Computers and Chemical Engineering 108 (2018) 11–23

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

A graph partitioning algorithm for leak detection in water distribution networks

Aravind Rajeswaran[1], Sridharakumar Narasimhan*, Shankar Narasimhan

Department of Chemical Engineering, Indian Institute of Technology Madras, India

ARTICLE INFO

Article history:
Received 7 January 2017
Received in revised form 10 August 2017
Accepted 18 August 2017
Available online 30 August 2017

Keywords:
Water distribution networks
Leak detection
Graph partitioning
Multi-objective optimization

ABSTRACT

Urban water distribution networks (WDNs) are large scale complex systems with limited instrumentation. Due to aging and poor maintenance, significant loss of water can occur through leaks. We present a method for leak detection in WDNs using repeated water balance and minimal use of additional off-line flow measurements. A multi-stage graph partitioning approach is used to determine where the off-line flow measurements are to be made, with the objective of minimizing the measurement cost. The graph partitioning problem is formulated and solved as a multi-objective mixed integer linear program (MILP). We further derive an approximate method inspired by spectral graph bisection to solve the MILP, which is suitable for very large scale networks. The proposed methods are tested on large scale benchmark networks, and the results indicate that on average, flows in less than 3% of the pipes need to be measured to identify the leaky pipe or joint.

1. Introduction

It is estimated that energy generation accounts for approximately 15% of global water use, while approximately 8% of energy is used for treating and transporting water (Garcia and You, 2016). The resulting water-energy nexus is well studied and the process systems engineering community has contributed significantly in formulating and solving complex problems, e.g., in design (Rojas-Torres et al., 2014), operation (Shi and You, 2016), and monitoring for leaks (Mulholland et al., 2014; Kim et al., 2016) and contaminants (Palleti et al., 2016; Mann et al., 2012).

With the ever growing complexity and scale of water distribution networks (WDNs), leak detection is of particular importance for effective water management and quality control (Colombo and Karney, 2002; Puust et al., 2010). Leaky distribution systems are inefficient due to water loss, energy wastage, and unreliable water quality: especially in case of underground leaks. These effects are even more pronounced when the networks are poorly instrumented and maintained, such as in low and middle income countries.

In WDNs, leaks and losses are quantified using unaccounted-for water (UFW). High levels of UFW are detrimental to the financial viability of the system. These losses are a combined effect of real losses like leaks in pipes or joints, as well as other causes such as water thefts and unauthorized consumption (Gonzalez-Gomez et al., 2011). Given the growing concern towards uncertainty in quality water supplies, the problem of leak detection and control has grown in importance. Various techniques based on acoustic methods and magnetic flux leakage (Mpesha et al., 2001; Sun et al., 2011; Colombo et al., 2009) are available to determine the location of defect (either small defect like corrosion, or large leaks) in a single pipe. However, these methods can be time consuming, expensive, or disruptive in nature. Thus, it is beneficial to use these techniques only after narrowing down the leak to a small part of the network.

One approach to leak detection involves the use of hydraulic models and simulators. Available measurements are used to estimate the location of a leak which match the sensor measurements closely. This method is generally called inverse analysis (Liggett and Chen, 1994) and requires a large optimization problem to be solved. In order to use this approach, measurements of flow rates and pressures at a large number of intermediate locations are required, in addition to source pressure and demand flows. A more severe limitation of methods that make use of pressure measurements, is that the predictions depend on precise estimates of model parameters, including pipe friction factors, which are difficult to obtain. Practical applicability of this method to large scale networks have proven to be a hard task, as reported by some researchers (Stephens et al., 2004, 2005). A similar inverse problem in real time contaminant detection has been formulated that identifies the set of potential

---

* Corresponding author.

E-mail addresses: aravraj@cs.washington.edu (A. Rajeswaran), sridharkrn@iitm.ac.in (S. Narasimhan), naras@iitm.ac.in (S. Narasimhan).

[1] Currently at University of Washington, Seattle.

http://dx.doi.org/10.1016/j.compchemeng.2017.08.007

Nomenclature

m	Number of edges in **G**
n	Number of vertices in **G**
$\mathbf{u}_1, \ldots, \mathbf{u}_n$	Eigenvectors of $\mathbf{JJ}^T$
w_k	Cost of edge k
$\mathbf{x} \in \{0, 1\}^n$	Indicator variable assigning each vertex to a partition
$\mathbf{z} \in \{-1, 1\}^n$	Indicator variable assigning each vertex to a partition
A	Adjacency matrix of **G**
D	Degree matrix of **G**
E	Set of edges
G(**N**, **E**)	Graph defined on **N** and **E**
J	Directed incidence matrix of **G**
L	Laplacian matrix of **G**
N	Set of vertices
$\mathbf{S}(\mathbf{N}_S, \mathbf{E}_S)$	Subgraph formed from **G**(**N**, **E**)
cut($\mathbf{S}, \bar{\mathrm{S}}$)	The cut-set of partition ($\mathbf{S}, \bar{\mathrm{S}}$)
$R(\mathbf{S}, \bar{\mathrm{S}})$	The cut-cost of partition ($\mathbf{S}, \bar{\mathrm{S}}$)
γ	Parameter used in goal programming formulation
$\lambda_1, \ldots, \lambda_n$	Eigenvalues of $\mathbf{JJ}^T$
μ	Weighting factor used in approximation formulation

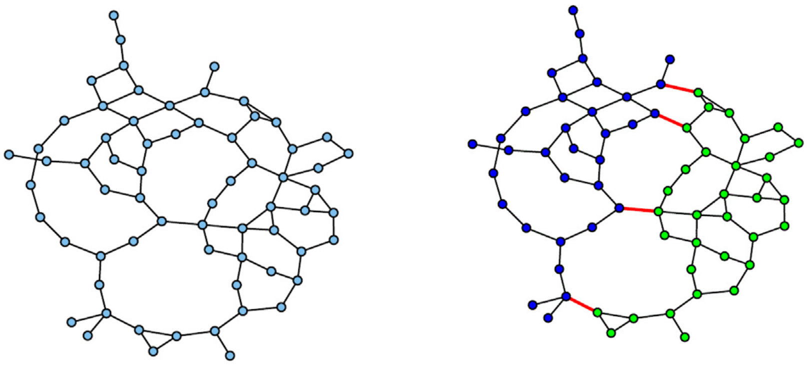
contamination locations (Mann et al., 2012). However, mixed integer linear programs (MILP) have to be solved in real time thus limiting this approach to networks with good communication and real time computational facilities.

Methods which do not use a hydraulic model explicitly in leak detection have also been proposed. One such method uses continuous pressure readings followed by filtering and statistical analysis to detect leaks, bursts and other abnormalities (Kim et al., 2016). On the other hand, Mulholland et al. (2014) describe a technique for leak detection using mass balances alone. The resulting system of equations is underdetermined and hence, a linear program with snapshots of data at different operating conditions is solved to obtain a most credible or plausible leak location. However, in the case of large networks, it is not clear if the leak can be localized to the desired degree of granularity. Furthermore, the method only exploits existing flow sensors and does not address the question of where to place additional flow sensors to address the problem of non-unique solutions and improve the leak resolution.

At present, most water networks are poorly instrumented with poor spatial and temporal resolution of data. In well instrumented networks, a small set of flow and pressure sensors are installed for the purpose of district metered area (DMA) sectorization, but these are few in number and consist of approximately 500–1000 connections. In addition, consumer consumption may be recorded over longer time scales while pressure changes are very rapid and occur over much shorter time scales. Thus, a leak if any, can be localized to a DMA or sub-network consisting of a large number of pipes or nodes, and further resolution of the leak is not trivial requiring significant time and effort.

In order to overcome the above difficulties, we propose a method for leak detection which uses only existing flow measurements, and some additional flow measurements which are repeatedly performed on-demand in field campaigns. We call this process of obtaining flow measurement in a pipe (on-demand) as **querying** the pipe for flow. Further, since the only property of leak we exploit is loss of material (water), the method is equally applicable to any form of loss including thefts. This method has the added advantage that it does not require knowledge of the pipe friction factors or pressure measurements.

To briefly illustrate the idea, consider the network shown in Fig. 1(a). Let us say that some node in this network is leaky, and our objective is to find it. By querying the red edges in Fig. 1(b), we can trace the leak to either of the two parts of the network, shown in blue and green. This is possible by exploiting water balance, or more broadly conservation laws, as will be shown in subsequent sections. By performing this operation repeatedly, we can arrive at a small part of the network which contains the leak. Querying a pipe requires access to it, which may be buried underground at a depth of about one to two meters. Hence, there is a non-negligible cost associated with every query. Therefore, it is important to minimize the number of queries, or query cost, which requires a strategic field campaign. An ideal field campaign should possess the following characteristics: (i) it should be systematic and arise out of a clear objective; (ii) it should scale to large sectors or the whole network in absence of DMAs; (iii) must be capable of assimilating information from other sources (like existing sensors); (iv) should be optimal, requiring only few queries and (v) should require minimal real time computations since communication and computational availability during a field campaign is likely to be limited. An algorithmic procedure for developing such a field campaign is the subject of this paper. We propose to solve this problem using a divide and conquer approach, where we divide the problem of leak detection using graph partitioning, and conquer the sub-problems using water balance. A high level description of the algorithm is shown in Fig. 2 which will be described in detail subsequently. |*S*| is the size of sub-network with leak, and *Th* is a pre-defined threshold. The partitioning algorithm determines which pipes to measure at each stage, such that the network is partitioned into two sub-networks around each of which a flow balance can be applied. The proposed method


(a) Network with a leaky node

(b) Querying a set of pipes

Fig. 1. Illustration of querying the edges for flows and identifying the leaky part of the network.