A Two-Phase Model to Detect and Localize Water Distribution System Leakages

Kyoung Won Min[1], Taewook Kim[2], Young Hwan Choi[3], Donghwi Jung[4], Joong Hoon Kim[5*]
[1,2] Graduate student, Department of Civil, Environmental and Architectural Engineering, Korea University, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea
[3] Research Professor, Research Institute for Mega Construction, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea
[4] Assistant professor, School of Civil, Environmental and Architectural Engineering, Korea University, Anam-ro 145, Seongbuk-gu, Seoul, Republic of Korea
[5] Professor, School of Civil, Environmental and Architectural Engineering, Korea University, Anam-ro 145, Seongbuk-gu, Seoul, Republic of Korea

**corresponding.jaykim@korea.ac.kr*

ABSTRACT

Water Distribution Systems (WDSs) is to supply the required quantity of customer's water demand under adequate pressure and acceptable water quality. Leakage in WDSs due to excessive pressure, pipe aging, and earthquakes leads to problems such as repair costs, disruption of water supply and economic losses. Adding the volume of water loss to customer's demand increases overall pipe flow rates and head losses throughout the system, which finally result in a low pressure at withdrawal point and the degradation of system functionality. The goal of the Battle of the Leakage Detection and Isolation Methods (BattleDIM) is to propose a method to detect and pinpoint the leakage events in L-town in 2019, as fast and accurately as possible. The SCADA measurements of flow and pressure sensor is given with the repaired date of leakages events during 2018. This study presents a new approach of two-Phase: (1) detecting the period of the individual leakage events, (2) pinpointing leak locations. In Phase 1, the data (e.g., pipe flow, tank level, nodal pressure) selected from correlation analysis is provided to the K-means clustering algorithm and Western Electric Company rules, by which normal and abnormal period of times are determined. The leakage events are assumed as the previous period for the repair completion, by which, the performance is compared with respect to the detection results of the two techniques. In Phase 2, the sensitivity analysis of applying an emitter to each node is performed in the calibrated L-town network with the pipe roughness and demand pattern. The leakage location is identified with the minimum flow variation between the 2019 SCADA measurements and the results of the calibrated network applied an emitter.

Keywords: Leakage detection, Correlation analysis, K-Means clustering, Western Electric Company, Sensitivity analysis

Introduction

Water Distribution Systems (WDSs) is to supply the required quantity of customer's water demand under adequate pressure and acceptable water quality. Leakages, generally occurred due to excessive pressure, pipe aging, and earthquakes, can cause the loss of water out of the system. Adding the volume of water loss to customer's demand increases overall pipe flow rates and head losses throughout the system, which finally result in a low pressure at withdrawal point and the degradation of system functionality.

The goal of the Battle of the Leakage Detection and Isolation Methods (BattleDIM) is to detect and pinpoint the leakage events occurred in L-Town in 2019 based on information obtained from the 2018 historical measurements of flow and pressure sensor and the repaired date. Since the characteristics of time and space are different, appropriate data mining techniques must be used.

While a clear abnormal signal of leakage is seen only at data of few sensitive pipe locations (and minor abnormalities scattered over the system), a promising strategy is to solve the detection and localization problems separately in a two-Phase scheme. In Phase 1, a methodology for detecting the period of the individual leakage events. The data is classified by the criteria such as the season and weekday/weekend. Through correlation analysis, only the critical hydraulic elements with low correlation are left. Then the leakage events are detected through K-means clustering algorithm and Western Electric Company rules for remaining critical hydraulic components. Then, the performance of the two techniques is compared for the detection results of normal and abnormal period of times. In Phase 2, a methodology for pinpointing leak locations is presented. The L-town network is calibrated with the pipe roughness and demand pattern based on the 2018 normal hydraulic measurements from Phase 1 results. The sensitivity analysis of applying an emitter to each pipe is performed in calibrated network. The leakage pinpointing is determined where is the minimum hydraulic variation between the 2019 SCADA measurements and the results from the calibrated network applied an emitter.

Methodology

The flowchart of the overall methodology is shown in Figure 1.

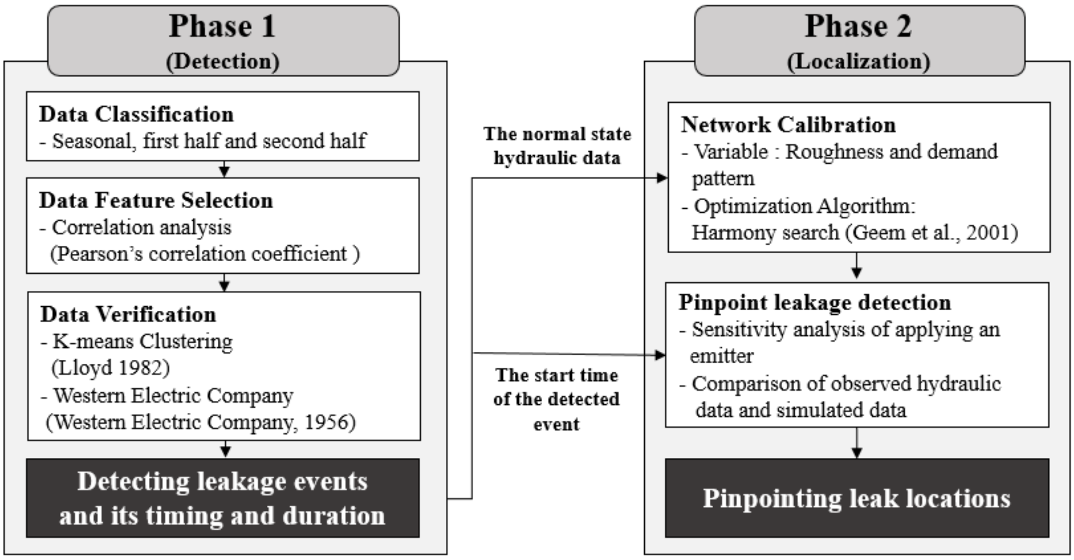

Figure 1. A Schematic of the Proposed Two-Phase Model

Phase 1 : Detecting leakage events and its timing and duration

Phase 1 presents a methodology to detect the period of the leakage. The normal states and the leakage states are detected through K-means clustering algorithm classifies two groups. Rather than analyzing a large amount of data, selecting and using the critical hydraulic components enables data noise reduction and efficient calculation. So the data classification and feature scaling are required.

Data classification

The hydraulic historical data have been collected for the period 2018-01-01 00:00 until 2018-12-31 23:55, at 5-minute time steps. The data is classified considering the season, weekday/weekend, first half/second half.

Data feature scaling

In this step, to select the critical hydraulic components, each hydraulic data (e.g., pipe flow, tank level, nodal pressure) in 5-minute time step is converted into the mean, maximum, minimum and standard deviation for each day, then the correlation analysis is performed. The correlation is calculated by Pearson's correlation coefficient as Equation 1. By this approach, the low sensitive elements are removed, leaving the high sensitive elements which are called as the critical elements in this study.

$$r_{xy} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (y_i - \bar{y})^2}} \quad (1)$$

Data verification

To detect the un-known leakage in 2019, the proposed approach needs to verification based on the observed leakage events. In this stage, the selected features as the critical elements from the previous simulations are used in the verification process by K-means clustering approach and Western Electric Company rules. The K-means clustering (Lloyd, 1982) can be categorized into the binary classes, the meaning of each category means in this study is the normal and the leakage events. The normal states or the leakage events is determined considering the repaired date. In this procedure, the occurred leakage events are assumed as the previous period for the repair completion. Western Electric Company rules (Western Electric Company, 1956) are decision rules based on the Shewart control chart. The critical hydraulic elements data under normal state according to K-means result are collected and then were calculated the 5-minute time step mean and the 5-minute time step standard deviation. Then, Western Electric Company rules were applied.

Phase 2 : Pinpointing leak locations

In Phase 2, a methodology for pinpointing leak locations is presented. First, the model of L-town network in a normal state is created. The L-town network is calibrated with pipe roughness and demand pattern by using the 2018 normal state hydraulic data obtained from Phase 1. After the model calibration, the sensitivity analysis of applying an emitter to all nodes is performed as simulation that generates leakage for each pipe. The pinpoint which has the minimum hydraulic difference between the leakage state data in 2019 and the results of the calibrated model applied to an emitter is determined as the leakage pinpoint.

Network calibration

L-town network have parameters which has difference from the actual WDSs parameters about 10%. There are several parameters to be calibrated, such as nodal demands, pipe roughness coefficient and valve setting (Lansey et. al., 1991). But among them, pipe roughness coefficient and demand pattern were used to calibrate the WDSs. Harmony Search (Geem et. al., 2001) optimization algorithm is used in the calibration process with 2018 normal state hydraulic data obtained from Phase 1. Objective function is the minimization of the difference between 2018 normal state hydraulic data and simulated data obtained from EPANET during total simulation time as an Equation 2 and the decision variables are pipe roughness and demand pattern:

$$Min\ \ f(x) = \sum_{t=1}^{T} \frac{\sqrt{(Q_{obs} - Q_{sim})^2}}{T} \quad (2)$$

where, t is time step, T is total time duration, Q_{obs} and Q_{sim} are observed and simulated flow data, respectively.

Leakage location detection

Based on the calibrated model representing behavior of L-town in normal state, the sensitivity analysis of applying an emitter coefficient to each node is performed. Sensitivity analysis shows that the change of L-town behavior, which was a normal state, can be confirmed when leakage occurs in a specific area. The flow data has changed depending on the location where the leakage occurred. This allows for obtaining simulated data when a leak occurred. The observed flow data when a leakage occurred in 2019 can be obtained from Phase 1. Root Mean Square Error (RMSE) value can be obtained by comparing all simulated data with observed data. If the error between the simulation result when the emitter coefficient is given to a specific node and the observation result is the smallest, it can be judged that a leak occurred in the pipe attached to the node.

Application and results

Detecting the period of the leakage events

In this study, for the detection the period of the individual leakage events, the data classification, feature scaling, and verification process for the 2018 hydraulic results are performed by changing the standard of the various combinations (e.g., the season, weekday/weekend, first half/second half). Moreover, the leakage occurring time can describe as the previous period of the repair completion time based on the repair date in 2018 hydraulic results. As a result of verification, the data can be classified as two parts: (1) January-March, (2) April-December and Table 1. shows the critical hydraulic components in each two parts.

Table 1. Feature scaling results

Month	Feature scaling results
Jan~Mar	**P227** (Mean, Min, Std), **Pump** (Mean, Max, Std), **Tank** (Mean, Max, Min, Std), **n1** (Mean, Max, Min, Std), **n4** (Mean, Max, Min, Std), **n215** (Mean, Std)
Apr~Dec	**P227** (Mean, Min, Std), **P235** (Mean, Min), **Pump** (Max, Min), **Tank** (Mean, Max, Min, Std), **n1** (Min), **n4** (Min), **n54** (Max, Std), **n105** (Std), **n215** (Mean, Max, Min, Std), **n296** (Max)

There are two ways to produce results, how to detect the leakage events by applying K-means clustering to critical hydraulic components in 2019 data and how to the leakage events by extracting normal state data through K-means clustering of critical hydraulic components in 2018 and then creating a WECO shewart control chart to apply in 2019. As a result, the detection performance of the former was the best with 81% and the detection performance of the latter was with 72%.

In Figure 2, the vertical black line shows the daily leakage events as K-means clustering approach based on the hydraulic results in 2018, and the detected events distinct clearly between before and after the date of repair. Moreover, in Figure 2, the flow variation of P227 and P235 shows the big difference at the leakage condition, but the difference of flow variation is decreased after the repaired date. Therefore, the residual between the two pipe flow data can be used as a leakage indicator.

The result of the leakage detection in 2019 through the values of hydraulic elements selected in 2018 shows in Figure 3. In 2019 year, the leakage occurs three times (Event 1: January 25, 2019; Event 2: March 4, 2019; Event 3: July. 25, 2019). In the second half of the year for Event 3, it can be described that it was still leaking because it was not repaired.

Pinpointing leak locations

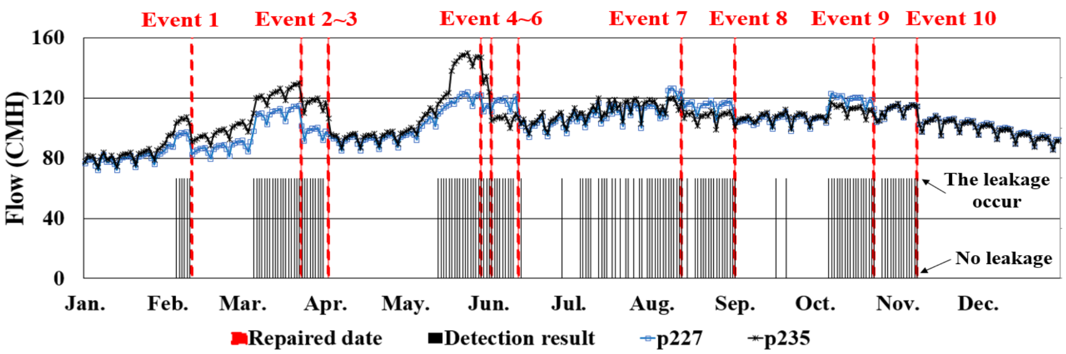

Figure 2. Data verification results from Phase 1

In this study, the leak locations are pinpointed through the model calibration with pipe roughness and demand pattern and the sensitivity analysis of applying an emitter to all nodes. The following steps are performed for pinpointing about on leak event: (1) a node is assumed to have to have a leak by assigning its emitter coefficient of 2.5, (2) a network hydraulic simulation is performed under the leak condition, (3) an error between simulated flow data and observed flow data of leakage state in 2019 is obtained, (4) then the nodal emitter coefficient is set to zero (back to normal condition) and the next node is assumed to have a leak, (5) repeat the steps (2)~(4) until all nodal leak conditions are considered. The flow values of the simulated P227 and P235 according to each nodal leak condition are different from each other. The error between the observed leakage state flow data of P227 and P235 in 2019 and the simulated flow data by the nodal leak condition are calculated in step (3). The small error between the simulated results and observed results of P227 and P235 means that a leak occurred near the node. In other words, the leak location was determined for the pipe with the smallest sum of the absolute error of P227 and P235 becomes a pinpoint. When pinpointing for one leak event is completed, the results for the next leak event are derived. The results of leakage locations are shown in Figure 3. The exact point where the leakage occurred is P629 for event 1, P34 for event 2 and P898 for event 3.

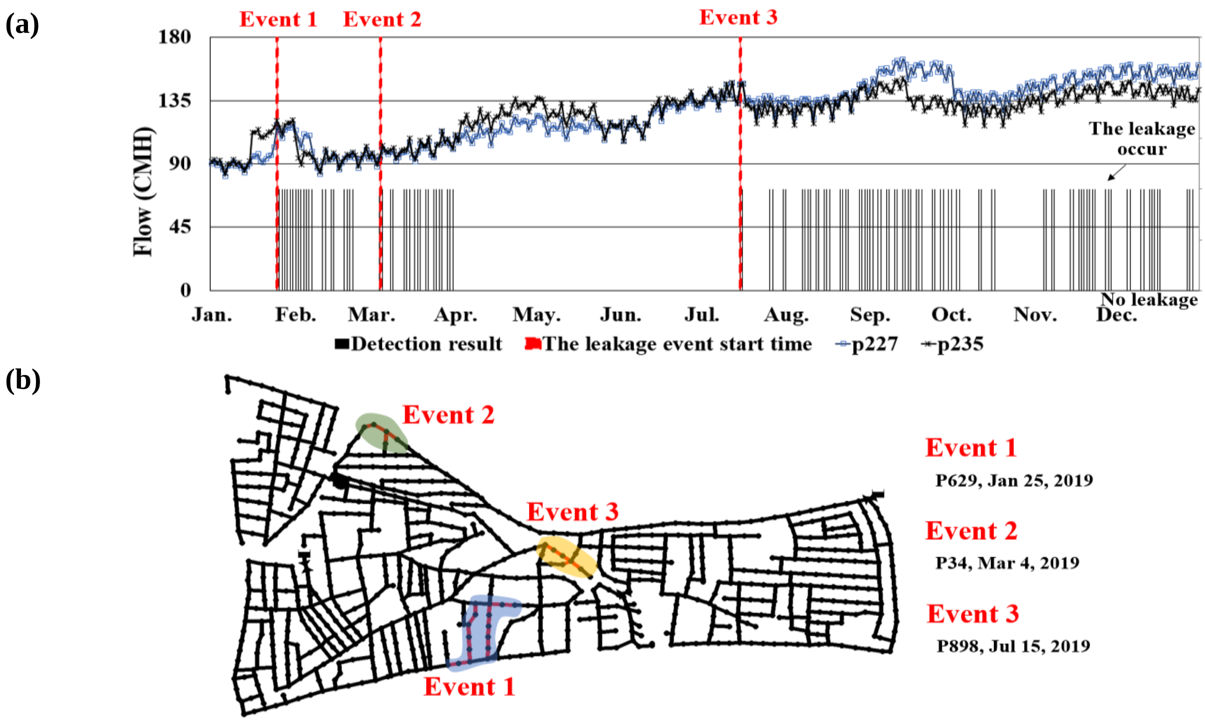

Figure 3. The leakage detection results of 2019 data (a) The detection the period of the leakage events results (b) The leakage pinpointing results

Conclusion

This paper proposes a methodology to find out where and when the leakage events occurs in WDSs. In future works, the normal state hydraulic data can be extracted through statistical techniques, unsupervised learning and the hybridization of unsupervised learning techniques and statistical techniques. In calibration process of EPANET INP file, the difference between observed flow data and simulated flow data was compared, but it could be compared with factors such as pressure and demand.

Reference

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. simulation, 76(2), 60-68.

Lansey, K. E., & Basnet, C. (1991). Parameter estimation for water distribution networks. Journal of Water Resources Planning and Management, 117(1), 126-144.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE transactions on information theory, 28(2), 129-137.

Western Electric Company. (1956). Statistical quality control handbook. Western Electric Company.

SUMMARY

Water Distribution Systems (WDSs) is to supply the required quantity of customer's water demand under adequate pressure and acceptable water quality. Leakage in WDSs due to excessive pressure, pipe aging, and earthquakes leads to problems such as repair costs, disruption of water supply and economic losses. Adding the volume of water loss to customer's demand increases overall pipe flow rates and head losses throughout the system, which finally result in a low pressure at withdrawal point and the degradation of system functionality. The goal of the Battle of the Leakage Detection and Isolation Methods (BattleDIM) is to propose a method to detect and pinpoint the leakage events in L-town in 2019, as fast and accurately as possible. The SCADA measurements of flow and pressure sensor is given with the repaired date of leakages events during 2018. This study presents a new approach of two-Phase: (1) detecting the period of the individual leakage events, (2) pinpointing leak locations. In Phase 1, the data (e.g., pipe flow, tank level, nodal pressure) selected from correlation analysis is provided to the K-means clustering algorithm and Western Electric Company rules, by which normal and abnormal period of times are determined. The leakage events are assumed as the previous period for the repair completion, by which, the performance is compared with respect to the detection results of the two techniques. In Phase 2, the sensitivity analysis of applying an emitter to each node is performed in the calibrated L-town network with the pipe roughness and demand pattern. The leakage location is identified with the minimum flow variation between the 2019 SCADA measurements and the results of the calibrated network applied an emitter.

.