

LEAK LOCALIZATION IN WATER DISTRIBUTION NETWORKS USING DATA-DRIVEN AND MODEL-BASED APPROACHES

Luis Romero¹, Joaquim Blesa^{1,2}, Débora Alves², Gabriela Cembrano¹, Vicenç Puig^{1,2}, Eric Duviella³

¹ Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain

² Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politècnica de Catalunya, Campus de Terrassa, Gaia Building, Rambla Sant Nebridi, 22, 08222 Terrassa, Barcelona, Spain

³ Informatics and Automatics Department, IMT Lille Douai, Lille, France

¹luis.romero.ben@upc.edu

ABSTRACT

The worldwide growing demand of water supply requires a proper management of the available hydraulic resources. One of the major concerns in the operation of water distribution networks (WDNs) is the existence of leakages, due to the high operational costs for the water utilities. Leaks can produce substantial economic losses, infrastructure damage and even health risks. Therefore, leak detection and isolation methodologies are widely researched.

On the one hand, model-based approaches exploit the existence of a hydraulic model of the considered WDN, as well as the availability of hydraulic measurements like inlet flow and pressure, and sensorized inner nodes pressure, to tackle the leak localization task. The suitability of these methods has been confirmed by numerous works during the years. On the other hand, the sources of information in the majority of water networks are rather limited, and other interesting measurements are not available, like water demands at the junctions, flows between inner nodes, etc. Thus, data-driven approaches, which have a reduced or non-existent dependency on a hydraulic model, can be helpful to locate leaks in WDNs that lack the mentioned measurements and modelling.

This abstract presents the combined utilization of a model-based and a novel data-driven methodology to locate leaks in the concrete case of the challenge proposed at BattLeDIM 2020. The division of the introduced network (L-Town) in three areas allows to determine the usage of one of the approaches at each one of these areas, depending on their concrete characteristics.

Besides, both methods allow to solve the multi-leak problem in a proper way, which entails a further step with regard to the classical single-leak assumption.

Keywords: leak localization, water distribution network, model-based, data-driven, multi-leak problem

INTRODUCTION

In the last ten years, our research team, with a solid experience in fault diagnosis applied to different real applications, have developed several leak localization methods that include model-based methods (Pérez et al., 2011), data-driven methods (Soldevila et al., 2020) and hybrid methods (Soldevila et al., 2016), which have successfully been applied to real WDNs. These methods considered inlet pressure and flow measurements, that are usually gathered in WDN for control purposes, and some pressure measurements in inner nodes from installed sensors that are easy and relatively cheap to install. The

L-Town challenge in BattLeDIM 2020 introduces multi-leak scenarios including abrupt and incipient leaks. Additionally, the three areas composing the network present different kind and amount of sensors. These facts have motivated the development of two new leak localization methods: A new data-driven approach that benefits from the high density of pressure sensors installed in Area A and a new model-based method that benefits from demand measurements provided by Automated Metered Readings (AMRs) in most of the nodes of Area C. The model-based method has also been applied to Area B, where there is a very low density of pressure measurements to apply the data-driven method. In the following, both leak localization methods that are able to tackle the multi-leak problem are described in more detail.

DATA-DRIVEN APPROACH

An explicative diagram of the novel data-driven leak localization procedure is displayed in Figure 1.

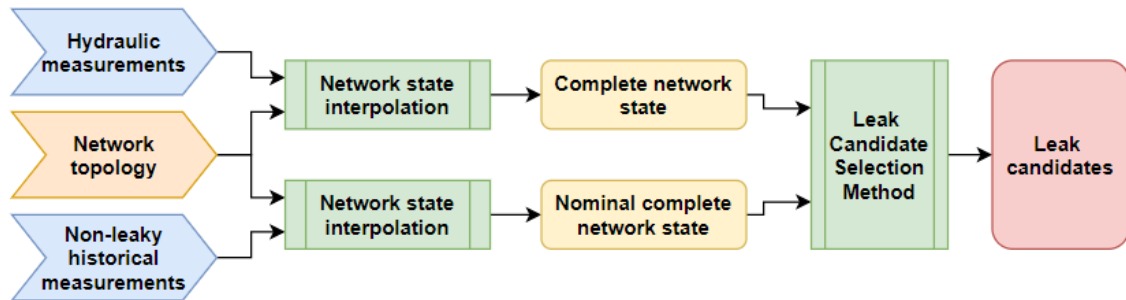


Figure 1: Workflow scheme of the complete data-driven leak localization approach.

It consists of a two-phase process:

1. From the hydraulic measurements gathered by the installed pressure sensors, as well as the topology of the WDN, the complete state of the network is estimated. The sensorized nodes state value will correspond to their exact hydraulic head, whereas the state of the rest of the nodes will consist of an approximation of the actual hydraulic head.
2. From the analysis of the current and nominal complete states, achieved from the network state interpolation of the current leaky data and historical non-leaky data respectively, a specific area of the network is considered as the source of the leak. An estimation of a more precise location of the leak is derived from the distance metric that is exploited to select the candidates.

As mentioned above, this methodology has been applied to Area A, due to the existence of 29 sensors distributed among the 659 nodes contained in the area. To face the presented challenge, the method is extended to locate pipe leaks by assuming the leak to be located in the pipe that connects the best node candidates.

1.1. Network state interpolation

The proposed state estimation method is based on two main ideas:

1. To introduce a simplification in the relation between the states of nodes of the network. The information of the measured nodes would be extended to the rest of nodes by means of a simple and linear equation. Hence, the state of the network does not represent exact hydraulic variables (except for the sensorized nodes), although it properly estimates them.
2. To exploit the graph associated to the WDN structure to perform the estimation of the states. This translates to the utilization of the nodes adjacency and the network flows directionality to induce a better approximation of the states.

Both objectives can be reached by solving the following optimization problem to obtain the network state estimator \mathbf{x} :

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2} [\mathbf{x}^T H \mathbf{x} + \tau \gamma^2] \\ \text{s. t.} \quad & B \cdot \mathbf{x} \leq \gamma \\ & S \cdot \mathbf{x} = \mathbf{x}^s \end{aligned}$$

where $H = LD^{-2}L$, and D and L are the degree matrix and the Laplacian of the graph associated to the WDN topology, respectively. In this way, the first part of the cost function introduces the simplification of the relation among nodes by means of the harmonization property, i.e., $\mathbf{x}^T H \mathbf{x} = \sum_{i=1}^n \left[x_i - \frac{1}{d_{ii}} \mathbf{w}_i \mathbf{x} \right]^2$, being n the total number of nodes. \mathbf{w}_i stands for the i^{th} row of a weight matrix W , achieved from the lengths of the pipes of the network (these distances must be inverted so that close nodes share a high weight value). The second part of the cost function, together with the inequality constraint, allow to impose a condition about the flow directionality by means of the incidence matrix B , in order to obtain more accurate state interpolations. Finally, the equality constraint sets the sensorized nodes states to their respective head value, achieved from the sensors measurements.

An example of the obtained results by means of this interpolation method is shown in Figure 2.

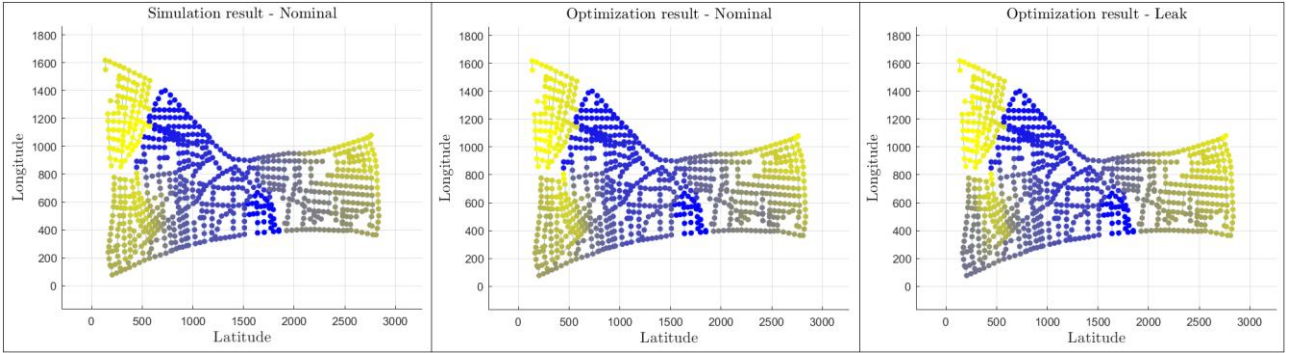


Figure 2: Comparison between the complete network state estimation, obtained for three different scenarios: a hydraulic simulation of a nominal situation, a graph-based interpolation of a nominal situation and a graph-based interpolation of a situation with a leak.

1.2. Leak candidate selection method (LCSM)

The key idea of the second part of the method consists of considering the state of a healthy situation of the network, \mathbf{x}_{nom} , and the state of the scenario with leak, \mathbf{x}_{leak} , to be the $x - y$ coordinates of n points in \mathbb{R}^2 . Then, the best fitting line for the obtained points cloud is computed, and the perpendicular distance, with sign, from each individual point to the line is computed. The importance of the sign lies in the assumption that leaky points tend to undergo a larger pressure value reduction in comparison to non-leaky nodes. This effect would be more noticeable in nodes which are close to the leaky one. The produced line directly depends on the values of all the nodes, and hence the distance between a certain node location in \mathbb{R}^2 and this line not only encodes information about the change of state in that node, but also the relation among states of nodes of the network.

Due to the large set of nodes that may fulfil the requirement of having a positive distance value, they need to be filtered. To solve this problem, the standard deviation of the distance vector is computed, using the achieved value as a threshold.

The set of final candidates for the presented example is displayed in Figure 4. Additionally, a detailed image is presented in the form of a colour map. The colour of each node depends on the distance value provided by the LCSM method: the greener the node, the higher the distance.

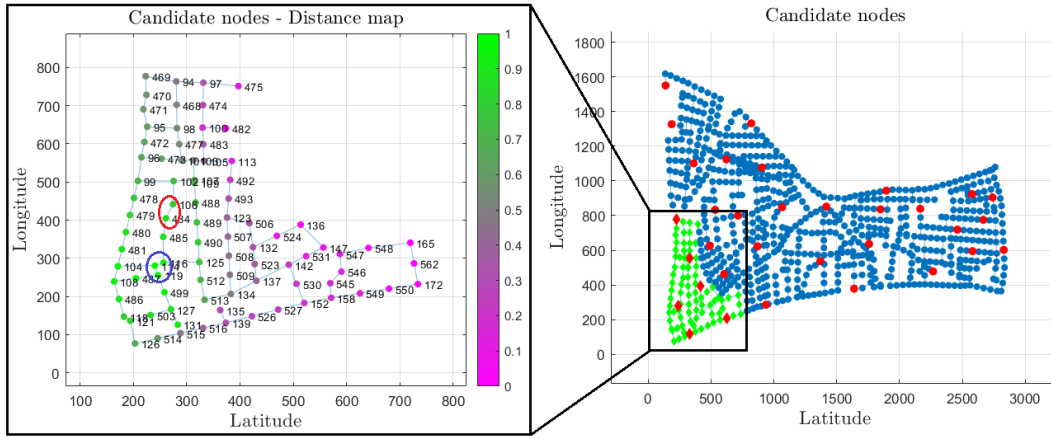


Figure 3: Representation of the set of candidates (right) together with a colour map of the distance value for each candidate (left). The red ellipse encircles the actual leaky pipe, while the blue ellipse indicates the zone with higher distances in the LCSM method.

The proximity of the red and blue ellipses illustrates the good performance of the proposed method. Nevertheless, notice that the goal of the method is to find a location area that contains the leak, but some degree of error may exist.

MODEL-BASED APPROACH

The model-based leak localization method developed and applied to Areas C and B relies on the scheme depicted in Fig. 4, based on computing pressures for different the leak scenarios (hypothesis) and matching actual pressure measurements with the most probable leak hypothesis.

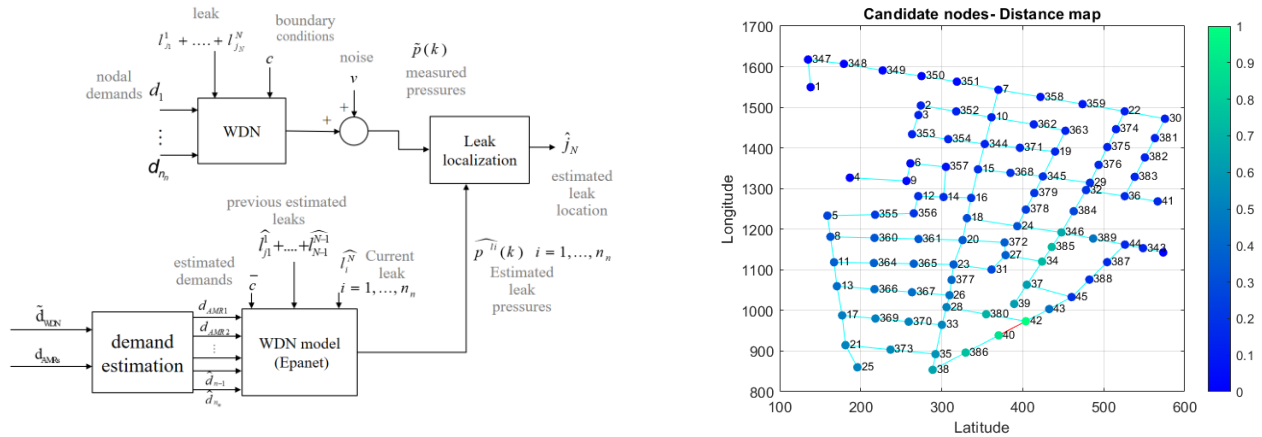


Figure 4: Workflow scheme of the model-based leak localization approach (left) and localization result for the fixed leak in Area C during 2018 (right).

Note that pressures of leak hypothesis are computed by an Epanet WDN model that considers user demands measured (if there are AMRs) and non-leak demands computed by the expected total demand distribution of the effective inlet flow (measured inlet flow minus leak estimation magnitudes) where there are not AMR measurements. In addition, the estimation of leaks produced in previous instants $l_{j1}^1 + \dots + l_{jN-1}^{N-1}$ are introduced in the nodes where the leak have been located in

previous steps and the magnitude of the leak produced at the current step l_i^N is introduced in the different possible leak places (hypothesis) $i = 1, \dots, n_n$. Therefore, a leak detection algorithm is necessary to determine at which time a leak is present in every area and to provide an estimation of its magnitude. In this work, leak detection and estimation is carried out by means the analysis of two inlet flows evolutions: the total inlet flow minus the pump flow to reservoir of Area C to detect and estimate leaks in area A and B and the output flow of reservoir of Area C compared with AMRs measurements to detect and estimate leaks in Area C.

BATTLE APPLICATION and CONCLUSIONS

This abstract presents a combined leak localization model-based/data-driven approach, applying each technique depending on the characteristics of the network. This methodology was used for the BattLeDIM 2020 challenge, and was applied, as stated, for the training and testing datasets. About the training results, some key aspects can be highlighted:

- About the data-driven method, the eight leaks of 2018 in Area A, that are said to be located and fixed by the water utility of L-TOWN, were found by the leak localization approach, as they were contained by the candidates set. There was a mean distance of 5.6 pipes from the leak to the best candidate, regarding that two of the leaks (p232 & p369) highly increase this mean value. If they are not considered, the mean is reduced to 3.5 pipes. It was discovered that this behaviour is caused by the pipe distance from the leaks to the sensors. If considering only the problematic leaks, the mean distance from the nearest sensors to the leak is 337.1 m, while this value is 158.7 m for the mean of the rest of the leaks.
- About the model-based technique, the reported leak of 2018 in Area C was detected and located at the pipe level. In addition, a non-reported leak was also detected and located. On the other hand, the leak of 2018 in Area B, was detected and located with an error of one pipe.

The testing results are included at the attached text file. The 2019 dataset implies a higher level of difficulty, due to the existence of several incipient leaks that are not fixed, which hinder the performance of the data-driven method, as well as the leak estimation for the model-based method.

The proposed approach is applicable to a broad class of water networks and leak situations, since it considers different options, e.g. for networks where no hydraulic model is available, taking advantage of a deployment of pressure sensors, or for networks with AMR and a reliable hydraulic model. Similarly, it considers single-leak and multiple-leak situations.

References

- Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., and Peralta, A. (2011). Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*, 19(10), 1157–1167.
- Soldevila, A., Blesa, J., Tornil-Sin, S., Duviella, E., Fernandez-Canti, R.M., and Puig, V. (2016). Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, 55, 162–173.
- Soldevila, A., Blesa, J., Jensen, T.N., Tornil-Sin, S., Fernandez-Canti, R.M., and Puig, V. (2020). Leak Localization Method for Water-Distribution Networks Using a Data-Driven Model and Dempster-Shafer Reasoning, in *IEEE Transactions on Control Systems Technology*, doi: 10.1109/TCST.2020.298