t\to 0}\frac{\log\left([\exp(t\mathbf{D})]_{(u,v)}\right)}{t} = \lim_{t\to 0}\frac{[(t\mathbf{D})]_{(u,v)}}{t} = \mathbf{D}_{(u,v)} \quad \text{if } u \neq v.$$

Example 3. *Recall the result in Example 2. Since $[\mathbf{A}^3]_{(b,i)} = 3$ and the sum of the pipe section lengths over all paths with $d = 3$ hops between vertices b and i is 44, the "average distance" is*

$$[\mathbf{W}^{(3)}]_{(b,i)} = 44/3. \quad \square$$

Theorem 2 provides an efficient way of evaluating the "average distance" $[\mathbf{W}^{(d)}]_{(u,v)}$ with the *fixed* number d of hops by using distance matrix $\mathbf{D}$ and adjacency matrix $\mathbf{A}$. Based on $[\mathbf{W}^{(d)}]_{(u,v)}$, we can obtain the "average distance" matrix $\mathbf{S}^{(L)}$ *within* L hops as follows.

Definition 2. *Let $0 < \lambda < 1$ be a user-controlled decay factor. Given a water network G, its "average distance" matrix $\mathbf{S}^{(L)}$ within L hops $(L = 2, 3, \cdots)$ is defined by*

$$[\mathbf{S}^{(L)}]_{(u,v)} = \begin{cases} \frac{1}{\beta}\left[\lambda\mathbf{D} + \lambda^2\mathbf{W}^{(2)} + \cdots + \lambda^L\mathbf{W}^{(L)}\right]_{(u,v)}, & (u \neq v);\\ 0, & (u = v).\end{cases} \quad \textit{with}\ \beta = \sum_{i=1}^{L}\lambda^i \cdot \mathbf{1}_{\{[\mathbf{W}^{(i)}]_{(u,v)}\neq 0\}} \tag{7}$$

where $\mathbf{1}_{\{[\mathbf{W}^{(i)}]_{(u,v)}\neq 0\}}$ is an indicator function, which returns 1 if $[\mathbf{W}^{(i)}]_{(u,v)} \neq 0$, and 0 otherwise.

Intuitively, $[\mathbf{S}^{(L)}]_{u,v}$ captures the weighted average distance within L hops between vertices u and v. In Equation (7), the first term $\lambda\mathbf{D}$ signifies that the paths of 1 hop have a contribution of λ to $\mathbf{S}^{(L)}$; the second term $\lambda^2\mathbf{W}^{(2)}$ means that the paths of (longer) 2 hops have a (smaller) contribution of λ^2 to $\mathbf{S}^{(L)}$, and so forth. The parameter $\frac{1}{\beta}$ is a normalization factor, which guarantees that the sum of all the weighted factors $\{\lambda, \lambda^2, \cdots, \lambda^L\}$ in Equation (7) is 1.

The constant λ is between 0 and 1, which can be thought of as a confidence level. Empirically, it is often set to 0.6–0.9, which gives the rate of decay as wave spreads across the pipe sections.

Example 4. *Recall the water network in Figure 1 and its distance matrix* $\mathbf{D}$ *and adjacency matrix* $\mathbf{A}$ *in Example 1. We choose* $\lambda = 0.85$ *and* $L = 5$*. By Definition 2, the "average distance" matrix* $\mathbf{S}^{(5)}$ *can be obtained as follows:*

$$\mathbf{S}^{(5)} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \end{array} \!\!\begin{array}{c} \begin{array}{cccccccccc} \;\;\;a\;\;\; & \;\;\;b\;\;\; & \;\;\;c\;\;\; & \;\;\;d\;\;\; & \;\;\;e\;\;\; & \;\;\;f\;\;\; & \;\;\;g\;\;\; & \;\;\;h\;\;\; & \;\;\;i\;\;\; & \;\;\;j\;\;\; \end{array} \\ \begin{bmatrix} 0 & 17.9024 & 20.6098 & 14.9965 & 14.1626 & 19.0000 & 22.7073 & 15.7290 & 19.9024 & 18.7290 \\ 17.9024 & 0 & 9.6673 & 12.9995 & 12.2146 & 8.5122 & 13.9983 & 14.5122 & 18.4246 & 17.5122 \\ 20.6098 & 9.6673 & 0 & 12.6098 & 14.9734 & 6.1593 & 13.5122 & 17.3068 & 17.6667 & 17.0000 \\ 14.9965 & 12.9995 & 12.6098 & 0 & 9.3958 & 14.6098 & 14.7073 & 10.7929 & 11.9024 & 10.7290 \\ 14.1626 & 12.2146 & 14.9734 & 9.3958 & 0 & 13.0000 & 17.4650 & 11.4341 & 11.2714 & 14.4341 \\ 19.0000 & 8.5122 & 6.1593 & 14.6098 & 13.0000 & 0 & 15.5122 & 16.0000 & 19.6667 & 19.0000 \\ 22.7073 & 13.9983 & 13.5122 & 14.7073 & 17.4650 & 15.5122 & 0 & 11.8536 & 12.9024 & 11.7724 \\ 15.7290 & 14.5122 & 17.3068 & 10.7929 & 11.4341 & 16.0000 & 11.8536 & 0 & 10.2734 & 8.9514 \\ 19.9024 & 18.4246 & 17.6667 & 11.9024 & 11.2714 & 19.6667 & 12.9024 & 10.2734 & 0 & 10.3821 \\ 18.7290 & 17.5122 & 17.0000 & 10.7290 & 14.4341 & 19.0000 & 11.7724 & 8.9514 & 10.3821 & 0 \end{bmatrix} \end{array}$$

As opposed to the previous work [10] that considers only one single path of the shortest length, $\mathbf{S}^{(L)}$ can capture multiple paths of different length between every two sensor locations by fully exploiting the network topology information. Thus, if the "average distance" $\mathbf{S}^{(L)}$ is used to quantify the wave traveling distance from a burst location to a sensor location, water loss events can be positioned more accurately, as will be shown in the next section.

2.2. Effectively Positioning Water Loss Event

Having evaluated the "average distance" matrix $\mathbf{S}^{(L)}$, we next present an efficient algorithm to position a water loss event with higher accuracy. We assume that the sensor points of the water network are time synchronized. Our basic idea is to measure the difference in "average distance" to two sensor locations that detect the burst transient at known times. Specifically, let $\bar{\nu}$ denote the average wave speed, and let t_u and t_v be the time points when the burst transient event is detected at sensor locations u and v, respectively. Note that the time of the burst event t_x is unknown in advance, but such a burst event must occur before $\min\{t_u, t_v\}$ (earlier than either of the detected time at locations u and v). We observe that the time gap between $(t_u - t_x)$ and $(t_v - t_x)$ (which can be calculated as $|t_u - t_v|$) is mainly due to the difference in "average distance" from the burst (source) location x to both sensor locations u and v. Hence, ideally we have the following equations:

$$t_u - t_v = (t_u - t_x) - (t_v - t_x) \quad \Rightarrow \quad \bar{\nu}(t_u - t_v) = \underbrace{\bar{\nu}(t_u - t_x)}_{dist(u,x)} - \underbrace{\bar{\nu}(t_v - t_x)}_{dist(v,x)}$$

which implies that

$$\bar{\nu}(t_u - t_v) = dist(u, x) - dist(v, x) \approx [\mathbf{S}^{(L)}]_{u,x} - [\mathbf{S}^{(L)}]_{v,x} \tag{8}$$

Then, we can enumerate each sensor location in V to find out the top-k (k is often set to 3–5 in practice) best approximate solutions $\hat{X} \subseteq V$ of x to Equation (8), that is,

$$\hat{X} = \arg\,(\text{top-}k) \min_{x \in V} \{|\bar{\nu}(t_u - t_v) - ([\mathbf{S}^{(L)}]_{u,x} - [\mathbf{S}^{(L)}]_{v,x})|\} \tag{9}$$

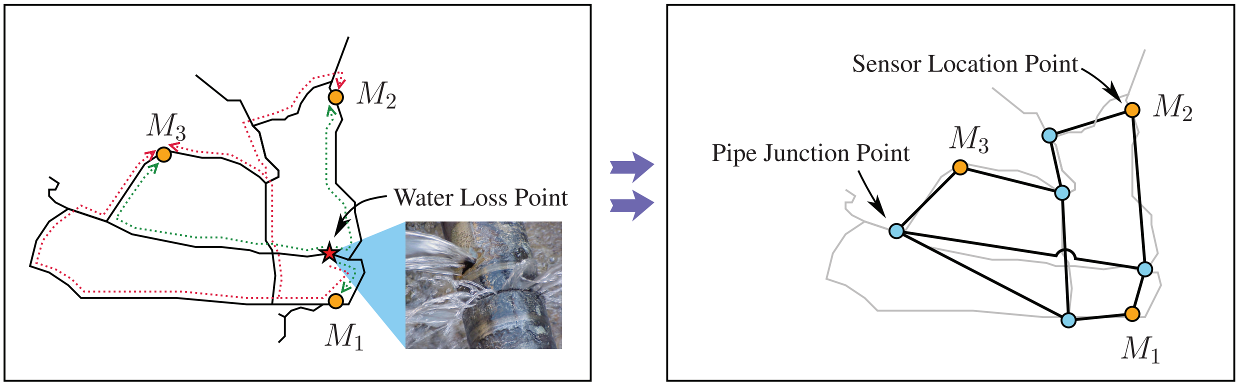

Figure 2. The real-life pipe network layout (left) and its heterogeneous graph (right), where yellow vertices represent pipe junctions, blue vertices are sensor locations. In the left figure, the green dotted lines denote the wave paths traversed by Srirangarajan *et al.*'s method [10], whereas both green and red dotted lines are those traversed by our approach.

Thus, the elements in $\hat{X}$ form a "hyperbolic curve" with two focal points u and v. To determine the precise location along this "hyperbolic curve", we need to choose another pair of sensor locations, say u and w, as two focal points, with the aim to produce the another "hyperbolic curve", that is, to find out another set of the top-k best approximate solutions $\hat{Y} \subseteq V$ to the following equation:

$$\hat{Y} = \arg\,(\text{top-}k) \min_{y \in V} \{ |\bar{\nu}(t_u - t_w) - ([\mathbf{S}^{(L)}]_{u,y} - [\mathbf{S}^{(L)}]_{w,y})| \} \tag{10}$$

The intersection of the two "hyperbolic curves" $\hat{X} \cap \hat{Y}$ will produce a small number of possible locations where a water loss event may occur. Finally, we can search locally for the most likely water loss position along pipe sections connected to the closest sensor locations in $\hat{X} \cap \hat{Y}$.

3. Experimental Study

In this section, we experimentally demonstrate the effectiveness of our water loss positioning scheme on the real test bed (WaterWiSe@SG) deployed on the water network system by Whittle *et al.* [12].

The test bed consists of sensors measuring hydraulic (pressure, flow) and water quality parameters. The pipe network layout is depicted in Figure 2 covering an area of 1km^2. It consists of $|V| = 8$ vertices ($|V_S| = 3$ pressure sensors M_1, M_2, M_3 that can detect the burst transients, and $|V_J| = 5$ pipe junctions). The measurement points are time synchronized using the GPS pulse per second (PPS) signal leading to a distance error of $\pm 2\text{m}$ [10]. To detect burst events, we also implement the CUSUM change detection test by Misiunas *et al.* [13].

The following parameters are used by default: (a) the decay factor $\lambda = 0.6$; (b) the total number of hops $L = 5$; (c) the top-k size $k = 3$.

Ten burst events are created during the evening from 21:00 to 23:00 hours. The results are reported in Table 1. For each burst event, we compute the arrival time difference for every pair of sensor locations. To estimate its burst location, we compare the localization errors of our proposed scheme with those of Srirangarajan *et al.*'s shortest distance-based method [10]. It can be discerned that, for every burst event, our method consistently exhibits 13.5%–62.7% higher accuracy than Srirangarajan *et al.*'s. The

Table 1. Results of Positioning Water Loss Events

Burst Event	Difference in Arrival Time (sec)		Water Loss Positioning Error (m)		Improved Ratio (%)
	$t_{M_2} - t_{M_1}$	$t_{M_3} - t_{M_1}$	Shortest Path Method [10]	Our Scheme	
1	0.27897	0.58687	42.07	32.49	62.70%
2	0.31484	0.52347	34.51	29.14	32.72%
3	0.32334	0.64109	43.42	33.53	56.30%
4	0.30235	0.55671	30.76	29.57	17.92%
5	0.33241	0.50157	26.11	25.33	13.50%
6	0.28782	0.57702	37.05	30.52	58.44%
7	–	0.58347	–	–	–
8	0.27647	0.47209	23.51	20.73	24.51%
9	0.32780	–	–	–	–
10	0.31478	0.52631	31.19	25.94	26.49%
Average	0.30653	0.55207	32.73	28.25	36.57%

average error of our water loss positioning method is 28.25 meters, which has 36.57% improvement over the Srirangarajan *et al.*'s approach. This is because our graph-based topology distance measure can comprehensively take into account the weighted contributions of paths with different hops between two sensor locations, whereas Srirangarajan *et al.*'s distance measure accommodates only one path of the shortest length in a biased manner. In addition, our techniques can produce the top-k ($k = 5$) best approximate solutions along a "hyperbolic curve", thus producing a better candidate set for local search.

Notice that 2 out of 10 burst events are not positioned, denoted as "–" in Events 7 and 9 of Table 1, due to the missing reading of sensors. Thus, in the above experiment, the percentage of burst events positioned by this method is $\sim$80%. Ideally, this percentage can be improved further if the sensors readings are good enough.

Currently, our algorithm is highly efficient to position burst events rather than long-term leakage, as the detection algorithm we adopted is based on a rate of sudden change criterion.

4. Conclusions

In this paper, an efficient scheme has been investigated to position water loss event more accurately by taking advantage of the water network topology. First, a novel graph topology-based measure is proposed, which can recursively quantify the "average distances" between every two senor locations simultaneously in a water network. Then, based on this measure, an efficient search algorithm is devised, which can integrate our "average distances" measure with the difference in the arrival times of the pressure variations detected at sensor locations. The viable experimental study on real-life test bed (WaterWiSe@SG) demonstrates that our proposed positioning scheme can position water loss event more reliably with an improvement of up to 62.7% accuracy over the best-known algorithm.

For future work, we aim to develop optimization techniques that can substantially accelerate the computation of our proposed scheme, aiming to position water loss events very quickly on a large-scale water supply system. Another interesting problem is to reduce its memory usage. We will incorporate some of our preliminarily results on graph analysis [14–19] into the water flow and pressure behavior, to achieve the scalability of our proposed algorithm.

Acknowledgment

The work presented in this paper is supported by the European Commission under the 7th Framework Programme Grant Agreement No.619795 and NEC Japan.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lomborg, B. *Global crises, global solutions*; Cambridge University Press, 2004.
2. Akiner, S. Turkmenistan: Strategies of Power, Dilemmas of Development. *Europe-Asia Studies* **2014**, *66*, 146–148.
3. Miller, R.J.; Low, P.F. Threshold gradient for water flow in clay systems. *Soil Science Society of America Journal* **1963**, *27*, 605–609.
4. Farmer, E. System for monitoring pipelines, 1989. US Patent 4,796,466.
5. Radder, A. On the parabolic equation method for water-wave propagation. *Journal of fluid mechanics* **1979**, *95*, 159–176.
6. Herrera, M.; Canu, S.; Karatzoglou, A.; Pérez-García, R.; Izquierdo, J. An approach to water supply clusters by semi-supervised learning. *Proceedings of International Environmental Modelling and Software Society (IEMSS)* **2010**.
7. Shaffer, J.P. Multiple hypothesis testing. *Annual review of psychology* **1995**, *46*, 561–584.
8. Geiger, G. State-of-the-art in leak detection and localization. *Oil Gas European Magazine* **2006**, *32*, 193.
9. Misiunas, D.; Lambert, M.; Simpson, A.; Olsson, G. Burst detection and location in water distribution networks. *Water Science and Technology: Water Supply* **2005**, *5*, 71–78.
10. Srirangarajan, S.; Allen, M.; Preis, A.; Iqbal, M.; Lim, H.B.; Whittle, A.J. Water main burst event detection and localization. Proc. 12th Water Distribution Systems Analysis Conference (WDSA), 2010.
11. Sniedovich, M. Dijkstra's algorithm revisited: The dynamic programming connexion. *Control and cybernetics* **2006**, *35*, 599.
12. Whittle, A.J.; Girod, L.; Preis, A.; Allen, M.; Lim, H.B.; Iqbal, M.; Srirangarajan, S.; Fu, C.; Wong, K.J.; Goldsmith, D. WATERWISE@SG: A testbed for continuous monitoring of the water distribution system in singapore. Proc. 12th Water Distribution Systems Analysis Conference (WDSA), 2010.
13. Misiūnas, D. *Failure Monitoring and Asset Condition Asssessment in Water Supply Systems*; Vilniaus Gedimino technikos universitetas, 2008.
14. Yu, W.; Lin, X. IRWR: Incremental random walk with restart. SIGIR, 2013, pp. 1017–1020.
15. Yu, W.; Lin, X.; Zhang, W. Fast incremental SimRank on link-evolving graphs. ICDE, 2014, pp. 304–315.
16. Yu, W.; Lin, X.; Zhang, W.; McCann, J.A. Fast All-Pairs SimRank Assessment on Large Graphs and Bipartite Domains. *IEEE Transactions on Knowledge and Data Engineering* **2015**, *27*, 1810–1823.
17. Yu, W.; McCann, J.A. Co-Simmate: Quick Retrieving All Pairwise Co-Simrank Scores. ACL, 2015, pp. 327–333.
18. Yu, W.; McCann, J.A. High Quality Graph-Based Similarity Search. SIGIR, 2015, pp. 83–92.
19. Yu, W.; McCann, J.A. Efficient Partial-Pairs SimRank Search for Large Networks. *PVLDB* **2015**, *8*, 569–580.