FAST LOCALIZATION OF MULTIPLE LEAKS IN WATER DISTRIBUTION NETWORK JOINTLY DRIVEN BY SIMULATION AND MACHINE LEARNING

Zhirong Li[1,2], Kunlun Xin[1,2]
[1]College of Environmental Science and Engineering, Tongji University, 200092, Shanghai, China
[2]Smart Water Joint Innovation R&D Center, Tongji University, 200092, Shanghai, China
[1]Xkl@tongji.edu.cn

ABSTRACT

INTRODUCTION

Water distribution network (WDN) is one of the most important infrastructures, undertaking the task of transporting safe and reliable water to millions of users for production and life. However, due to many reasons such as the poor-quality pipes, aging facilities and man-made destruction, pipelines are vulnerable to be damaged, thus resulting in energy waste and potential contamination.

Traditional leakage detection relies on the hardware equipments, such as listening stick and correlator, requiring a lot of time, labor and financial support. With the development of supervisory control and data acquisition (SCADA) systems and computer science, it became possible to use real-time pressure and flow data measured in WDN to identify and locate leaks. A number of methodologies based on different kind of sensor measurements and various algorithms have been proposed to detect and locate leakage events in recent years. The key principle of the above methods is the deviation between the real-time data and the simulation data that obtained from the hydraulic model. The leakage recognition and localization models are then established by analyzing the deviations through methods such as mathematical models, heuristic approaches, and machine learning. For example, Farley et al. [1] identifies the leak location in WDN or a district metering area(DMA) through calculating sensitivity matrix of pressure/flow measurements after applying leaks to potential burst locations. Sun et al. [2] combines two different machine learning classifiers based on linear discriminant analysis and neural networks with Bayesian temporal reasoning to find the node that is most likely to leak. Andrew et al. [3] realizes the identification of multiple leaky nodes by successive linear approximation. In addition, Casillas et al. [4] proposes a leak location framework using genetic algorithm (GA) and particle swarm optimization (PSO) based on the optimized sensor placement.

In this research we propose a novel Multiple Leak Detection and Isolation Framework (MLDIF) based on existing sensor layout. This framework contains three parts: calibration, detection and isolation. First, a gradient iteration algorithm with variable steps is used in the calibration part. Then, Seasonal-Trend decomposition procedure based on Loess (STL decomposition) and K-means clustering algorithm are applied to identify the occurrence time of leaks and approximate quantity of leakages. To not only reduce the probable leakage range but also locate the leakage position accurately in the shortest time, leak feature of each pipe is modeled using residual vector obtained from comparing the pressure measurements with their estimation using a hydraulic model. Finally, by matching the real leakage characteristics with the modeled feature of each pipe, the most likely leaky pipe can be figured out. The isolation result is less sensitive to the identification error of leakage quantity, so the leak events can be effectively located. Compared with the methods that use pure machine learning, the overall framework is time-saving, especially for complex pipeline networks with long-term monitoring data and multiple leaks, because the former is highly targeted to the operating conditions of the pipeline network.

METHODOLOGY

Fig. 1 shows the flowchart of the proposed MLDIF and detailed introduction of each part is as follows:

Calibration

The model-based approaches rely heavily on the accuracy of the hydraulic model, which will be greatly affected by the estimation of node consumptions and pipelines parameters. In order to calibrate parameters before leakage identification, a novel and efficient method based on the pressure sensitivity matrix and gradient-based algorithm is proposed, which can simultaneously identify existing regional leaks and perform calibration of pipe roughness. By dividing the leakage areas and pipes into groups using Dijkstra algorithm, an overdetermined linear problem is constructed based on the Jacobian matrix. The quantity of regional leakage and pipe roughness of each group can thus be obtained after several iterations through the least square method.

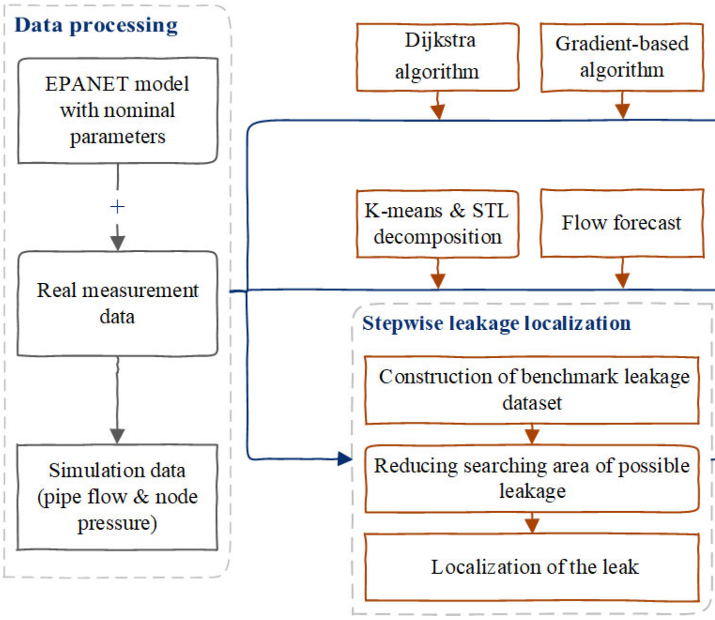

Figure 1. A flowchart of MLDIF.

Identification

To identify leaks that occur at different time or even a combination of multiple leaks during a long period of time, it is necessary to first recognize the normal state of the pipeline network. An approach integrating the STL decomposition with K-means algorithm is applied. After decomposing pressure difference with STL decomposition, the K-means algorithm is used to cluster the trend component to identify the leakage conditions of the WDN, and the residual component can indicate the approximate time period when the leak starts to occur. Besides, the pressure difference is defined as the difference between the predicted pressure without considering leakage and the measured pressure:

$$\Delta \boldsymbol{P} = \boldsymbol{P}_0 - \boldsymbol{P} \tag{1}$$

where $\boldsymbol{P}_0$ and $\boldsymbol{P}$= predicted and measured pressures, respectively.

By performing the real-time simulations during the possible period that the new leak occurs, differences between the predicted and measured data can be analyzed through the statistical process control (SPC) method and the exact time that the leak started can be determined. Similarly, the quantity of leakage can be estimated by analyzing the flow prediction results. The detected occurrence time and possible quantity of the leakage will then be used in the leakage localization part of MLDIF.

Isolation

For large WDNs, identifying leaks that occur at different time is a very arduous and time-consuming task. To overcome this, a method using reference leak dataset to quickly narrow the range of leak area and accurately locate the leak point is proposed. The stepwise leakage localization includes the following steps:

- **Step1 Construction of benchmark leakage dataset$\Delta \boldsymbol{P}^*$**: Hydraulic simulations are carried out for the cases that there is no leakage and that each pipe has a certain size of leakage individually, the node pressure can then be calculated in the presence and absence of leakage in each pipeline. EPANET2 [5] is employed to perform extended period simulations (EPS) in this study. The average reference vector of pressure difference caused by the leakage in any pipe can be defined by Eq. (2):

$$\Delta \boldsymbol{P}_j = \frac{1}{T}\sum_{t=1}^{T}\frac{\boldsymbol{P}_{j0}(t)-\boldsymbol{P}_j{}^f(t)}{f}, \qquad j \in [1, \quad \cdots \quad , N] \tag{2}$$

where j = pipe index, f = the magnitude of leakage, $\boldsymbol{P}_{j0}(t)$ and $\boldsymbol{P}_j{}^f(t)$= pressure before and after adding leakage in pipe j at time t ,T = the number of the simulation time steps and N = the total number of candidate leak pipes. In addition, $\Delta \boldsymbol{P}^*$ is the set of all$\Delta \boldsymbol{P}_j$.

• **Step2 Reducing searching area of possible leakage:** According to the recognition results of identification part, the predicted pressure value of the pipe network assuming that the new leakage has not occurred can be obtained by the hydraulic simulator. Then pressure difference between prediction and measurements with and without the new leak can be calculate by Eq. (1). By sorting the cosine distance (d_j^c) between $\Delta \boldsymbol{P}$ and each$\Delta \boldsymbol{P}_j$, only the pipes with the d_j^c greater than the reliable distance, d^*, are considered to have the risk of breakage and need real-time leakage simulation.

• **Step3 Localization of the leak:** Simulate real-time leakage on each pipe obtained in **step2** with the leak magnitude estimated before. Pipe leakage feature$\Delta \boldsymbol{S}_j$, correspond to pipe j for each time step can be computed as:

$$\Delta \boldsymbol{S}_j(t) = \frac{\Delta \boldsymbol{P}_j(t)}{\Delta P_{jm}(t)} \quad j \in [1, \quad \cdots \quad , N] \tag{3}$$

where $\Delta P_{jm}(t)$=the m^th^ element of vector $\Delta \boldsymbol{P}_j$ at time t and the choice of m depends on the sensitivity of each node that has been installed pressure sensors, while the real leakage characteristic $\Delta \mathbf{S}(t)$ can be defined as:

$$\Delta \mathbf{S}(t) = \frac{\Delta \mathbf{P}(t)}{\Delta \mathrm{P}_m(t)} \tag{4}$$

By calculating and ranking the Euclidean distance between $\Delta \mathrm{S}$ and the entire simulated leakage features$\Delta \boldsymbol{S}_j, j \in [1, \cdots, N]$, the pipe that is most likely to leak can be obtained finally by Eq. (5):

$$\mathrm{linkID} = \arg \min_{j \in [1,\cdots,N \;]} \left(\sum_{t=1}^{T} d_j(t)\right) \tag{5}$$

where arg min is the argument of the minimal distance evaluated and $d_j(t)$ is Euclidean distance between $\Delta \mathrm{S}$ and $\Delta \boldsymbol{S}_j$ at time t.

Inevitably, there will be multiple leak points in WDN and it is impossible to simulate all kinds of leak combinations. However, in fact, the possibility of two leakages that occur simultaneously is very small. Therefore, the identification and localization process of multiple leaks can be converted to a relatively simple problem that execute single leak localization based on the identified leaks.

RESULTS AND DISCUSSION

The proposed MLDIF is applied to the leakage identification and isolation for L-Town.

Flow forecast and Calibration

With the help of Automated Metered Readings (AMRs) and flow measurements, it become possible to predict the water consumption of each area based on the basedemand of each node. And leaks in each area can be separately identified and located then.

According to the analysis of the difference between the predicted regional water demand and the flow measurement, it can be estimated that there was almost no leakage in L-Town in early 2018. Therefore, pipe roughness is considered to be the main parameter that needs calibrated using the gradient iteration algorithm. The predicted pressure value at monitoring node is better fitted to its observed value after calibration.

Leak detection and isolation

By performing flow forecast, STL decomposition and K-mean clustering of the long-term data, taking the data in 2019 as an example, leak conditions can be divided into several clusters, as shown in **Fig.2a**. There are four possible explanations for the changes in Leak Cluster Index (LCI): repair of leak pipe; occurrence of new leaks; growth of existing leaks and stopped development of existing leaks. In addition, repeated changes in LCI can be regarded as an unstable state of WDN but leaks have not changed.

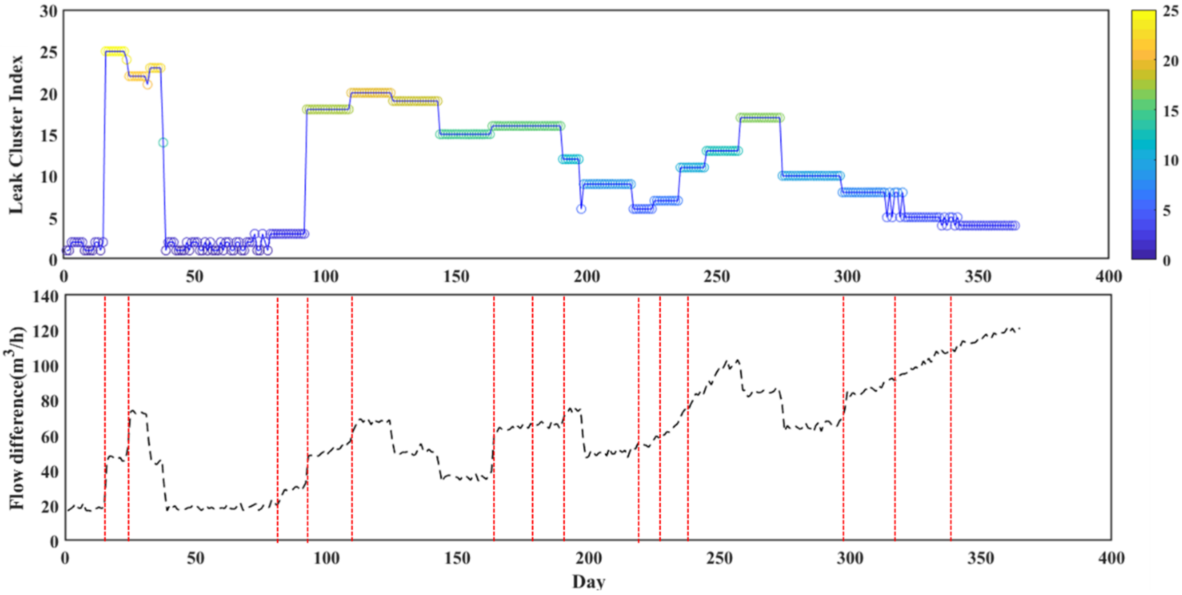

Figure 2. The classification of leak conditions (a) and final identification results of "area A" and "area B" in 2019 by analyzing flow and pressure difference (b).

After comprehensive analysis of all flow/pressure differences, start time of each leak in "area A" and "area B" in 2019 are figured out. **Fig.2b** shows the final identification results. Start time of leaks in "Area C" can be obtained in the same way.

Leaks that have been already detected and repaired in 2018 are identified and located by MLDIF in order to verify the effectiveness and efficiency of this method. It can be seen from **table 1** that all leaks can be detected in a short time. According to the regulation, one leak is considered to be located successfully if the real location of the leak appears in the top5 pipe list.

Table 1. Detection result and computation time of each leak in 2018

Real linkID	Rank	If in top5	Topological distance-to-top1	Computation time(s)
p461	1	√	0	20.40
p232	2	√	1	18.40
p673	1	√	0	27.19
p628	2	√	1	21.43
p538	1	√	0	11.36
p866	1	√	0	13.22
p31	1	√	0	22.50
p183	2	√	1	19.33
p158	2	√	1	19.26
p369	5	√	2	19.11

Although the real leak location is not always identified as the top1 of possible leak pipes, it can be found form **table 1** that the topological distance between the top1 location and the real location is very close, therefore, it is still credible to directly select the top1 location as the most likely location of leakage. After the above calculation and analysis, the accuracy, efficiency and good applicability of the proposed framework in identifying and isolating leaks can be proved.

All the leakage events that occurred after 2018 are detected and isolated by the MLDIF, including several unfixed leaks. In addition to the leaks that have been discovered and fixed by water utility of L-Town in 2018, the proposed method finds four more leaks in 2018 undetected and unfixed and 16 new leaks in 2019, including 16 leaks in "Area A", 1 leak in "Area B" and 3 leaks in "Area C", as shown in **Fig.3**. Computation time of each leak is less than 30s. Start time and pipe ID of all the detected leaks are summarized in the attachment.

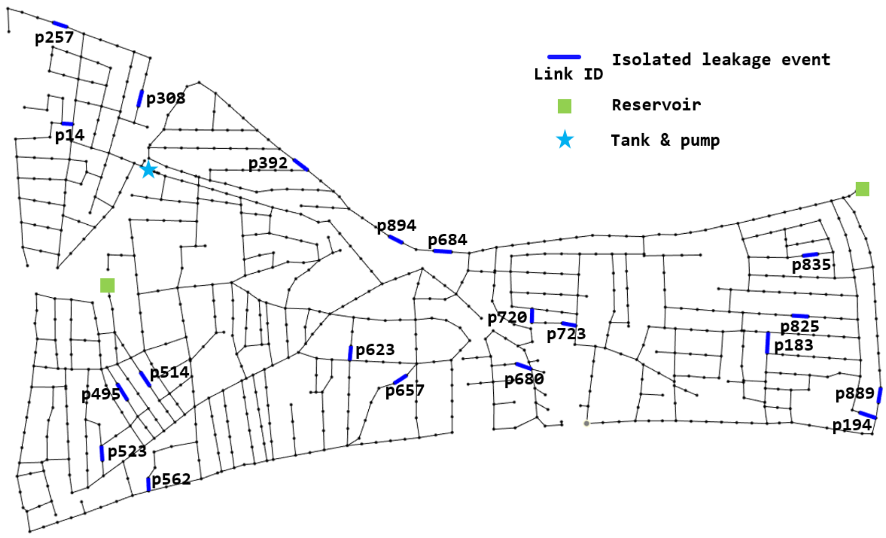

Figure 3. Locations of all detected and isolated leaks.

ACKNOWLEDGMENTS

This work was financially supported by National Key Research and Development Plan (Grant No. 2016YFC0400602) and National Natural Science Foundation of China (Grant No. 51978494，51678425).

REFERENCE

[1] Farley, B., Mounce, S. R., Boxall, J.B., 2013. Development and field validation of a burst localization methodology. J. Water Resour. Plan. Manag., 139, 604–613.
[2] Sun, C., Benjamí, P., Vicenç, P., et al. 2020. Leak localization in water distribution networks using pressure and data-driven classifier approach. Water, 12, 54.
[3] Andrew, B., Venkata, S. A., Downey, B., et al. 2017. Successive linear approximation methods for leak detection in water distribution systems. J. Water Resour. Plan. Manag., 143(8): 04017042.
[4] Casillas, M. V., Luis, E. G., Vicenç, P., et al. 2015. Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms. Water, 7, 6496-6515.
[5] Rossman, L. A. 2000. Vol. 38 of EPANET 2: Users manual. Cincinnati: Cincinnati US EPA National Risk Management Research Laboratory.

Keywords: model calibration, leakage identification, leakage localization, hydraulic model

SUMMARY

The leakage control of the water supply network is of highly concerned in the water supply industry. Leakage localization methods can guide water utilities to repair the leak and reduce water waste in time. Although methods have been reported on leak detection and isolation, there is still a lack of studies on accurate localization of multiple leaks within a water distribution network. This article presents a novel Multiple Leak Detection and Isolation Framework (MLDIF) based on existing pressure and flow measurements. This framework contains three parts: calibration of the hydraulic model, leakage detection and leakage isolation. Firstly, a gradient iteration algorithm with variable steps is used to calibrate the hydraulic model. Then the simulation data and the real pressure/flow measurement data are used to perform a joint analysis with the method of STL decomposition and K-means algorithm, to identify the changes in the network state that might indicate possible leaks. Finally, the pipe with the highest probability of leakage can be locked by ranking the matching degrees between the actual leakage feature and the simulated leakage features.

The final location result obtained by MLDIF is less sensitive to the identification error of leak quantity, and MLDIF is timesaving compared with the methods that only use machine learning, especially for complex pipeline networks with multiple leaks. By combining the MLDIF method with a hydraulic model of the system and a dataset of long-term measurements (flows and pressures) of SCADA, all the leakage events that occurred after 2018 are detected and isolated individually. The comparative analysis of the recognition results with the provided leak report in 2018 proved the accuracy, efficiency and good applicability of MLDIF in identifying and isolating leaks.