AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR028285

Efficient Leak Localization in Water Distribution Systems Using Multistage Optimal Valve Operations and Smart Demand Metering

Yuan Huang[1], Feifei Zheng[1], Zoran Kapelan[2], Dragan Savic[3,4], Huan-Feng Duan[5], and Qingzhou Zhang[1]

[1]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China, [2]Department of Water Management, Delft University of Technology, Delft, Netherlands, [3]KWR Water Research Institute, Nieuwegein, Netherlands, [4]Centre for Water Systems, University of Exeter, Exeter, UK, [5]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong

Key Points:
- A multistage method is developed to localize pipe burst leaks using valve operations
- The graph theory is employed to efficiently identify optimal valve operations
- The proposed method is demonstrated by practical application to a real DMA

Supporting Information:
- Supporting Information S1
- Data Set S1
- Data Set S2
- Data Set S3

Correspondence to:
F. Zheng,
feifeizheng@zju.edu.cn

Citation:
Huang, Y., Zheng, F., Kapelan, Z., Savic, D., Duan, H.-F., & Zhang, Q. (2020). Efficient leak localization in water distribution systems using multistage optimal valve operations and smart demand metering. *Water Resources Research*, *56*, e2020WR028285. https://doi.org/10.1029/2020WR028285

Received 7 JUL 2020
Accepted 2 OCT 2020
Accepted article online 6 OCT 2020

Abstract This paper proposes a multistage method for burst leak localization through valve operations (VOs) and smart demand metering in district meter areas (DMAs) of water distribution systems (WDSs). Each stage includes partitioning of the DMA into two subregions using VOs and identification of potentially leaking pipes within the subregions through water balance analysis based on smart demand meters. Such a process is performed repeatedly (multiple stages) to narrow down the spatial range for pinpointing leak locations. To improve efficiency, a bisection optimization problem is formulated to localize the minimum leak areas using the lowest number of VOs, which is solved by a graph theory-based method. The utility of the proposed method is demonstrated using two DMAs (DMA1 and DMA2) of a real WDS with different topological properties. Results show that the proposed method can efficiently localize artificial burst leaks in DMA1 within 7–15% of the total pipe length, implying that the proposed method is theoretically effective in localizing pipe burst leaks. The real application to DMA2 has identified two leak regions with 2.3 and 4.2 km of pipe length (around 3–6% of the entire DMA2) using 18 VOs. These two burst leaks have been subsequently confirmed and pinpointed using listening rods by practitioners of the local water utility. These results indicate that the proposed multistage method is effective and efficient for burst leak localization, which can be promising for wide practical applications due to rapid developments of smart WDSs (e.g., smart demand meters or control valves).

1. Introduction

Water is a scarce and critical resource to humankind for supporting population growth, economic development, and urbanization expansion. The Organization for Economic Cooperation and Development (OECD) has reported that the global water demand is expected to increase by 55% by 2050, resulting in that 40% of the global population can be under severe water stress (OECD, 2012). However, a substantial portion of clean or drinking water has been continuously lost from the water distribution systems (WDSs) nowadays. Specifically, according to the World Bank, worldwide water utilities lost nearly 48 billion m^3 of the treated water annually from the distribution networks, leading to about $14 billion of financial loss per year (Mutikanga et al., 2013). Therefore, it is of great importance to manage water losses of WDSs, not only for financial savings of the water utilities but also for mitigating the water crisis around the world.

Leakage is one of the main causes of water loss in WDSs, with the leak volumes varying widely across different systems, regions, and countries (Puust et al., 2010). For instance, water loss due to leakage can be as low as 3–7% of the total supplied water in some well-maintained systems, such as the Netherlands; yet it can be also as high as 50% or more in some developing countries such as many cities/towns in India and China (China Urban Water Association, 2015). In addition, leaks do not just cause waste of natural resources but also increase unnecessary energy cost and even induce potential health threats (e.g., contaminant or pathogen intrusion from leaks under low-pressure conditions, Besner et al., 2011). Consequently, leak localization has been one of the prioritized and critical tasks for water utilities to improve the operation and management of WDSs. To address this problem,

intensive studies have been undertaken over the past few decades, and hence a wide spectrum of methods and techniques have been developed to localize leaks in WDSs (Gupta & Kulat, 2018).

Real water losses in WDSs are typically classified as background and burst leaks (Puust et al., 2010). Background leaks are distributed, often relatively small, and cannot be detected; therefore, the leak localization methods developed so far mainly focused on burst leaks. These methods can be overall classified into two categories: equipment-based and model-based approaches (Qi et al., 2018). The equipment-based approaches are progressively developed from the traditional listening rods since the mid-1980s to the newly emerged equipment such as leak noise loggers, tethered systems, ground penetrating radars, and infrared thermography in recent years (Gupta & Kulat, 2018; Mutikanga et al., 2013). These equipment-based techniques have greatly improved the effectiveness of leak localization in WDSs. However, their practical applications are still challenged mainly due to the low efficiency (e.g., labor-intensive), high cost, and the need of great professional knowledge/experience within their implementations (Farah & Shahrour, 2017). As a result, the equipment-based approaches are often used to identify the leaks within a small local region of the WDSs (Rajeswaran et al., 2018).

The model-based approaches for leak localization are generally performed with the aid of WDS hydraulic models as well as the data analysis methods or optimization techniques (Zheng et al., 2016, 2017). These approaches can be further categorized into transient-based methods, optimization-based methods, and data-driven methods. The transient-based methods often analyze the transients induced by leaks using signal process techniques (e.g., Colombo et al., 2009; Wang et al., 2020); the optimization-based methods typically define the leak detection as a demand calibration problem (e.g., Sanz et al., 2015; Sophocleous et al., 2019; Zhang et al., 2016); the data-driven methods analyze the data from monitoring sites (e.g., pressure sensors and flow meters) to identify leaks (e.g., Romano et al., 2014; Zhou et al., 2019). While the model-based approaches are efficient and cost-effective in identifying leaks (Duan et al., 2011), they often exhibit equifinality and low-reliability issues in practical applications, especially for highly looped and complex networks (Sophocleous et al., 2019).

Over the past few decades, there has been a move toward the use of district meter areas (DMAs) to improve the efficiency of the leakage management of the complex WDS (Mutikanga et al., 2013). Such leak management is typically carried out by analyzing the monitored flows at inlets and outlets of the DMA during the nighttime (night flow analysis, Covas et al., 2006). While this DMA-based approach can be effective in determining whether leaks exist in a DMA (regional localization), it is still difficult to identify the exact localization of leaks within the DMA due to that many DMAs can be spatially large, especially in big cities. For example, DMAs can consist of 1,000–2,000 properties in the United Kingdom and even larger users in some cities of China (Qi et al., 2018).

To address the potential challenges associated with the existing leak localization methods as mentioned above, the present paper proposes a new multistage method for leak localization within the DMA with the aid of active valve operations (VOs) and smart demand metering. The proposed method is made possible due to that many water utilities have increasingly utilized smart and intelligent techniques in their WDSs (referred as smart WDSs), such as smart demand meters, remote control valves, and advanced informatics (Creaco et al., 2019; Nguyen et al., 2018). Such smart facilities provide opportunities for the development of innovative techniques to enable enhanced WDS operations and management (Farah & Shahrour, 2017). For instance, smart demand meters have been increasingly used by many water utilities to provide near-real-time demand data at end users, thus offering opportunities for accurate hydraulic modeling of the WDS (Creaco et al., 2016; Nguyen et al., 2018) and efficient leakage management (Bragalli et al., 2019; Nguyen et al., 2018). To the best of our knowledge, the present study is the first work that the smart demand meters within the WDS combined with the VOs are employed to efficiently and effectively localize the leak regions of a DMA.

Within the proposed method, a multistage bisection optimization approach is developed to locate burst leaks within the DMA, where VOs and water balance analysis (WBA) based on smart demand meters are iteratively performed to gradually narrow down the spatial regions associated with leaks. The proposed method requires a hydraulic model to evaluate the hydraulics status of the DMA after VOs. To improve the efficiency of the method, a graph theory-based approach is developed to identify optimal VO strategies that can enable the partitioning of the DMA in an efficient manner. The main contributions/novelties of the present work

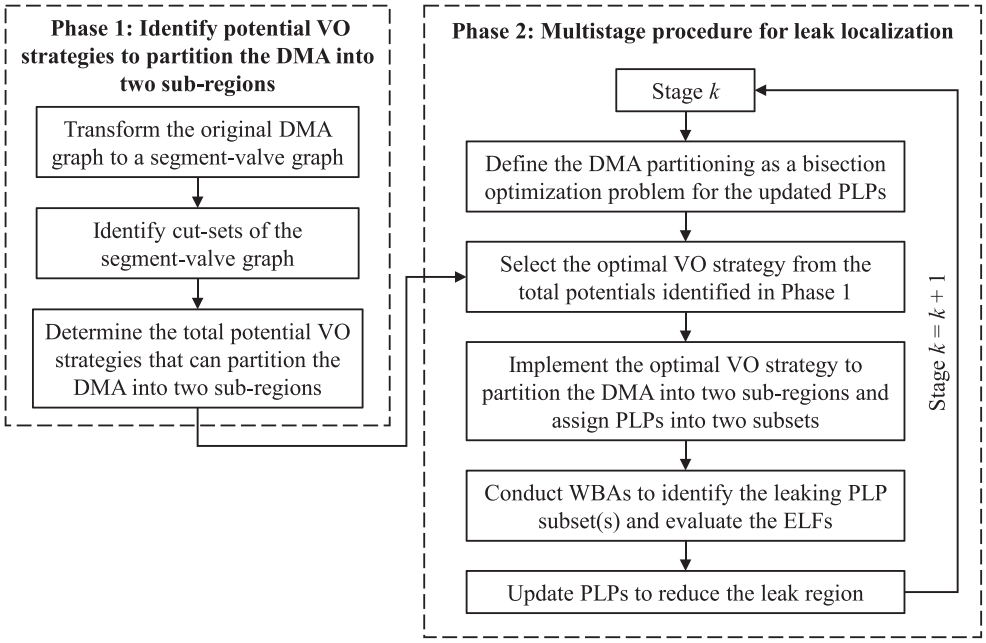


Figure 1. Overview of the methodology framework.

include (i) the proposed method can be significantly more efficient than the traditional equipment-based approaches in leak localization, and (ii) the proposed method can identify the leak region more reliable with a significantly finer spatial resolution than traditional model-based approaches. The details of the proposed multistage method are presented below, and two DMAs in a real WDS are used to demonstrate its feasibility and application procedures.

2. Methodology

2.1. Overview

The proposed method utilizes a strategy of two operations in sequence—the VOs and the WBA, conducted in a multistage manner, to localize the leaking pipes/regions in the DMA. More specifically, assuming that initially all pipes in a DMA are the potentially leaking pipes (PLPs), the method uses VOs (i.e., closing valves) to partition the DMA into two subregions. This is followed by conducting the WBA to identify whether the PLPs within the two subregions have leaks. This VO + WBA strategy is repeated stage by stage to narrow down the search for the spatial regions that have PLPs. This strategy is carried out until no further VOs are available for the identification, and the finally identified region represents the minimum spatial area associated with leaks in the DMA.

Following the above rationale, the proposed methodology can be divided into two phases, which are Phase 1 and Phase 2 as shown in Figure 1. Phase 1 identifies potential VO strategies that can partition the DMA into two subregions, which is carried out only once in the proposed methodology. Phase 2 is designed to enable leak localization using a multistage manner, that is, it is carried out repeatedly to identify leaks. In Phase 1, a graph theory-based approach is employed to identify VO strategies for DMA partitioning, involving the transformation of the original DMA graph to a segment-valve graph, and the identification of cut sets of the segment-valve graph as the VO strategies.

In Phase 2 (Figure 1), the leaks are localized using the following multistage procedures. Considering a particular stage k, the optimal VO strategy is firstly determined to partition the DMA into two subregions. This is attained by formulating the DMA partitioning as a bisection optimization problem that the PLPs in the DMA is bisected (by length) into two PLP subsets within two subregions using the minimum number of VOs. The rationale behind the selection of the bisection approach is that it is statistically efficient in identifying leaking pipes conditioned on the fact that no prior knowledge of the leak locations is available, which is typically the

case in engineering practice (Ferrari et al., 2014). Then the optimal VO strategy is selected from the VO strategies identified in Phase 1 (i.e., Phase 1 has approximately enumerated all potential VO strategies) based on the defined bisection optimization problem. Subsequently, the selected optimal VO strategy is implemented to partition the DMA into two subregions as well as assign the PLPs into two subsets. This is followed by conducting WBAs for the two subregions to identify the leaking PLP subset(s) as well as to evaluate the estimated leak flowrate(s) (ELFs). Based on the WBA results, the PLPs are updated by removing pipes with no leaks, thereby reducing the leak region of the DMA. The leak localization procedure moves to the next stage (i.e., $k = k + 1$), and these procedures are repeated in a multistage manner until no VOs are available.

It is possible that at a particular stage the WBA can identify that both the two subregions contain leaks, implying the existence of multiple leaks in the DMA. For such a case, the proposed method concentrates on the identification of the PLP subset with a larger ELF (i.e., localizing the leak with a larger flowrate) and puts aside the other PLP subset as candidate PLPs (referred to as CPLPs) temporarily. When the PLP subset with a larger ELF has been finally identified, the proposed method is further implemented to localize other leaks by setting the CPLPs as PLPs until all leaks in the DMA have been identified (the details are given in sections 2.3.3 and 2.4).

It should be acknowledged that VOs may disturb or even interrupt the (part of) water supply in the analyzed DMA, hence the proposed method is typically implemented during the minimum night flow (MNF) period (e.g., between 02:00 and 05:00, Mutikanga et al., 2013). This is practically feasible as the domestic demands during this period are generally negligible, and hence the interruption of water supply for domestic users within a short time period is considered acceptable. However, the demands at commercial nodes or large users can still be guaranteed during the application of the proposed method by adding pressure constraints at these nodes.

To enable effective leak localizations using the proposed method, the WBAs at each stage should be conducted in (near) real time. This requires the use of smart demand meters for all users in the DMA providing online meter information, at least at the same time resolution as the flow meters at inlets/outlets of the DMA (e.g., every 5 min). Such a large number of smart meters may not be available for most WDSs, but accurately measuring the user demands at a high temporal resolution is an important future trend (Bragalli et al., 2019). Actually, due to the rapid development of the Internet of Things technology in recent years, increasing installation of smart demand meters has been observed in WDSs around the world (Zheng et al., 2018). These real-time demand data have already been used for various engineering purposes, such as user demand pulse analysis and real-time leak analysis (Bragalli et al., 2019; Creaco et al., 2016). This is especially the case in China as reducing the leakage rate below 10% by 2020 is the objective explicitly requested by the central government. This has become an official target for all water utilities, as stated in the "Action Plan for Prevention and Control of Water Pollution" issued in 2015, thereby motivating the wide use of smart demand meters in water utilities. Therefore, it is anticipated that the proposed method can be practically promising in the near future. The details of the proposed methodology (Phases 1 and 2) are given below.

2.2. Phase 1: Identify VO Strategies to Partition the DMA Into Two Subregions

Consider the DMA as a connected graph $G(V, E)$, where the nodes and links represent vertices (V) and edges (E), respectively. The valves in the DMA can be considered as vertices or edges or in the graph depending on how they are formed in the network models. For instance, the valves are taken as links in the EPANET2.0 that has been widely used for WDS modeling (Rossman, 2000), thus they can be considered as edges in the graph. A VO strategy that partitions the DMA into two subregions can be regarded as a cut set of the graph in which the cut edges/vertices correspond to the closed valves. Therefore, only the edges/vertices corresponding to valves, rather than other types of edges/vertices, are considered to enable DMA partitioning in the proposed method. To attain this goal, the proposed method transforms the original DMA graph to a segment-valve graph in which only the valves are considered as edges, with details given in the section below. As a result, the cut sets of the segment-valve graph can be identified as the VO strategies for DMA partitioning.

2.2.1. Transformation of the Original DMA Graph to a Segment-Valve Graph

The concept of a segment is usually used to represent the minimum region of a WDS that can be isolated by closing valves (Walski, 1993). While there are many algorithms available to identify segments in WDSs

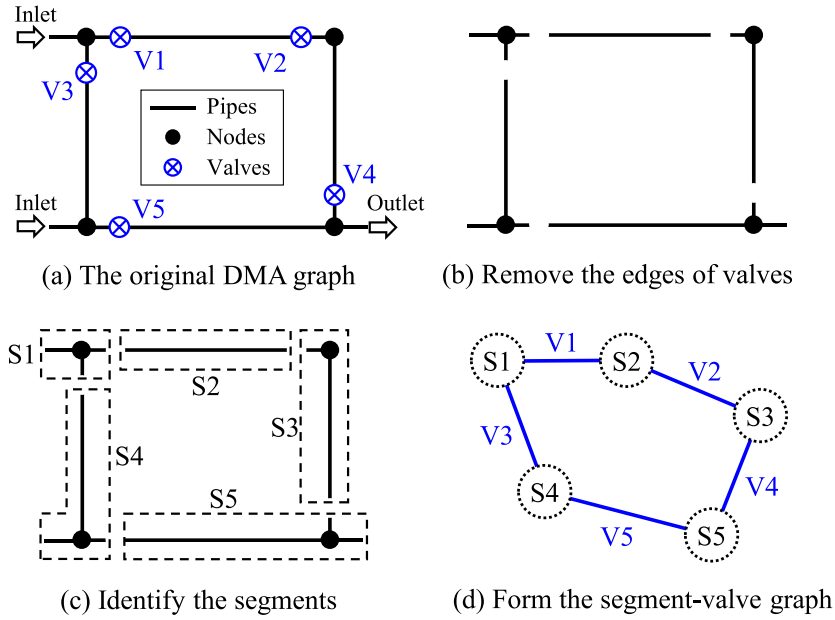


Figure 2. Illustration for the transformation of the original DMA graph to a segment-valve graph.

(Giustolisi & Savic, 2010), a graph theory-based approach is used in the proposed method due to its simplicity. The original DMA graph is transformed into a segment-valve graph in three steps, which are (i) the removal of all the valve edges in the original DMA graph $G(V, E)$ in order to generate a new graph G_{new} (V_{new}, E_{new}), (ii) the identification of all connected components of the graph G_{new} (V_{new}, E_{new}) using the depth-first search method, where each connected component corresponds to a segment of the DMA, and (iii) the formation of the segment-valve graph G_s (V_s, E_s) by taking the identified segments as vertices (V_s) and the connected valves as edges (E_s). For clarification, an illustrative example is presented in Figure 2. As shown in the figure, the original graph of the hypothetical DMA with five valves (Figure 2a) is transformed into a segment-valve graph $G_s = (V_s, E_s)$, where $V_s = \{S1, S2, S3, S4, S5\}$ in Figure 2c is the set of vertices and $E_s = \{V1, V2, V3, V4, V5\}$ in Figure 2d is the set of edges.

2.2.2. Identification of Cut Sets of the Segment-Valve Graph

As only valves are considered as edges in the segment-valve graph (Figure 2), the edges of a cut set represent the corresponding valves that divide the DMA into two subregions when they are closed. Therefore, it can be deduced that different cut sets of the segment-valve graph represent different VO strategies (closed valves) for DMA partitioning. In other words, at each stage, these cut sets can be considered as the possible solutions to the bisection optimization problem as illustrated in Figure 1. Motivated by this, an approach is proposed to enumerate all cut sets of the segment-valve graph as feasible VO strategies for DMA partitioning.

However, based on the graph theory, there are theoretically $2^{|V_s|-1} - 1$ cut sets for the segment-valve graph G_s (V_s, E_s) (Karger, 1993). Therefore, it can be computationally expensive, or even intractable, to find all the cut sets for large graphs (i.e., DMAs with a large number of valves). For example, for an undirected graph with 30 vertices, there are about 537 million cut sets, and the computational efforts for identifying such a massive number of cut sets can significantly go beyond the time budgets that are typically available. To address this challenge, the proposed method adopts Karger's (Karger, 1993) algorithm to identify a sufficient number of VO strategies for DMA partitioning, which are accordingly used to enable the identification of near-optimal VO strategies for the bisection optimization problem. The pseudocode of Karger's algorithm used for the cut set identifications is given in Figure 3. These identified VOs strategies (i.e., cut sets) will be used as potential solutions for the bisection optimization problem described in Phase 2 as shown below.

2.3. Phase 2: Multistage Procedure for Leak Localization

2.3.1. Bisection Optimization for DMA Partitioning

In Phase 2, the VO + WBA strategy is repeated to update the PLPs by removing nonleaking pipes (NLPs) to gradually reduce the leak region in the DMA. Specifically, at each stage of the proposed method, the optimal

Algorithm 1: Cut-set identifications for the segment-valve graph

Input: the segment-valve graph $G_s = (V_s, E_s)$
Output: the cut-sets *Cuts*
n ← 1
Repeat
  $G_{new} = (V_{new}, E_{new}) \leftarrow G_s$
  while $|V_{new}| > 2$ **do**
    randomly select $e \in E_{new}$
    merge the two vertices of e
    remove e from E_{new}
  end while //the contraction procedure
  $Cuts(n) \leftarrow$ the edges in G_{new}
  $n \leftarrow n + 1$
until $n > |V_s|^4 \log|V_s|$ //repeat the contraction procedure
Remove duplicate cut-sets in *Cuts*
return *Cuts*

Figure 3. Cut set identifications for the segment-valve graph.

VO strategy used to partition the DMA into two stages is determined by solving the bisection optimization problem as follows:

$$\text{Minimize: } F^k = w\frac{\sum_{n=1}^{N}\left|v_n^k - v_n^{k-1}\right|}{N} + (1-w)\frac{\left|\sum_{m=1,\, m \in \text{PLP}_1^k}^{M} L_m - \sum_{m=1,\, m \in \text{PLP}_2^k}^{M} L_m\right|}{\sum_{m=1}^{M} L_m} \tag{1}$$

Subject to

$$\text{Valve operations: } \mathbf{V}^k \in \{0, 1\}^N \tag{2}$$

$$\text{DMA partitioning: } G - G_1\left(\mathbf{V}^k \cup G_2\left(\mathbf{V}^k\right)\right) \tag{3}$$

$$\text{Assigning PLPs into two subsets: } \mathbf{PLP}^k = \mathbf{PLP}_1^k \cup \mathbf{PLP}_2^k \tag{4}$$

$$\text{Assigning CPLPs to a sub-region: } \mathbf{CPLP} \subseteq G_1\left(\mathbf{V}^k\right) \text{ or } \mathbf{CPLP} \subseteq G_2\left(\mathbf{V}^k\right) \tag{5}$$

$$\text{Pressure constraints at uninterruptible users: } \mathbf{H}_u\left(\mathbf{V}^k\right) \geq \mathbf{h}_u^{\min} \tag{6}$$

where the superscript k represents the kth stage of the multistage procedure and F^k is the objective function value; $v_n^k \in \{0, 1\}$ indicates the status of each valve at the kth stage with $v = 0$ and 1 representing a closed or open valve, respectively; $n = 1, 2, ..., N$ is the nth valve and N is the total number of valves that can be operated in the DMA; $\mathbf{V}^k = \left\{v_1^k, ..., v_n^k\right\}$ is the set of valve status at the kth stage; L_m is the length of pipe m and M is the total number of pipes in the DMA. $G_1(\mathbf{V}^k)$ and $G_2(\mathbf{V}^k)$ in Equation 3 represents the partitioned subgraphs (i.e., subregions in the DMA) from the original graph G by the VOs $\mathbf{V}^k$. **PLP** and **CPLP**, respectively represent the set of PLPs and CPLPs as previously defined. Equation 4 indicates the result of DMA partitioning is that the PLPs ($\mathbf{PLP}^k$) are assigned into two subsets $\mathbf{PLP}_1^k \subseteq G_1\left(\mathbf{V}^k\right)$ and $\mathbf{PLP}_2^k \subseteq G_2\left(\mathbf{V}^k\right)$, respectively.

The first term on the right side of Equation 1 (i.e., $\sum_{n=1}^{N}\left|v_n^k - v_n^{k-1}\right|$) indicates the number of VOs at the kth stage relative to its previous stage. The second term on the right side of Equation 1 (i.e., $\left|\sum_{m=1,\, m \in \text{PLP}_1^k}^{M} L_m - \sum_{m=1,\, m \in \text{PLP}_2^k}^{M} L_m\right|$) indicates the length difference between the PLPs in the two subsets partitioned by the VOs at the kth stage. The two terms are normalized on the range of 0–1 by dividing the total number of valves (N) and the total length of pipes ($\sum_{m=1}^{M} L_m$), respectively. Consequently, the

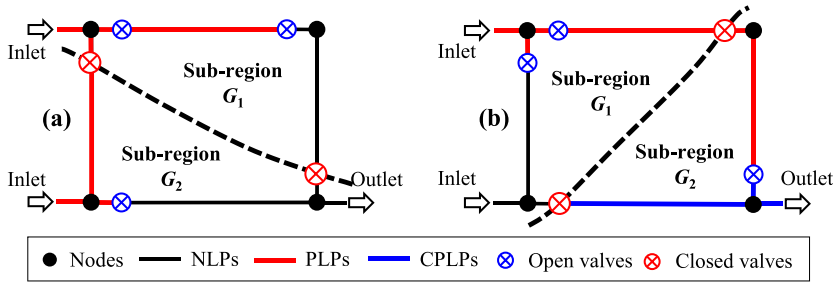


Figure 4. Illustration of different cases for DMA partitioning including (a) both subregions have inflows and (b) one subregion has inflows and the other does not.

minimization of Equation 1 can be physically interpreted as using the minimum number of VOs to assign the PLPs into two subsets with the minimum difference in the total length of pipes at each stage, thus achieving a statistically efficient leak localization as previously stated. The w in Equation 1 is the weight, indicating the relative importance of the two terms. In this study, $w = 0.5$ is used, representing equal importance for both terms in Equation 1, but the impacts of different w values on the optimization results are investigated later in the case studies.

As previously described, the CPLPs are a set of pipes temporarily removed from the **PLP** set in the DMA when both PLP subsets have leaks identified by the WBA. Equation 5 defines a constraint that ensures that all pipes in the CPLPs should be always in a single subregion. In other words, these pipes are not allowed to be separated before they are reset as PLPs for leak identification. This is because the ELF associated with the CPLPs have to be used for WBA analysis within the proposed method. The details of the CPLPs are given in section 2.3.3.

Equation 6 represents the pressure constraints for the users that require continuous water supplies even during the night flow period (i.e., sensitive or critical users) in the WDS. Therefore, the pressures at these users ($\mathbf{H}_u(\mathbf{V}^k)$) should be no lower than minimum allowable values ($\mathbf{h}_u^{\min}$) after the VOs ($\mathbf{V}^k$). To obtain the values of $\mathbf{H}_u(\mathbf{V}^k)$, the EPANET2.0 software (Rossman, 2000) is used in this study, where the simulations are carried out within the entire operation duration with a relatively high time resolution (e.g., every 5 min). The identified VO strategies in Phase 1 are repeatedly evaluated using the objective function in Equation 1 and the constraints in Equations 2–6, in order to identify the VO strategy with the lowest objective function value and the constraints satisfied at each stage.

2.3.2. WBA for Subregions

At each stage of Phase 2, the optimal VO strategies determined from Equations 1–6 are implemented to partition the DMA into two subregions (i.e., $G = G_1 \cup G_2$) as well as dividing the entire **PLP** set into two subsets (i.e., $\mathbf{PLP} = \mathbf{PLP}_1 \cup \mathbf{PLP}_2$). This is followed by the use of the WBA to identify which subset(s) (i.e., $\mathbf{PLP}_1$ or/and $\mathbf{PLP}_2$) has/have leaks, thereby updating the **PLP** by removing the pipes with no leaks. Generally, the WBA is conducted to estimate the nonrevenue water (NRW) for a given region (e.g., a DMA) based on the records of inflows, outflows, volumes changes in the storage tanks if available, as well as the measured water demands of the analyzed region (Mutikanga et al., 2013). Intuitively, a region is likely to have leaks if the estimated NRW value during the MNF period is high, if say, exceeding a predetermined threshold based on the MNF analysis (Covas et al., 2006).

Two assumptions are made in the WBA conducted in the proposed method. These are (i) the NRW is dominated by pipe burst leaks (i.e., the leaks considered in this paper), instead of background leaks or meter inaccuracies as leaking rates caused by the latter two are typically much lower (Covas et al., 2006), and (ii) pressure fluctuations resulting from VOs during the MNF period are insignificant in affecting the leak flowrates (Wright et al., 2015). Based on these two reasonable assumptions, the NRW obtained from the WBA can be approximately considered as the ELF, that is, ELF $\approx$ NRW. Accordingly, ELF(G), ELF(G_1), ELF(G_2), ELF ($\mathbf{PLP}_1$), ELF ($\mathbf{PLP}_2$) and ELF (**CPLP**) are defined as the ELFs of the entire DMA, the subregion G_1, the subregion G_2, the subset $\mathbf{PLP}_1$, the subset $\mathbf{PLP}_2$ and the set **CPLP**, respectively. These values are obtained from the NRW analysis by applying the WBAs to the DMA and the subregions. In this study, a threshold value ELF_{tol} is also defined to identify the existence of burst leaks in the presence of

Algorithm 2: Evaluation of ELF($\mathbf{PLP}_1$), ELF($\mathbf{PLP}_2$)

Input: G, G_1, G_2, **PLP**, $\mathbf{PLP}_1$, $\mathbf{PLP}_2$, **CPLP**, ELF(**CPLP**)
Output: ELF($\mathbf{PLP}_1$), ELF($\mathbf{PLP}_2$)

conduct WBA for G to obtain ELF(G)
if G_1 and G_2 have inflows //the case in Figure 3(a)
  conduct WBAs for G_1 and G_2 to obtain ELF(G_1) and ELF(G_2)
else if G_1 has inflows and G_2 is isolated //the case in Figure 3(b)
  conduct WBA for G_1 to obtain ELF(G_1)
  ELF(G_2) = ELF(G) – ELF(G_1)
end if
if $\mathbf{CPLP} = \emptyset$ //the case in Figure 3(a)
  ELF($\mathbf{PLP}_1$) = ELF(G_1)
  ELF($\mathbf{PLP}_2$) = ELF(G_2)
else if $\mathbf{CPLP} \subseteq G_1$
  ELF($\mathbf{PLP}_1$) = ELF(G_1) – ELF(**CPLP**)
  ELF($\mathbf{PLP}_2$) = ELF(G_2)
else if $\mathbf{CPLP} \subseteq G_2$ //the case in Figure 3(b)
  ELF($\mathbf{PLP}_1$) = ELF(G_1)
  ELF($\mathbf{PLP}_2$) = ELF(G_2) – ELF(**CPLP**)
end if
return ELF($\mathbf{PLP}_1$), ELF($\mathbf{PLP}_2$)

Figure 5. Evaluation of ELF ($\mathbf{PLP}_1$), ELF ($\mathbf{PLP}_2$).

background leaks and meter inaccuracies. More specifically, if ELF ($\mathbf{PLP}_i$) > ELF_{tol} ($i \in \{1, 2\}$), it indicates that the PLPs in $\mathbf{PLP}_i$ have leaks and vice versa. Note that the defined ELF_{tol} in the proposed method is similar to the threshold for MNF analysis in DMAs that accounts for background leaks and meter inaccuracies (Alkasseh et al., 2013; Farah & Shahrour, 2017).

The aim of the WBA is to estimate the ELFs of the two PLP subsets, that is, to identify whether leaks exist in the subsets. To achieve this, the following steps are implemented:

i. Estimate the ELF(G) using the WBA applied to the entire DMA before the leak localization.
ii. Evaluate ELF(G_1) and ELF(G_2). When evaluating ELF(G_1) or ELF(G_2), two different cases (see Figure 4) may exist, which are (a) both subregions have inflows, that is, both subregions contain at least one inlet of the DMA as shown in Figure 4a, and (b) one subregion has inflows and the other one does not have inflows (i.e., one subregion contains all the inlets of the DMA while the other is isolated by closing valves, as shown in Figure 4b). In Case (a), the WBAs are applied to two subregions to determine the ELFs, that is, ELF(G_1) and ELF(G_2) in Figure 4a. In Case (b), the WBA is applied to the subregion having inflows to compute the ELF (e.g., ELF(G_1) in Figure 4b). The ELF of the other subregion (e.g., ELF(G_2) in Figure 4b) is estimated as the difference between the original ELF of the entire DMA and the ELF of the subregion having inflows;
iii. Estimate the ELF ($\mathbf{PLP}_1$) and ELF ($\mathbf{PLP}_2$). This is carried out by using the estimated ELF(G_1) and ELF(G_2) and two possible cases exist. These are (a) if the analyzed subregion has no CPLPs (e.g., the subregion G_1 in Figure 4b), then ELF ($\mathbf{PLP}_i$) = ELF(G_i), $i \in \{1,2\}$; and (b) if the analyzed subregion has CPLPs (e.g., the subregion G_2 in Figure 4b), the ELF of the PLP subset in that subregion can be calculated as ELF ($\mathbf{PLP}_i$) = ELF(G_i)—ELF (**CPLP**), $i \in \{1, 2\}$. The pseudocode for the above procedures is given in Figure 5.

2.3.3. The Update for PLPs

By conducting the WBAs at each stage, ELF ($\mathbf{PLP}_1$) and ELF ($\mathbf{PLP}_2$) for the two PLP subsets are obtained to identify whether they have leaks, thereby reducing the size of the set of PLPs by removing NLPs. As

Algorithm 3: Update for the PLPs

Input: $\mathbf{PLP}_1$, $\mathbf{PLP}_2$, **CPLP**, **NLP**, ELF($\mathbf{PLP}_1$), ELF($\mathbf{PLP}_2$), ELF(**CPLP**), ELF_{tol}
Output: **PLP**, **NLP**, **CPLP**, **ELF(CPLP)**

```
if ELF(PLP_1) > ELF_tol and ELF(PLP_2) ≤ ELF_tol    //case (a)
    PLP = PLP_1
    NLP = NLP ∪ PLP_2
else if ELF_tol < ELF(PLP_2) < ELF(PLP_1)           //case(b), multiple leaks
    PLP = PLP_1
    CPLP = CPLP ∪ PLP_2
    ELF(CPLP)+= ELF(PLP_2)
end if
return PLP, NLP, CPLP, ELF(CPLP)
```

Figure 6. Update for the PLPs.

previously stated, two different cases may exist within the WBA: (a) only one PLP subset in a subregion is identified to have leaks, and (b) both PLP subsets in the two subregions are identified to have leaks, implying the existence of multiple leaks in the DMA. For case (a), the PLP subset having leaks are moved to the next stage for further partitioning, and the other PLP subset with no leaks are removed as so to reduce the spatial region of the leaking pipes. For case (b), the proposed method selects the PLP subset with a larger ELF for further identification and considers the PLP subset with lower ELP as CPLPs. These pipes are added to the set **CPLP** temporarily, and the ELF (**CPLP**) is accordingly updated, with details represented by the pseudocode in Figure 6.

2.4. Overall Procedures of the Proposed Multistage Method

The steps of the proposed multistage method to localize leaks in a DMA are shown in Figure 7, with details of each step given below.

Step 1: Initialize the method for leak localization including the consideration of all pipes in the DMA as PLPs (i.e., the set **PLP**), the set **NLP** = Φ, the set **CPLP** = Φ and the prespecification of the threshold value ELF_{tol}.
Step 2: Identify a sufficient number of VO strategies using the method described in section 2.2.
Step 3: Determine the optimal VO strategy to partition the DMA into two subregions by solving the bisection optimization problem (Equations 1–6 in section 2.3.1).
Step 4: Implement the optimal VO strategy to partition the DMA into two subregions as well as assigning the PLPs into two subsets (i.e., $G = G_1 \cup G_2$ and $\mathbf{PLP} = \mathbf{PLP}_1 \cup \mathbf{PLP}_2$).
Step 5: Conduct WBAs for the two PLP subsets to estimate their ELFs (i.e., ELF ($\mathbf{PLP}_1$) and ELF ($\mathbf{PLP}_2$)), with details given in section 2.3.2.
Step 6: Update the set **PLP** (removing the NLPs determined in Step 5) as well as the set **NLP** and **CPLP** following the method described in section 2.3.3.
Step 7: If there are more VOs available to further partition the set of PLPs into subsets, move to Step 3, otherwise the pipes in the set **PLP** are identified as the final leaking pipes in the DMA and the proposed method moves to Step 8;
Step 8: If there are CPLPs in the DMA (i.e., **CPLP** ≠ Φ), update the **PLP** using **CPLP,** and the method moves to Step 3, otherwise, all leaks in the DMA are identified and the proposed multistage procedure is terminated.

It is noted that the proposed method cannot handle the case where multiple leaks with estimated total leak flow are greater than ELF_{tol} but each one of these leaks has estimated flow smaller than ELF_{tol}. This case indicates that there are many tiny leaks in the WDS, which cannot be detected using the WBA as proposed in this paper.

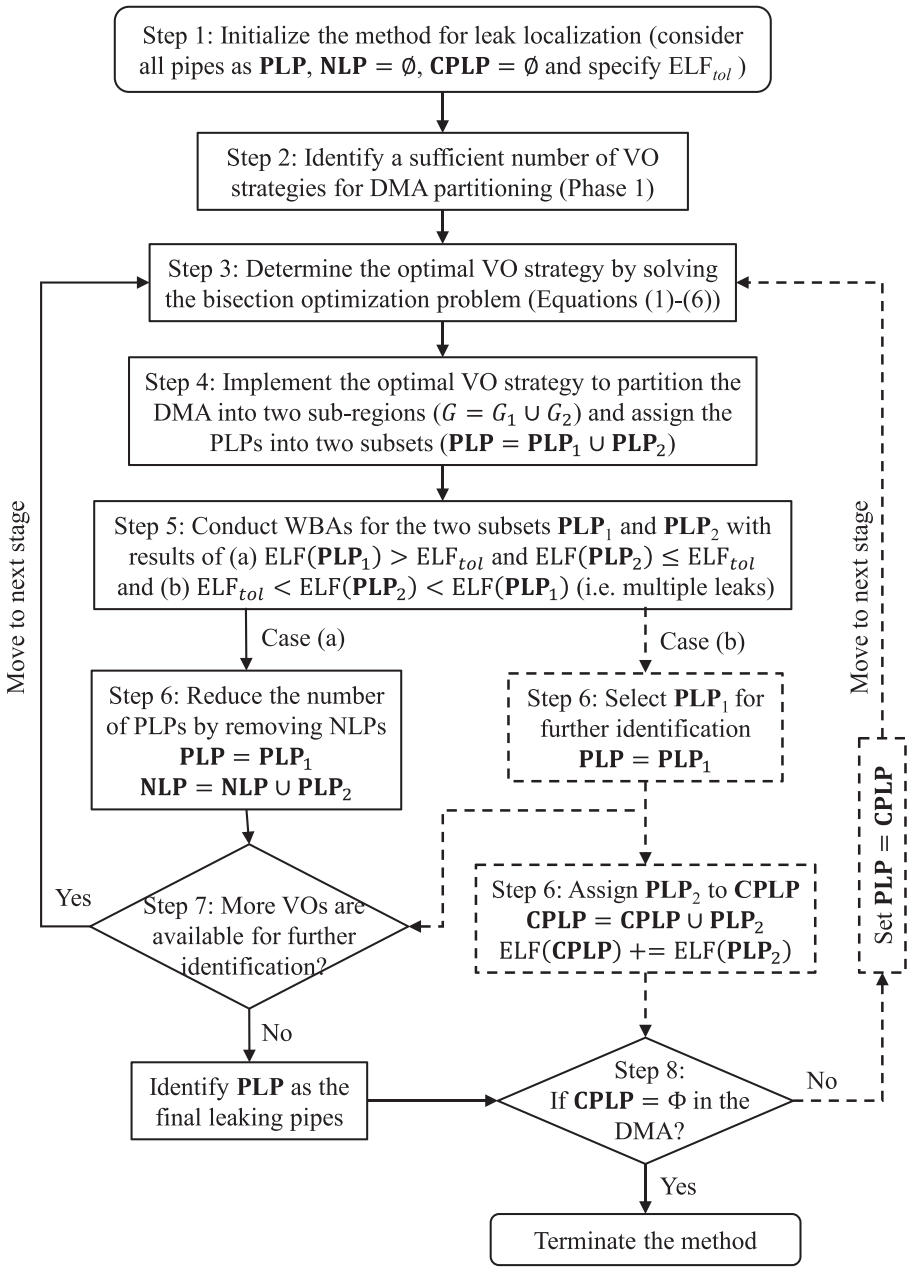


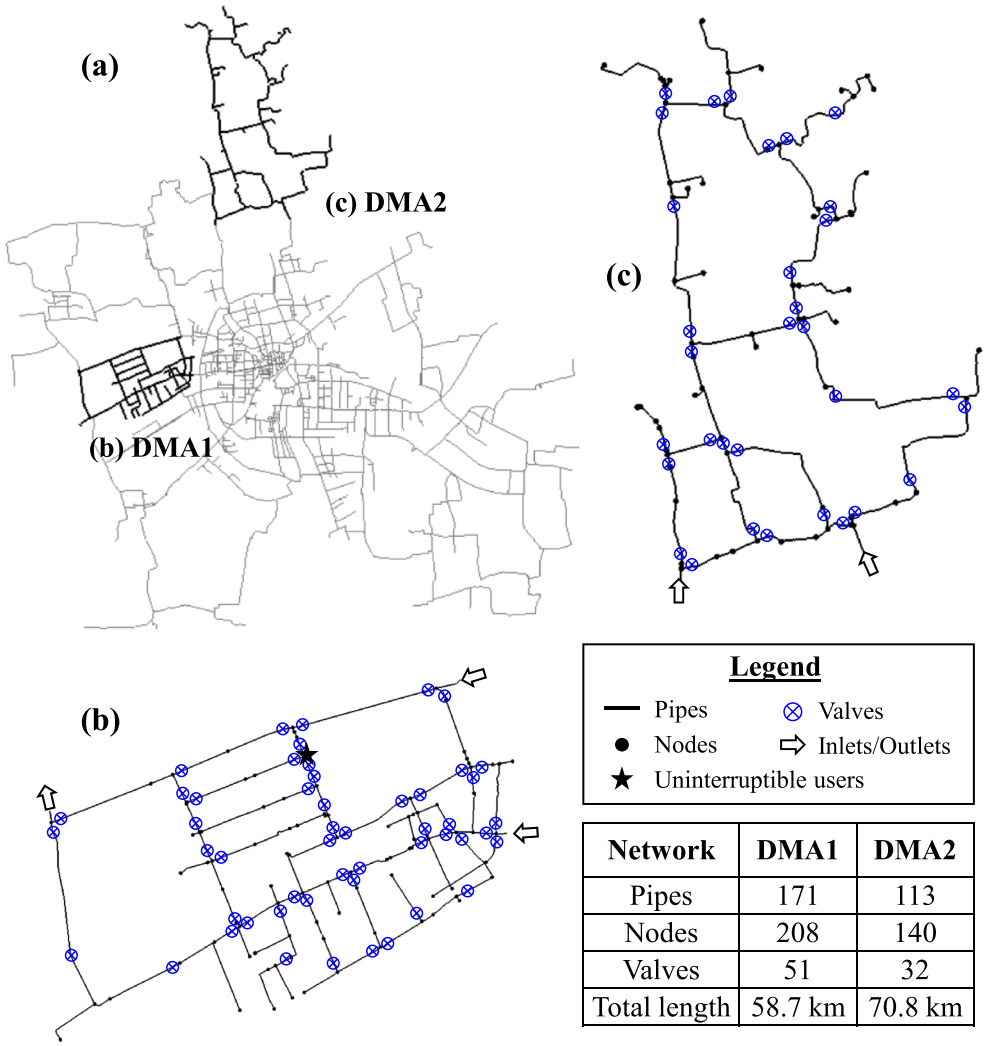
Figure 7. Overall procedures of the proposed multistage method for leak localization.

It should be highlighted that if the WBAs at Step 5 identifies that both PLP subsets have leaks (i.e., Case b in Figure 7), the proposed method is implemented to localize the leaks in the subset with a larger value of the estimated leak flow (ELF). When these identified leaking pipes have been handled (e.g., isolated or repaired), the other subset with a smaller ELF (i.e., the pipes in the **CPLP** set) would be reconsidered by executing the proposed method again.

3. Case Studies

3.1. Description of Two Cases

Two DMAs (DMA1 and DMA2) from a real WDS in China as shown in Figure 8 are used to demonstrate the utility of the proposed multistage leak localization method. The entire WDS serves a population of about 1.2 million with the maximum daily water demand of 0.47 million m^3. The two selected DMAs are of different



Network	DMA1	DMA2
Pipes	171	113
Nodes	208	140
Valves	51	32
Total length	58.7 km	70.8 km

Figure 8. Layouts of (a) a real-world WDS, (b) DMA1, and (c) DMA2.

complexities and scales. According to the hydraulic model of the entire WDS provided by the local water utility, DMA1 is highly looped, consisting of 171 pipes (58.7 km in length), 208 nodes and 51 available valves. DMA2 consists of 113 pipes (70.8 km in length), 140 nodes and 32 available valves as listed in the table in Figure 8. DMA1 has two inlets and an outlet, serving predominantly residential consumers with about 38,000 connections and a large uninterruptible user (requires continuous water supply averaging 51.2 m^3/h). DMA2 is located at the boundary of the city by providing water through two inlets and serves residential users with about 26,700 connections. Flows entering the two DMAs at inlets and outlets are measured and collected online (every 5 min) by the local water utility. Based on the collected data, the water consumptions vary between about 270–1,200 and 180–800 m^3/h for DMA1 and DMA2, respectively.

Smart demand meters have been intensively deployed in these two DMAs over the past few years. This is partly driven by the government pressure to reduce the leakage to 10% of the total water consumptions by the end of 2020 as previously mentioned, as well as because of the rapid development of smart metering technologies (e.g., cheaper meters and powerful data transmission internet, Zheng et al., 2018). For DMA1, a total of 11,600 smart meters have been installed, and the demands at the nodes without data from smart meters are calibrated using the measured data (e.g., flows, pressures, and known demands from smart meters). Prior to demand calibration for nodes without data from smart meters, a temporal pattern is assigned to these nodes based on engineering experience, thereby ensuring the calibration results to be

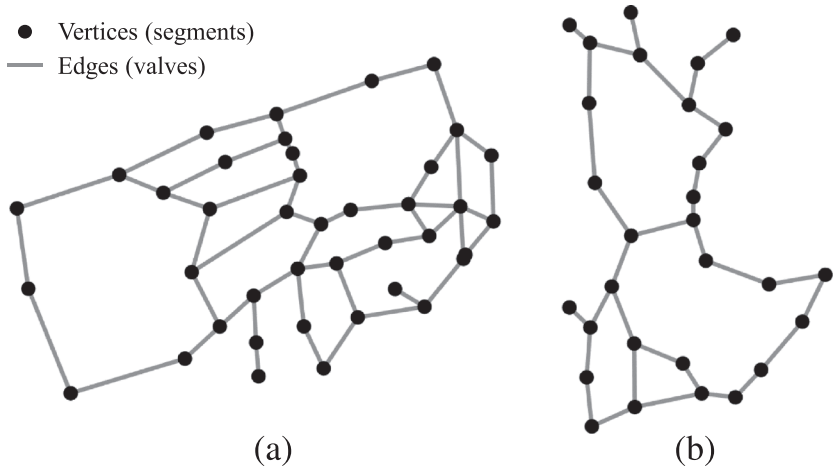


Figure 9. Transformed segment-valve graphs for (a) DMA1 and (b) DMA2.

practically meaningful. Such calibration is not elaborated in this study as this WDS hydraulic model has been already calibrated by the local water utility. Given that DMA2 is selected by the local water utility as a pilot to enable smart leakage management, all users have been measured by smart demand meters, leading to a total of 26,700 m. In this paper, DMA1 is used as a hypothetical case with artificial leaks to demonstrate the implementation procedures of the proposed method for localizing single and multiple leaks, and DMA2 is taken as a real case for the application of the proposed methodology in the field.

3.2. Identification of the Sufficient Number of VOs (Phase 1)

The identified segment-valve graphs of the two DMAs in Phase 1 of the proposed method are given in Figure 9, where the graph of DMA1 consists of 40 vertices (representing segments) and 51 edges (representing valves), and the graph of DMA2 has 29 vertices and 32 edges. The VO strategies were obtained through the cut set identification method as described in Figure 3. As a result, a total of 4,963 and 410 different cut sets (i.e., VO strategies) were identified for DMA1 and DMA2, respectively. The computation times for finding these cut sets of the two segment-valve graphs were about 35 and 9 min, respectively. These times were obtained using a PC with a 12-core Intel Core i9-7980XE (2.6 GHz) CPU. As previously stated, Phase 1 of the proposed method needs only to be carried out once before the implementation of the multistage procedure for leak localization (Phase 2). This also means that it would not affect the efficiency of the proposed method even for DMAs with highly complex configurations (e.g., DMA1).

3.3. Case 1: DMA1 With Artificial Leaks

As previous stated, DMA1 in Figure 8b is used as a hypothetical case to demonstrate the proposed method for localizing leaks. The hydraulic model of DMA1 has been calibrated by the local water utility, and the calibrated nodal demands at different periods of the day were considered as the online data measured by smart demand meters, in order to demonstrate the utility of the proposed method. These nodal demands combined with the inflows/outflow at the inlets/outlet of DMA1 are used to enable the WBA. Based on this calibrated model, artificial leaks were directly added to some pipes of this DMA as fixed nodal demands over time, representing the leak locations to be identified. In addition, the background leaks and measurement noise were ignored for this hypothetical case, that is, the threshold value $ELF_{tol} = 0$. As a result, the NRW obtained from the WBA was directly taken as the ELF, with ELF > 0 indicating the existence of leaks. However, it should be noted that the ELF_{tol} should be carefully determined in real cases as it often represent the impacts of background leaks and measurement noises (more discussions are given in section 4). As shown in Figure 8b, the pressure head at the uninterruptible user located in DMA1 is required to be no lower than 15.0 m to maintain sufficient water supply during the application of the proposed method.

3.3.1. A Single Leak

The key assumption here is that the MNF of DMA1 has increased due to a single leak of 15.0 m^3/h (5.6% of the MNF). The leak was assumed to be located at a pipe of DMA1 as shown in Figure 10a, and the proposed

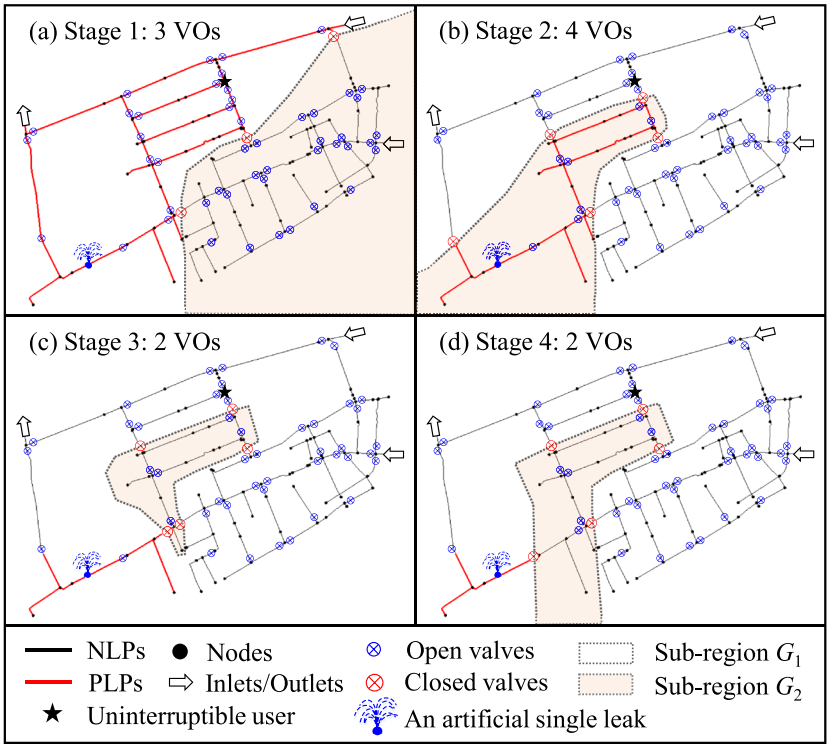


Figure 10. Valve operations (VOs) identified by the proposed method applied to DMA1 with a single leak.

method was used to localize the region associated with this leak following the procedure given in section 2.4. More specifically, the VOs were simulated by the hydraulic model and the WBA was then conducted based on the hydraulic simulations of the inflows, the outflow and the nodal demands for this assumed leak scenario. The identified optimal VOs at each stage of the proposed method are shown in Figure 10. The corresponding ELF results within the WBA and the update procedures at each stage are presented in Table 1. In Table 1, G represents the entire DMA, G_1 and G_2 indicate the two subregions in Figure 10, and $\mathbf{PLP}_1$ and $\mathbf{PLP}_2$ are the two subsets of PLPs corresponding to the two subregions G_1 and G_2, respectively.

As shown in Figure 10a, three valves were closed at Stage 1 (identified by the proposed method) to partition the entire DMA into two subregions, as well as to assign the PLPs (i.e., all pipes in the DMA) into two subsets. The length difference between these two subsets was 0.7 km, indicating the great effectiveness of the proposed graph theory based optimization method. The subsequent WBAs at Stage 1 (see Table 1) would identify that the PLP subset in the subregion G_1 have leaks (i.e., the red lines in Figure 10a) as the hypothetical leak was assumed to locate within this subregion. As a result, the spatial region of leaking pipes was reduced from 58.7 to 29.0 km. At the following Stages 2, 3, and 4, the updated PLPs (i.e., red lines at the previous stages as shown in Figure 10) were further partitioned into subsets with similar pipe length using optimal VOs identified by the proposed method. This led to the reduction of the leak region from 29.0 to 14.7, 6.8, and 4.3 km at Stages 2, 3, and 4, respectively. The assumed leaking pipe was localized at Stage 4 as no more VOs were available to enable a further region reduction.

Table 1
ELF Results of the Proposed Method Applied to DMA1 With a Single Leak

	WBA (m^3/h)					Update (m^3/h)	
Stages	ELF(G)	ELF(G_1)	ELF(G_2)	ELF ($\mathbf{PLP}_1$)	ELF ($\mathbf{PLP}_2$)	ELF (**PLP**)	ELF (**CPLP**)
1	15	15	0	15	0	15	—
2	15	0	15	0	15	15	—
3	15	15	0	15	0	15	—
4	15	15	0	15	0	15	—

In summary, a total of 16 VOs (eight different valves were operated) were involved in the four stages of the proposed method applied to DMA1 with an assumed single leak. This included 11 operations for DMA partitioning, and five additional operations to open all closed valves in order to restore the system to its original state. The proposed method successfully narrowed down the potential leak region from 58.7 to 4.3 km by VOs for DMA1 with a single leak. Consequently, it can significantly improve the efficiency of

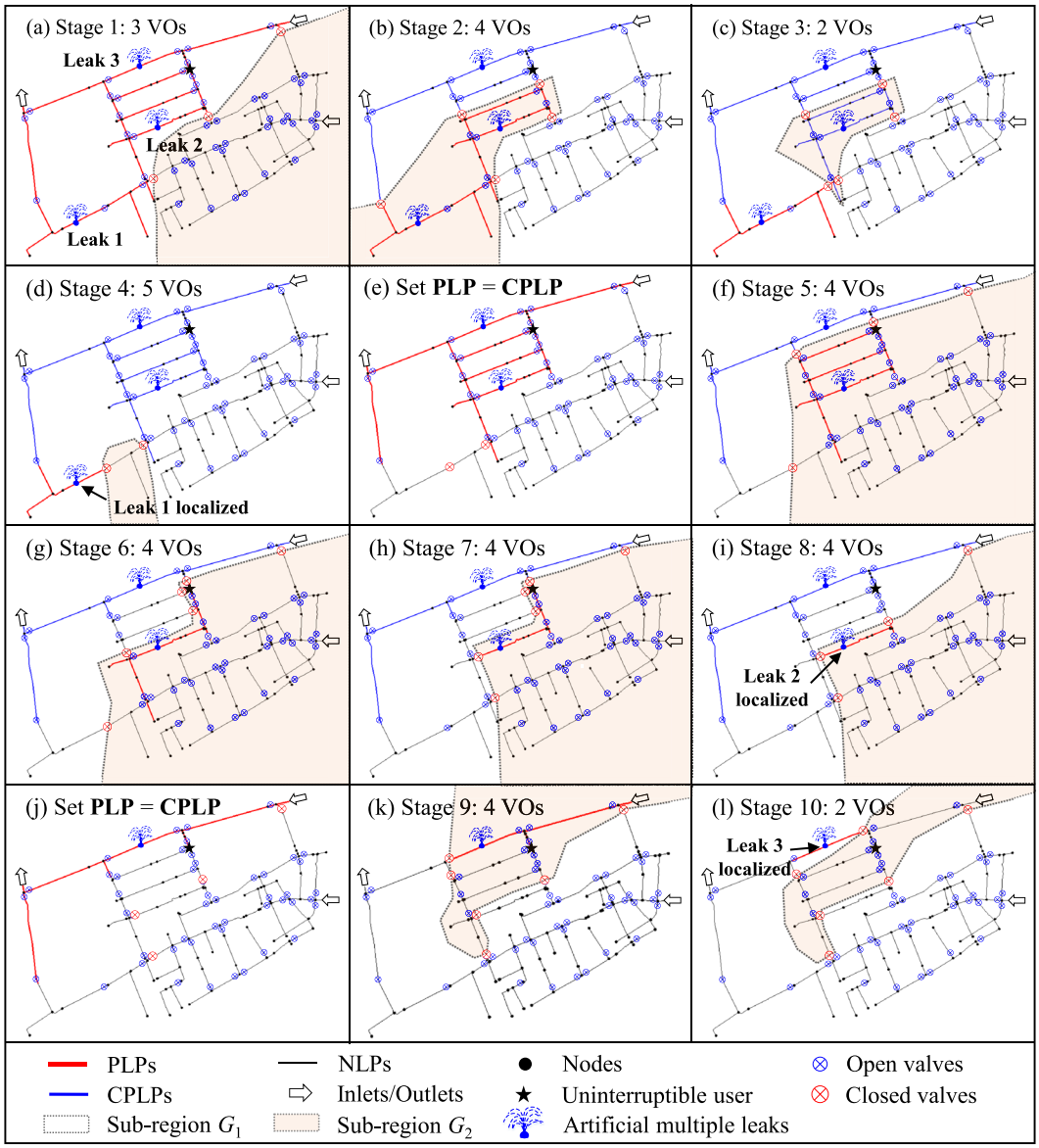


Figure 11. Valve operations (VOs) identified by the proposed method applied to DMA1 with three different leak locations.

leak localization as the equipment can be subsequently used to pinpoint the leak locations within the identified leak region rather than the entire DMA.

3.3.2. Multiple Leaks

Three different leak locations (Figure 11a) were considered for DMA1 as a hypothetical case to further demonstrate the utility of the proposed method in handling multiple leaks. More specifically, Leak 1 was the same as the leak in the previous case (the single leak with 15.0 m^3/h), and Leak 2 and Leak 3 were added with fixed leak flowrates of 10.0 and 5.0 m^3/h, respectively. As a result, the ELF of the DMA was 30.0 m^3/h for this hypothetical case. The proposed method was implemented to localize these artificial leaks following the procedures given in section 2.4. The identified optimal VOs were given in Figure 11, and the corresponding ELF results were presented in Table 2.

As shown in Figure 11 and Table 2, a total of 10 stages were required to identify the three leak regions in DMA1. Stages 1–4 shown in Figures 11a–11d were performed to identify the leak region associated with Leak 1 as it has the largest leak flow. Particularly, at Stages 2 and 3, the WBAs would identify that both

Table 2
ELF Results of the Proposed Method Applied to DMA1 With Three Different Leak Locations

	WBA (m^3/h)					Update (m^3/h)	
Stages	ELF(G)	ELF(G_1)	ELF(G_2)	ELF (**PLP**$_1$)	ELF (**PLP**$_2$)	ELF (**PLP**)	ELF (**CPLP**)
1	30	30	0	30	0	30	—
2	30	5	25	5	25	25	5
3	30	20	10	15	10	15	15
4	30	30	0	15	0	15	15
Identify a leak region (Leak 1), set **PLP** = **CPLP**						15	—
5	15	5	10	5	10	10	5
6	15	5	10	0	10	10	5
7	15	5	10	0	10	10	5
8	15	5	10	0	10	10	5
Identify another leak region (Leak 2), set **PLP** = **CPLP**						5	—
9	5	0	5	0	5	5	—
10	5	5	0	5	0	5	—
Identify the third leak region (Leak 3)							

PLP subsets have leaks, as shown in Table 2. Taking Stage 2 for example, as ELF (**PLP**$_1$)<ELF (**PLP**$_2$), the **PLP** set was updated as **PLP** = **PLP**$_2$ for further partition (the red lines in Figure 11b). The subset **PLP**$_1$ was considered as the **CPLP** set as shown by the blue lines in Figure 11b with an ELF (**CPLP**) = 5.0 m^3/h (see Stage 2 in Table 2).

After Stage 4, as CPLPs were available (i.e., **CPLP** ≠ Φ), the **PLP** set was reset as **PLP** = **CPLP** for further partition, where the blue lines in Figure 11d changed to red lines in Figure 11e. Before reexecution of the proposed method, it was assumed that the identified Leak 1 was repaired, thus ELF(G) of the entire DMA was reduced to 15.0 m^3/h, as shown in Table 2. Stages 5–8 were further executed to localize the leak region associated with Leak 2 as illustrated in Figures 11f–11i. The WBAs at Stage 5 would identify that both PLP subsets have leaks and the PLPs within the subregion G_1 was considered as CPLPs (blue lines in Figure 11f) as ELF (**PLP**$_1$) < ELF (**PLP**$_2$) (Table 2). Finally, the proposed method (Stages 9 and 10) was implemented again to the PLPs in Figure 11j, in order to identify the third leak region that had Leak 3.

Consequently, a total of 10 stages involving 42 VOs based on 18 different valves (including six VOs for restoring the system in the end) were used to localize three leaks in the DMA. The finally identified leak regions were 4.3, 2.3, and 1.9 km, respectively, which were also significantly lower than the total length of pipes in the DMA (i.e., 58.7 km). This result indicated that the proposed method was also, theoretically, effective and efficient for the localization of multiple leaks.

3.4. Sensitivity Analysis

3.4.1. Influence of the Valve Density

Given that the number of available valves in a DMA can be variable, the number of different valves that can be operated in DMA1 was considered for a sensitivity analysis to investigate the performance variation of the proposed method for leak localization. Four scenarios with different valve densities in DMA1 were considered, including the original 51 valves, with the three reduced scenarios of 40 valves, 30 valves and 20 valves, as shown in Figure 12. For each valve scenario, assuming that a single leak existed at each of the pipes in the DMA, leading to a total of 171 hypothetical cases (i.e., 171 pipes in DMA1), representing all possible leak locations in the DMA. The proposed method was implemented to all these cases for leak localization and the results for each valve scenario were summarized in Figure 13.

Figure 13 presents the statistical results of the number of VOs used for leak localization and the identified leak region (in length) for the four-valve scenarios in boxplot form (the crosses indicate the average values). As shown in this figure, the proposed method was able to localize leaks to leak regions of 1.0–13.0 km (i.e., 98.3–77.9% reduction in spatial scale relative to the entire DMA) within a range of 10 to 25 VOs. This implied that the proposed method was effective for leak localization, even for the DMA with a limited number of available valves (e.g., 20 valves in DMA1). A clear and reasonable trend was also

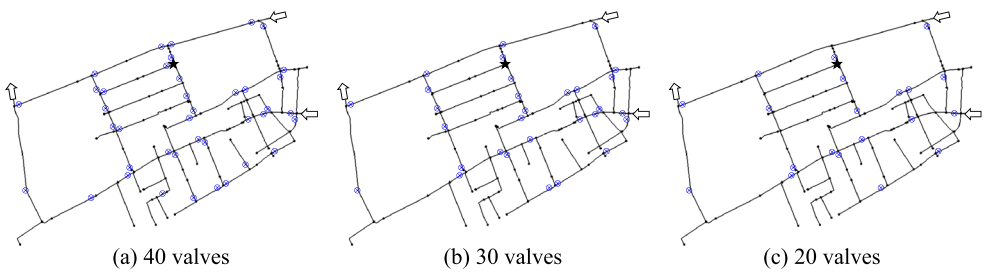


Figure 12. Scenarios of available valves for DMA1: (a) 40 valves, (b) 30 valves, and (c) 20 valves.

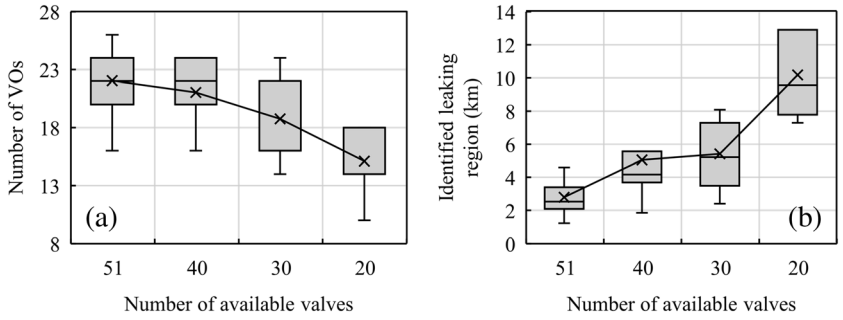


Figure 13. Boxplot of (a) the number of VOs for leak localization and (b) the identified leak region (in length) for the four-valve scenarios (the crosses indicate the average values).

observed from Figure 13 that the number of VOs for leak localization decreased and the identified leak region increased with the reduction of available valves from 51 to 20. This revealed a trade-off relationship between the cost (related to the number of VOs) and the efficiency of the proposed method for leak localization, which should be considered by practitioners in practical applications.

3.4.2. Influence of the Weight in Equation 1

As previously stated, the variation of the weight w in Equation 1 may have impacts on the optimization results. While $w = 0.5$ was used in the proposed method, the influence of different w values was investigated. Five different weight values ($w = 0.1$, 0.3, 0.5, 0.7 and 0.9) were considered to enable a sensitivity analysis. For each weight value, all possible leak locations in DMA1 were considered to be localized using the proposed method. This was done by assuming each of the 171 pipes in DMA1 as a hypothetical case with a single leak at the pipe. Figure 14 presents the boxplot of the number of VOs of all leak cases for each weight value. In this figure, the average value represents the mean of the number of VOs over all leak scenarios, with a lower value indicating an overall higher efficiency. The range of the box in Figure 14 represents the variation of the number of VOs across different leak scenarios, with a lower range indicating a better robustness in efficiency. It can be observed from Figure 14 that the value of $w = 0.5$ represented the overall best performance as it possessed the lowest average number of VOs over all leak scenarios, as well as rather high efficiency robustness across different leak scenarios. However, it can be also deduced from Figure 14 that a slight variation of w value (if say $w = 0.3$ or 0.7) would not significantly affect the results of the proposed method.

3.5. Case 2: Real Case Study—DMA2

The proposed method was applied to the real case of DMA2 to demonstrate its utility in practical applications. Flow measurements from two inlets and smart metering demands at all users within the DMA were available with a 5 min time resolution. According to the collected data of the DMA throughout June 2019, the recorded minimum inflows at night hours between 02:00 and 05:00 increased to approximately 180.0 m^3/h while the corresponding demands measured by smart demand meters were about 151.8 m^3/h. Therefore, the NRW of the DMA at night hours was 28.2 m^3/h (i.e., 15.7% of the total inflow), implying that this DMA was highly likely to have a leak(s). Since no pipe bursts were reported to the water utility, it was deduced that leaks may be invisible (leaks to the underground). Therefore, the proposed multistage method was employed to localize leaks in DMA2 with the approval of the local water utility.

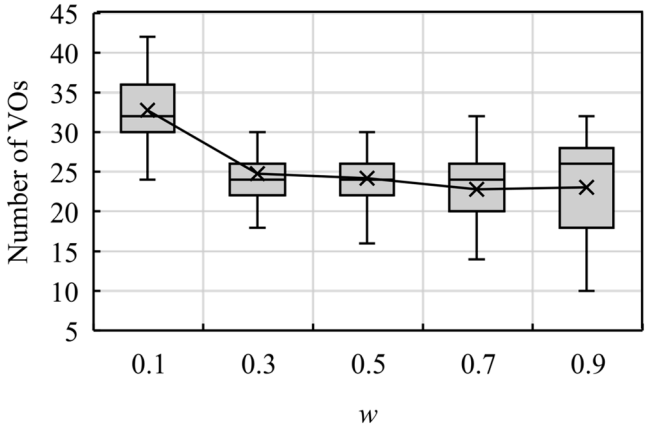


Figure 14. Boxplot of the number of VOs for leak localization for different weight values (the crosses indicate the average values).

The ELF for this DMA was estimated as ELF(G) = 28.2 m^3/h, where G represented the entire DMA. Meanwhile, a threshold value ELF_{tol} = 7.2 m^3/h (i.e., 2.0 L/s) was recommended by the staff in the local water utility based on their engineering experience (based on the historical data of flows and demands in this DMA), which approximately represented background leaks and meter inaccuracies of DMA2. The impacts of

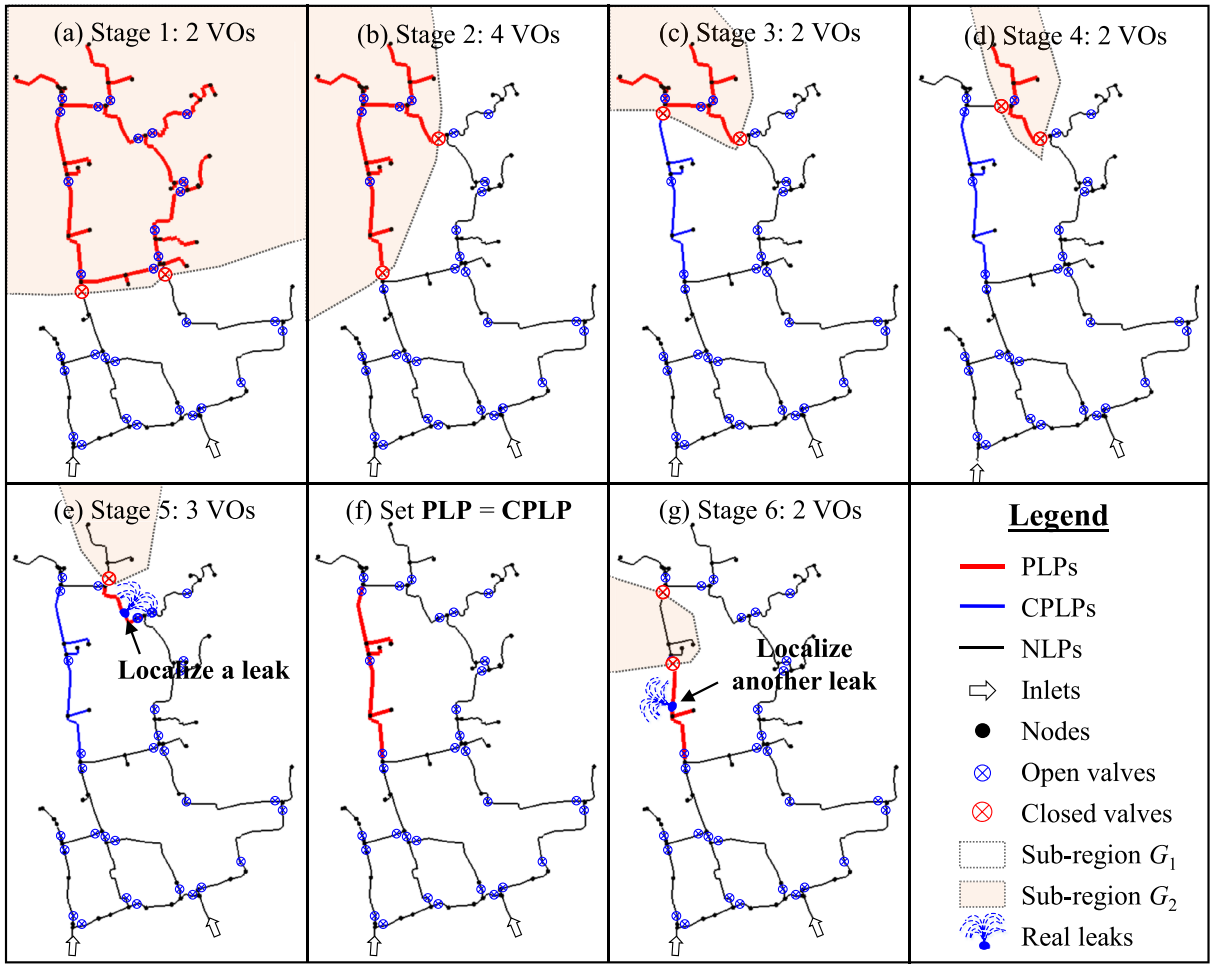


Figure 15. Valve operations (VOs) and identification results at each stage for localizing real leaks in DMA2.

the use of ELF_{tol} are further discussed in section 4. The proposed method was implemented during the night hours of 5 July 2019 starting from 02:00 am. At each stage of the proposed multistage method, the optimal VOs were implemented by the staff of the local water utility. To avoid inducing severe transient flows during VOs, the valves were operated slowly over a period of 3 to 10 min. This was followed by real-time WBAs applied to the DMA after a period of time (e.g., 15–30 min) of the VOs, in order to obtain stable and reliable ELF results. The results were given in Figures 15 and 16, where Figure 15 presented the VO operations at each stage and Figure 16 outlined the processes of the WBAs for this real application of the proposed method.

As shown in the two figures, a total of six stages of valve manipulations were used to identify ultimately two leaks in DMA2. More specifically, the proposed method utilized five stages (i.e., Stages 1–5) to identify a leak region (2.3 km in pipe length) as indicated in Figure 15e with red lines. At Stage 3 (Figure 15c), the WBA indicated that ELF(G_1) and ELF(G_2) were on average 12.2 and 15.8 m^3/h, respectively (Figure 16). Given that ELF (**PLP**$_1$) = ELF(G_1) and ELF (**PLP**$_2$) = ELF(G_2), ELFs for both subsets were greater than ELF_{tol}. This implied that both the PLP subsets had leaks, that is, the case of multiple leaks. As a result, the **PLP**$_1$ subset was temporarily considered as the **CPLP** set (the blue lines in Figure 15c) due to ELF (**PLP**$_1$) < ELF (**PLP**$_2$). Consequently, when a leak region was identified after Stage 5, CPLPs existed in the DMA with ELF (**CPLP**) = 12.2 m^3/h as illustrated in Figure 16. Therefore, the leak localization method should be carried on to further identify other leaks, but this was done after the identified leaking pipes were handled.

Based on the identified leak region, several engineers from the local water utility pinpointed the leak spot using listening rods and repaired this leak. This was followed by the application of the proposed method

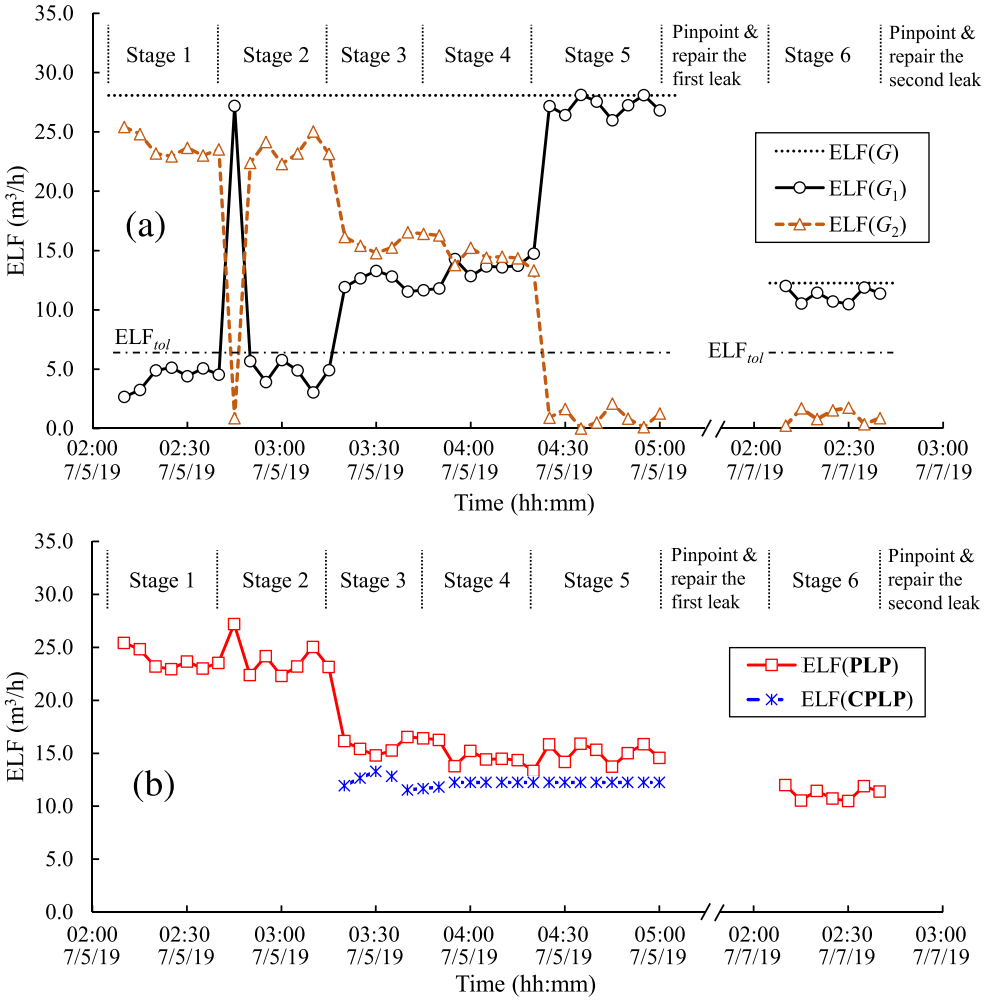


Figure 16. ELF results of (a) the WBAs and (b) the update at each stage for localizing real leaks in DMA2.

again during the night hours of 7 July 2019 by setting **PLP** = **CPLP**, as shown in Figure 15f and Figure 16. As a result, the ELF of the entire DMA became ELF(G) = ELF (**CPLP**) = 12.2 m^3/h (Figure 16). Within this application, only one stage with two VOs (i.e., Stage 6) was used to identify another leak region with the pipe length of 4.2 km in the DMA (the red lines as shown in Figure 15g). Given that no CPLPs were available in the DMA, the leak localization method was terminated, and this was followed by that engineers were sent again to pinpoint the exact location of the leak within the identified region, as indicated in Figure 15g.

Consequently, DMA2 with an ELF of 28.2 m^3/h at night hours was identified having two leaks by the proposed method, where 18 operations of 8 different valves (i.e., 15 VOs for DMA partitioning and 3 VOs for system restoring) were involved. Two leak regions were identified with the total pipe length of 2.3 and 4.2 km. As a result, the spatial scale of the leak region was reduced by 90.8% relative to the entire DMA2 (i.e., 70.8 km). This has significantly saved the efforts used for the subsequent applications of the equipment-based method for pinpointing the leak locations (e.g., the use of listening rods in this case).

The flowrates of the two identified leaks were estimated to be about 15.0 and 10.0 m^3/h, respectively, from the WBA results (i.e., the ELF (**PLP**) at Stages 5 and 6). Therefore, the benefit of the presented method in water savings can be approximately evaluated compared to the use of listening rods for leak localization. More specifically, the method of listening rods has to traverse the entire DMA2 about 3 weeks as stated by the engineers in the local water utility based on their experience, while the proposed method can localize

Table 3
Hydraulic Variations Induced by Valve Operations for the Presented Cases

Cases	Valve operations	Pressure fluctuation (m)		Pipes with reversed flow(%)
		Average	Maximum	
The single leak in DMA1	16	0.4	0.5	2.2
Multiple leaks in DMA1	42	1.4	2.8	21.7
The real case in DMA2	18	1.3	2.3	2.4

these two leaks in two nights. Accordingly, the water saving of the proposed method could be roughly estimated as the total leak volumes in 19 days at a flowrate of 25.0 m^3/h, which was 11,400 m^3 in total (demands of about 25,000 families for a day).

4. Analysis of the Real Application of the Proposed Method

The presented cases with artificial and real leaks demonstrated that the proposed method can effectively localize leaks in a DMA by narrowing down the potential leak region from the entire DMA to a very limited area (e.g., around 10% in spatial scale). The method only utilizes VOs and WBA, hence it is very efficient and easy to implement for water utilities with smart demand meters installed. Nevertheless, a drawback of the presented method lies in the VOs for DMA partitioning, which may lead to a disturbance or even an interruption to supply in (part of) the DMA. To mitigate such effect induced by VOs, the proposed method is implemented during the night flow period. Furthermore, hydraulic analyses were performed to examine the hydraulic conditions (i.e., nodal pressures and pipe flows) as a result of the VOs within the multistage method applied to the two DMAs. The average and maximum pressure fluctuations across all DMA nodes as well as the lengths of pipes with the reversed flow (relative to the total length of the DMA) were summarized in Table 3.

The results in Table 3 generally indicated that the hydraulic status of nodes and pipes were not significantly affected by VOs, except for those within the isolated regions during the VOs for a short period of time. This is especially the case for the single leak in DMA1 and the real case in DMA2. It thus implied that the VOs of the proposed method had rather limited hydraulic impacts on the DMAs. For DMA1, it was also found that the ratio of pipes with reversed flow increased from 2.2% to 21.7% as the number of VOs increased from 16 to 42. The pipes with reversed flow directions may have water quality issues (Qi, Zheng, Guo, Zhang, et al., 2018), which should be also carefully realized before a large number of VOs.

The utility of the proposed method depends on the accuracy of the WBA applied to the DMA and the hydraulic model used. In other words, the selection of the threshold ELF_{tol} is vital to enable the effective application of the method. As previously described, the ELF_{tol} is used to approximate the background leaks as well as meter inaccuracies of flow measurements at inlets/outlets and smart demand meters of users. The background leak rate is typically correlated with the pipe conditions (e.g., age and length), the number of connections and the pressure (Alkasseh et al., 2013). That rate can be roughly evaluated by empirical and semiempirical methods developed in literature (Hunaidi, 2010; Tabesh et al., 2009). For the meter inaccuracies, the best policy is to calibrate them frequently to minimize the errors and evaluate these errors by conducting sampling tests. In addition, as the proposed method is implemented at night hours when water usage is at a minimum rate, the errors caused by demand meters are minimized as well. Prior to the application of the proposed method, the ELF_{tol} can be estimated based on long-term monitoring and analysis of the flows and demands based on the operation and maintenance of the DMA (e.g., the pipe and meter conditions). In addition, if this threshold can be evaluated (and hence adjusted) at different stages associated with subregion sizes, this should be done as this would further improve the effectiveness of the leak localization method.

It should be noted that the proposed method is similar to the traditional step-testing approach (Puust et al., 2010) in utilizing the VOs for leak localization. However, the VO strategy of the proposed method is dynamically optimized for each stage of the method implementation. That makes it significantly different from the traditional step-testing approach where VO strategies are typically prespecified during the planning stage (i.e., VOs are empirically determined). Consequently, the proposed method is capable of handling complex DMAs with highly looped structures and a large number of valves while the traditional step-testing approach is difficult, if not impossible, in such cases. In addition, the proposed method utilizes a multistage bisection approach to localize leaks in the DMA; thus, it is very efficient for leak localization with only a few VOs. However, for the traditional step-testing approach, the leaks are generally identified by traversing all pipes from the terminal of the DMA to the source, leading to a large number of VOs. Therefore, the traditional step-testing approach is very labor-intensive and thus has been used sparingly by water utilities.

5. Summary and Conclusions

Techniques for leak localization have been consistently developed in water industries worldwide, especially with the advancement of smart WDSs in recent years. This paper presented a multistage approach for leak localization in district metering areas (DMAs) with the aid of VOs and smart demand metering. More specifically, the proposed method utilized the operations of existing valves to partition the DMA into two subregions. This was followed by the application of the real-time WBA based on smart demand meters to identify whether the PLPs within the two subregions have real leaks. This process was repeatedly performed (i.e., multiple stages), thereby reducing the number of PLPs in the DMA and accordingly narrowing down the spatial regions with leaks.

The utility of the proposed method has been demonstrated using two real DMAs with different topology properties and scales. DMA1 was considered as a hypothetical case, where a single leak and multiple leaks (i.e., three leaks) were assumed to be localized, respectively. The application results indicated that the proposed method was able to localize the leaks to a smaller region with 4.3 km of pipes using 16 VOs for the case of a single leak. For the three leak locations, the proposed method was able to identify leak regions with 4.3, 2.3, and 1.9 km of pipes, respectively, using a total 42 VOs (the total pipe length of DMA1 is 58.7 km). For DMA2, the method was applied on a real case for which the demands within the DMA have been measured by smart meters. Results showed that the proposed method with 18 VOs was capable of identifying two leak regions within 2.3 and 4.2 km of pipe lengths (the total pipe length of DMA2 is 70.8 km). The leaks within the two leaking regions were then efficiently pinpointed by engineers. This can lead to a water saving of approximate 11,400 m^3 compared to the traditional method that uses listening rods to traverse the entire DMA for leak localization. The two presented cases confirmed that the proposed method is effective and efficient for leak localization. Such a multistage localization method as proposed in this paper is very promising given that smart meters and remote control valves are being increasingly installed in water distribution networks.

It is noted that the implementation of the proposed method may induce interruptions of water supply for parts of the DMA (except for uninterruptible users) as well as water quality issues (e.g., discoluration) as a result of VOs and reversed flows. However, short-term water supply interruptions during low flow periods are often acceptable in practice, especially if the local residents can be notified before the applications of the operational strategies. An ad hoc method for dealing with the potential water quality issue is the assessment of this risk with the aid of hydraulic simulations (e.g, determine which pipes can have inversed flows due to VOs) before the application of the methodology. That allows engineering practices to be adopted to address this problem (e.g., evaluating the risk of water quality issues and potential mitigation strategies).

Data Availability Statement

The data/models used in this study have been submitted as the supporting documents.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments
This work is funded by the National Natural Science Foundation of China (Grant No. 51922096), and the Excellent Youth Natural Science Foundation of Zhejiang Province, China (LR19E080003).

References

Alkasseh, J. M., Adlan, M. N., Abustan, I., Aziz, H. A., & Hanif, A. B. M. (2013). Applying minimum night flow to estimate water loss using statistical modeling: A case study in Kinta Valley, Malaysia. *Water Resources Management*, *27*, 1439–1455. https://doi.org/10.1007/s11269-012-0247-2

Besner, M., Michèle, P., & Stig, R. (2011). Assessing the public health risk of microbial intrusion events in distribution systems: Conceptual model, available data, and challenges. *Water Research*, *45*(3), 961–979. https://doi.org/10.1016/j.watres.2010.10.035

Bragalli, C., Neri, M., & Toth, E. (2019). Effectiveness of smart meter-based urban water loss assessment in a real network with synchronous and incomplete readings. *Environmental Modelling & Software*, *112*, 128–142. https://doi.org/10.1016/j.envsoft.2018.10.010

China Urban Water Association (2015). *Statistical yearbook of urban water supply*. Beijing, China: China Statistical Press.

Colombo, A. F., Lee, P., & Karney, B. W. (2009). A selective literature review of transient-based leak detection methods. *Journal of Hydro-Environment Research*, *2*(4), 212–227. https://doi.org/10.1016/j.jher.2009.02.003

Covas, D., Jacob, A., & Ramos, H. (2006). Bottom-up analysis for assessing water losses: A case study. *Water Distribution Systems Analysis Symposium*, 2006.

Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P. R., & Walski, T. (2019). Real time control of water distribution networks: A state-of-the-art review. *Water Research*, *161*, 517–530. https://doi.org/10.1016/j.watres.2019.06.025

Creaco, E., Kossieris, P., Vamvakeridou-Lyroudia, L., Makropoulos, C., Kapelan, Z., & Savic, D. (2016). Parameterizing residential water demand pulse models through smart meter readings. *Environmental Modelling & Software*, *80*, 33–40. https://doi.org/10.1016/j.envsoft.2016.02.019

Duan, H. F., Lee, P. J., Ghidaoui, M. S., & Tung, Y. K. (2011). Leak detection in complex series pipelines by using the system frequency response method. *Journal of Hydraulic Research*, *49*(2), 213–221. https://doi.org/10.1080/00221686.2011.553486

Farah, E., & Shahrour, I. (2017). Leakage detection using smart water system: Combination of water balance and automated minimum night flow. *Water Resources Management*, *31*(15), 4821–4833. https://doi.org/10.1007/s11269-017-1780-9

Ferrari, G., Savic, D., & Becciu, G. (2014). Graph-theoretic approach and sound engineering principles for design of district metered areas. *Journal of Water Resources Planning and Management*, *140*, 04014036. https://doi.org/10.1061/(asce)wr.1943-5452.0000424

Giustolisi, O., & Savic, D. (2010). Identification of segments and optimal isolation valve system design in water distribution networks. *Urban Water Journal*, *7*(1), 1–15. https://doi.org/10.1080/15730620903287530

Gupta, A., & Kulat, K. D. (2018). A selective literature review on leak management techniques for water distribution system. *Water Resources Management*, *32*(10), 3247–3269. https://doi.org/10.1007/s11269-018-1985-6

Hunaidi, O. (2010). *Leakage management for water distribution infrastructure-report 2: Results of DMA experiments in Ottawa, ON*. Canada: National Research Council.

Karger, D. (1993). Global min-cuts in RNC and other ramifications of a simple Mincut algorithm. *Proc. 4th Annual ACM-SIAM Symposium on Discrete Algorithms*.

Mutikanga, H. E., Sharma, S. K., & Vairavamoorthy, K. (2013). Methods and tools for managing losses in water distribution systems. *Journal of Water Resources Planning and Management*, *139*(2), 166–174. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000245

Nguyen, K. A., Stewart, R. A., Zhang, H., Sahin, O., & Siriwardene, N. (2018). Re-engineering traditional urban water management practices with smart metering and informatics. *Environmental Modelling & Software*, *101*, 256–267. https://doi.org/10.1016/j.envsoft.2017.12.015

OECD (The Organisation for Economic Co-operation and Development) (2012). *OECD environmental outlook to 2050: The consequences of inaction*. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264122246-en

Puust, R., Kapelan, Z., Savic, D. A., & Koppel, T. (2010). A review of methods for leakage management in pipe networks. *Urban Water Journal*, *7*(1), 25–45. https://doi.org/10.1080/15730621003610878

Qi, Z., Zheng, F., Guo, D., Maier, H. R., Zhang, T., Yu, T., & Shao, Y. (2018). Better understanding of the capacity of pressure sensor systems to detect pipe burst within water distribution networks. *Journal of Water Resources Planning and Management*, *144*(7), 04018035. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000957

Qi, Z., Zheng, F., Guo, D., Zhang, T., Shao, Y., Yu, T., et al. (2018). A comprehensive framework to evaluate hydraulic and water quality impacts of pipe breaks on water distribution systems. *Water Resources Research*, *54*, 8174–8195. https://doi.org/10.1029/2018WR022736

Rajeswaran, A., Narasimhan, S., & Narasimhan, S. (2018). A graph partitioning algorithm for leak detection in water distribution networks. *Computers & Chemical Engineering*, *108*, 11–23. https://doi.org/10.1016/j.compchemeng.2017.08.007

Romano, M., Kapelan, Z., & Savić, D. A. (2014). Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management*, *140*(4), 457–467. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000339

Rossman, L. A. (2000). *EPANET 2: User's manual*. Cincinnati: USEPA.

Sanz, G., Perez, R., Kapelan, Z., & Savic, D. (2015). Leak detection and localization through demand components calibration. *Journal of Water Resources Planning and Management*, *142*, 04015057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000592

Sophocleous, S., Savic, D., & Kapelan, Z. (2019). Leak localization in a real water distribution network based on search space reduction. *Journal of Water Resources Planning and Management*, *145*(7), 04019024. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001079

Tabesh, M., Yekta, A. H. A., & Burrows, R. (2009). An integrated model to evaluate losses in water distribution systems. *Water Resources Management*, *23*(3), 477–492. https://doi.org/10.1007/s11269-008-9284-2

Walski, T. M. (1993). Water distribution valve topology for reliability analysis. *Reliability Engineering and System Safety*, *42*(1), 21–27. https://doi.org/10.1016/0951-8320(93)90051-Y

Wang, X., Waqar, M., Yan, H. C., Louati, M., Ghidaoui, M. S., Lee, P. J., & Karney, B. (2020). Pipeline leak localization using matched-field processing incorporating prior information of modeling error. *Mechanical Systems and Signal Processing*, *143*, 106849. https://doi.org/10.1016/j.ymssp.2020.106849

Wright, R., Abraham, E., Parpas, P., & Stoianov, I. (2015). Control of water distribution networks with dynamic DMA topology using strictly feasible sequential convex programming. *Water Resources Research*, *51*, 9925–9941. https://doi.org/10.1002/2015WR017466

Zhang, Q., Wu, Z. Y., Zhao, M., Qi, J., Huang, Y., & Zhao, H. (2016). Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *Journal of Water Resources Planning and Management*, *142*(11), 04016042. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000661

Zheng, F., Tao, R., Maier, H. R., See, L., Savic, D., Zhang, T., et al. (2018). Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Reviews of Geophysics*, *56*, 698–740. https://doi.org/10.1029/2018RG000616

Zheng, F., Zecchin, A., Maier, H., & Simpson, A. (2016). Comparison of the searching behavior of NSGA-II, SAMODE, and Borg MOEAs applied to water distribution system design problems. *Journal of Water Resources Planning and Management*, *142*, 04016017. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000650

Zheng, F., Zecchin, A., Newman, J., Maier, H., & Dandy, G. (2017). An adaptive convergence-trajectory controlled ant colony optimization algorithm with application to water distribution system design problems. *IEEE Transactions on Evolutionary Computation*, *21*, 773–791. https://doi.org/10.1109/TEVC.2017.2682899

Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K., & Fu, G. (2019). Deep learning identifies accurate burst locations in water distribution networks. *Water Research*, *166*, 115058. https://doi.org/10.1016/j.watres.2019.115058