Active Contamination Detection in Water Distribution Systems

Stelios G. Vrachimis[1], Ron Lifshitz[2], Demetrios G. Eliades[3], Marios M. Polycarpou[4], and Avi Ostfeld[5]

[1]Ph.D. Candidate, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus; Email: vrachimis.stelios@ucy.ac.cy

[2]MSc, Faculty of Civil and Environmental Engineering, Technion Israel Institute of Technology, Haifa 32000, Israel; PH: +972-4-8292630; FAX: 972-4-8228898; E-mail: lifshitz.ron@gmail.com

[3]Research Assistant Professor, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus; Email: eldemet@ucy.ac.cy

[4]Professor, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus; Email: mpolycar@ucy.ac.cy

[5]Professor, Faculty of Civil and Environmental Engineering, Technion Israel Institute of Technology, Haifa 32000, Israel; PH: +972-4-8292782; FAX: 972-4-8228898; E-mail: ostfeld@tx.technion.ac.il

ABSTRACT

In this paper, we propose a novel methodology for altering the area monitored by water quality sensors in Water Distribution Systems (WDS) when there is suspicion of a contamination event. The proposed Active Contamination Detection (ACD) scheme manipulates WDS actuators, i.e. by closing and opening valves or by changing the set-points at pressure controlled locations, to drive flows from specific parts of the network in predetermined paths,

and enable the sensors to monitor the quality of water from previously unobserved locations. As a consequence, the monitoring coverage of the sensors is increased and some contamination events occurring within those areas can be detected. The objective is to minimize the contamination impact by detecting the contaminant as soon as possible, while also maintaining the hydraulic requirements of the system. Moreover, the methodology facilitates the isolation of the contamination propagation path and its possible source. We demonstrate the ACD scheme on two networks, analyze the results and open the discussion for further work in this area.

INTRODUCTION

Since antiquity, humans have created water distribution infrastructures to deliver quality drinking water in required amounts from water sources to consumers. These vast infrastructures were, and still are, at the core of human development in fields of health and sanitation, life quality, urban and agricultural development and possibly every other aspect of modern human environment (Gorchev and Ozolins 1984). Water distribution systems (WDS) are continuously growing bigger and more complex. The management and operation of such complex systems pose great challenges to engineers and researchers alike. A large part of the population relies on the safe and reliable operation of WDS (Krause et al. 2008; Ostfeld et al. 2008). To ensure the above, various techniques for fault prevention, detection and response have been developed and improved over the past decades. Faults in WDS may affect the hydraulics, such as the case of equipment failures and leakages, or may affect water quality such as the case of a contaminant entering the water. This work focuses on contamination faults.

Traditionally, one of the most popular approaches for contamination detection in WDS uses a set of fixed sensor locations over the WDS to monitor and alert faults (Eliades and Polycarpou 2007; Kessler et al. 1998). This type of event detection methodology is also common in other infrastructure systems such as power, transportation and communication networks. For each given WDS topology, a set of possible sensor locations is available and the problem of finding the optimal set has been investigated in depth over the past decades (Hart and Murray 2010; Ostfeld and Salomons 2004; Taormina et al. 2018; Zhao et al. 2016). The methodology of fixed sensor locations for fault detection can be improved, in terms of detectability, by choosing optimal sensor locations based on hydraulic and topological analysis of the system. Sensor placement for monitoring faults takes into consideration, mostly, four main goals: (1) maximal coverage of system components, (2) early detection of fault events, (3) deriving information on the event source and (4) minimal number of sensors for economic reasons. Since the resulted set of sensor locations applies minimal to no change on the network topology or hydraulics, it is regarded as a Passive Contamination Detection (PCD) scheme, in which no manipulation on the systems' original (pre-setup) operation condition is being made. The topology and hydraulics of the WDS are regarded as given and cannot be changed or altered for detectability purposes.

In search for improved detection methodologies, studies in which the locations of the above sensors are perfected to better achieve the four main goals of detection have been published in the past (Eliades and Polycarpou 2010; Dorini et al. 2010; Ung et al. 2017). As to date, most studies in the PCD field focus on improving technical parameter analysis using state of the art algorithms. The challenges posed in this field by the always-growing complexity of WDS calls for original, novel thinking and methodologies. As opposed to the PCD scheme described above, an Active Contamination Detection (ACD) scheme considers deliberate reconfiguration of system components and hydraulics to achieve better detectability. ACD is based on a general concept in the fault-diagnosis literature termed *active fault detection*, which states that an auxiliary input — a set of known, predetermined input procedures in known times — can be applied to a system, to improve the detection ability of a given sensor set (Campbell and Nikoukhah 2004). The auxiliary input is designed in a way such that specific faults trigger different responses from the system. Using a suitable fault-detection methodology, the different responses are identified and linked to a fault, thus improving fault detection time and fault isolation. Applying an *active fault detection* scheme to different types of systems is not straightforward and poses a challenging task. It has been successfully applied in the field of electrical engineering where it has been mainly studied in the past two decades (Campbell et al. 2002; Hood and Ji 1997).

The field of ACD in WDS lacks previous major researches, however related work that can be considered a natural precursor of this work has been done in redesigning WDS for containment of possible contaminations (Grayman et al. 2016). In this work we focus on the motivation for generating ACD methodologies and the benefits when applied. The general problem of actively detecting a suspected contamination, while minimizing the impact of that contamination and maintaining hydraulic requirements of the network, is mathematically formulated. Then, a simplified version of the ACD problem is solved using heuristic algorithms. The proposed solution is demonstrated on two benchmark networks and the results are analyzed and discussed.

MOTIVATIONAL EXAMPLE

The concept of ACD in WDS is demonstrated using a simple 6 (six) node network (Fig. 1) with arrows representing the flow direction in each of the connecting pipes between nodes. Fig. 1(A) describes the original state of the network with no sensors or hydraulic manipulation, the flow directions are as marked by the arrows. It can be observed from the flow directions that no single sensor location can cover the entire network. For example, when a sensor (S) is located in node 5 as in Fig. 1(B), node 6 is unobservable. Moving the sensor (S) to node 6 as in Fig. 1(C), results in the loss of observability at nodes 3 and 5. The covered nodes in each scenario are marked in gray. A full system coverage, in this layout and flow regime, cannot be achieved using only 1 (one) fixed sensor location. Fig. 1(D), on the other hand, shows the potential of using hydraulic manipulation in the form of pipe closure. When the connecting link between nodes 2 and 4 is closed, the flows are forced to generate a continuous path passing through all the system nodes. Therefore, if one of the nodes is infected, the contamination will surely reach the sensor at some point of time.

Moreover, if the flow regimes shown in Fig. 1(C) and (D) are applied alternately over the system operational times, a better source identification can be achieved. If a contamination is detected during Fig. 1(C) operational times, it can be located at nodes 1,2,4 or 6. If a switch to Fig. 1(D) causes a detection in the sensor, it can also be located at nodes 3 or 5.

It may be argued that redirecting suspected contaminating water from various parts of the network, may spread the contamination to more consumers. This is the case when the sensor is located as in Fig. 1(C), node 3 (three) is suspect of being contaminated and the control action of Fig. 1(D) is applied. This scenario results in nodes 3,4,5,6 being contaminated, while originally only nodes 3,5 would be contaminated, i.e. the contamination coverage has increased. However, a better metric of contamination impact is how much contaminated water is consumed, which is also a function of the detection time. If the alternative for confirming the existence of a contaminant is to perform manual sampling at node (3), a procedure which may take up to 24 hours, then the total contaminated water consumed by nodes 3,5 may exceed the total contaminated water consumed by nodes 3,4,5,6 when the contamination is detected in only a few hours by the sensor.

PROBLEM FORMULATION

The topology of a Water Distribution Network (WDN) is described by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where the nodes $\mathcal{N}$ represent water tanks or reservoirs, junctions of pipes and water demand locations, while the links $\mathcal{L}$ represent pipes and pumps. Let the subset of nodes, indicated by $\mathcal{N}_s \subset \mathcal{N}$, represent water contamination *sensor nodes* and the subset of nodes $\mathcal{N}_c \subset \mathcal{N}$ represent *nodes suspect of contamination*, of which only one can be the source of contamination while the others represent the uncertainty of the contaminant source location.

The hydraulic state of the network is described by the flow-states $\boldsymbol{q}$, where q_j is the water flow in link $j \in \mathcal{L}$, and the head-states $\boldsymbol{h}$, where h_i is the piezometric head at node $i \in \mathcal{N}$. The piezometric head (referred to as just head from now on) consists of a component analogous to the pressure p_i at node i, and of the node elevation z_i in respect to a geodesic reference. Each node i in $\mathcal{N}$ is associated with a water demand at the node location, indicated by $d_i \in \mathbb{R}^+$.

The main hydraulic requirement in a water distribution network is that all consumer demands are satisfied, which is achieved by ensuring a defined minimum pressure $p_{\min}$ at all nodes. Additionally, a maximum pressure $p_{\max}$ should be ensured to reduce the risk of pipe failures. The main hydraulic actuators in WDN are valves which can modify flow in pipes, valves which can modify pressure at nodes and pumps which add energy in the form of pressure in the network.

Let the subset of links $\mathcal{L}_v \subset \mathcal{L}$ indicate pipes that have valves which can open or close by request. The status of pipe $j \in \mathcal{L}_v$ depends on the status of the valve on this pipe, indicated by $v_j \in \{0, 1\}$. The flow q_j in pipe j is restricted to be equal to zero when the valve is closed, i.e. when $v_j = 0$, while the flow is unrestricted when $v_j = 1$. Thus, flows in the network are dependent on the input vector $\boldsymbol{v} \in \{0, 1\}^{n_v}$, where n_v is the number of valves in the network.

Additionally, let the subset of nodes $\mathcal{N}_p \subset \mathcal{N}$ be the *pressure control nodes* in the network where pressure is regulated at a specified set-point. In this work, the head h_i of node $i \in \mathcal{N}_p$ relates to the pressure set-point indicated by u_i such that $h_i = u_i + z_i$, where z_i is the elevation of the node. These nodes represent the output of Pressure Reduction Valves (PRVs) which are usually placed at the entrances of District Metered Areas (DMAs) and their output pressure can be selected. Another way the pressure control nodes are realized, is by using pumps able of *pump head control*, i.e. the pumps are equipped with pressure sensors and are able to regulate the pressure at the pump output. The pressure set-point vector $\boldsymbol{u} \in \mathbb{R}^{n_p}$, where n_p is the number of pressure control nodes, is constrained by the physical properties of the corresponding actuating device, i.e. for PRVs the maximum pressure set-point is the PRV input pressure, and for pumps it is the pump input pressure plus the maximum pressure the pump can add. We define $\boldsymbol{u}_{\min}$ and $\boldsymbol{u}_{\max}$ as the lower and upper bounds respectively of the pressure set-point vector.

The hydraulic state of a WDN is calculated using the conservation of energy and mass equations (Lansey and Mays 1999) in discrete time, with the hydraulic step Δt corresponding to the discrete time step $k \geq 0$. The pressures and flows in the network for each time step k

are calculated using the *hydraulics function* $f_h(\cdot)$ given by:

$$\begin{bmatrix} \boldsymbol{p}(k) \\ \boldsymbol{q}(k) \end{bmatrix} = f_h\left(\boldsymbol{d}(k), \boldsymbol{h}(k-1), \boldsymbol{u}(k), \boldsymbol{v}(k); \mathcal{G}\right) \tag{1}$$

This work aims to develop an algorithm which enables the water quality monitoring of the nodes in $\mathcal{N}_c$, using the available stationary sensors in $\mathcal{N}_s$ by manipulating the valve $\boldsymbol{v}$ and pressure $\boldsymbol{u}$ settings in the network. Monitoring a node in $\mathcal{N}_c$ is defined as the event when water originating from that node passes through any of the sensor nodes in $\mathcal{N}_s$ at a later time step. Specifically, let $f_{tr}(\cdot)$ be defined as the *contaminant trace function*, which tracks the spread of a contaminant from a node $i \in \mathcal{N}_c$ to other nodes in $\mathcal{N}$. The set of contaminated nodes due to node i at time step k is denoted by $\mathcal{N}_c^i(k)$ and given by:

$$\mathcal{N}_c^i(k) = f_{tr}\left(\boldsymbol{Q}(k), \boldsymbol{V}(k), i \in \mathcal{N}_c; \mathcal{G}\right), \tag{2}$$

where $\boldsymbol{Q}(k)$ is the sequence of flow vectors such that $\boldsymbol{Q}(k) = [\boldsymbol{q}(1), \boldsymbol{q}(2), ..., \boldsymbol{q}(k)]$ and $\boldsymbol{V}(k)$ is the sequence of valve control vectors such that $\boldsymbol{V}(k) = [\boldsymbol{v}(1), \boldsymbol{v}(2), ..., \boldsymbol{v}(k)]$. Finally, monitoring of a node $i \in \mathcal{N}_c$ is achieved when $\mathcal{N}_c^i(k) \cap \mathcal{N}_s \neq \emptyset$. The earliest time step at which a node $i \in \mathcal{N}_c$ has been monitored is defined as the *detection time step* and is given by $k_d^i = \min\left\{k : \mathcal{N}_c^i(k) \cap \mathcal{N}_s \neq \emptyset\right\}, \forall i \in \mathcal{N}_c$.

It is necessary to define a maximum allowable detection time step, denoted by $\bar{k_d}$. In addition, the set $\mathcal{K} = \{1, \cdots, \bar{k_d}\}$ is defined as the set of allowable time steps such that only when $k_d \in \mathcal{K}$ is a detection considered successful. The reason for defining $\bar{k_d}$ may be to ensure that a contaminant will reach a sensor location in a detectable concentration even if it decays in the water. However, it can also be determined using other time constraints defined by the utility operator. In some cases, depending on the available sensors or network topology, it may be preferable to perform manual sampling instead of redirecting the contaminant to a sensor. The decision of whether to perform manual sampling depends on the estimated time it would take to perform this procedure. This information is incorporated into the maximum detection time step $\bar{k}_d$ of the proposed ACD methodology.

The primary objective of the proposed methodology is to minimize the impact of any suspected contamination. In this work, for simplicity, the *contamination impact* refers to the volume of contaminated water consumed and is calculated using the impact formula $f_{imp}(\cdot)$. The formula can be adapted to include different aspects of the damage caused by contaminant consumption (Eliades et al. 2011). It is assumed that when detection is confirmed, preventative actions immediately take place that prevent further contamination. The contamination impact at time step k of a contaminant originating from node $i \in \mathcal{N}_c$, is denoted by $I^i(k)$ and calculated as follows:

$$I^i(k) = f_{imp}\left(k_d^i, \boldsymbol{D}(k); \mathcal{N}_c^i(k)\right) = \begin{cases} \Delta t \sum\limits_{\tau=1}^{k_d^i} \sum\limits_{j \in \mathcal{N}_c^i(\tau)} d_j(\tau),\ k_d^i \leqslant k \\ \Delta t \sum\limits_{\tau=1}^{k} \sum\limits_{j \in \mathcal{N}_c^i(\tau)} d_j(\tau),\ k_d^i > k \end{cases} \tag{3}$$

where $\boldsymbol{D}(k)$ is the sequence of demand vectors such that $\boldsymbol{D}(k) = [\boldsymbol{d}(1), \boldsymbol{d}(2), ..., \boldsymbol{d}(k)]$ and $\mathcal{N}_c^i(\tau)$ is the set of contaminated nodes in the network at time step τ due to contaminants originating from node $i \in \mathcal{N}_c$. Due to unknown source of contamination, it is necessary to calculate the impact from a contaminant originating from any of the nodes in $\mathcal{N}_c$. The final impact is the maximum of the possible impacts, calculated as follows:

$$I(k) = \max_i \left\{ f_{imp}\left(k_d^i, \boldsymbol{D}(k), \mathcal{N}_c^i(k)\right)\right\}, \forall i \in \mathcal{N}_c \tag{4}$$

As a secondary objective, the proposed methodology minimizes the valve control actions, as it may be infeasible to close/open a large number of valves, especially when these are not remotely controlled. The number of valve control actions taken at time step k can be calculated by taking the L^1-norm of the difference $|\boldsymbol{v}(k) - \boldsymbol{v}(k-1)|_1$. Let $\Delta V(k)$ indicate the total number of valve control actions taken until time step k, and calculated as follows:

$$\Delta V(k) = \sum_{\tau=1}^{k} |\boldsymbol{v}(\tau) - \boldsymbol{v}(\tau-1)|_1. \tag{5}$$

Closing valves can significantly affect the pressures in the network. The methodology should ensure that a hydraulic solution within the pre-defined pressure requirements is feasible. This can be achieved by letting the algorithm choose the set-points $\boldsymbol{u}(k)$ for pressure controlled nodes, for which the corresponding pressure constraints should also apply.

We then define the following multi-objective optimization problem:

Problem 1 *Given a water network with a set of sensor nodes $\mathcal{N}_s$ and a set of nodes suspect of contamination $\mathcal{N}_c$ at time $k = 0$, find the valve control vector sequence $\boldsymbol{V}(\bar{k_d}) = [\boldsymbol{v}(1), \boldsymbol{v}(2), ..., \boldsymbol{v}(\bar{k_d})]$ and the pressure set-point sequence $\boldsymbol{U}(\bar{k_d}) = [\boldsymbol{u}(1), \boldsymbol{u}(2), ..., \boldsymbol{u}(\bar{k_d})]$ such that:*

$$\begin{aligned}
\underset{\boldsymbol{V}(\bar{k_d}),\boldsymbol{U}(\bar{k_d})}{\operatorname{argmin}} \quad & \left\{I(\bar{k_d}), \Delta V(\bar{k_d})\right\} \\
& I(\bar{k_d}) = \max_i \left\{ f_{imp}\left(k_d, \boldsymbol{D}(\bar{k_d}), \mathcal{N}_c^i(k)\right)\right\}, \forall i \in \mathcal{N}_c \\
& k_d^i = \min\left\{k \in \mathcal{K} : \mathcal{N}_s \cap \mathcal{N}_c^i(k) \neq \emptyset\right\}, \forall i \in \mathcal{N}_c \\
s.t. \quad & \Delta V(\bar{k_d}) = \sum_{\tau=1}^{\bar{k_d}} |\boldsymbol{v}(\tau) - \boldsymbol{v}(\tau-1)|_1 \\
& \left[\boldsymbol{p}(k)^\top \ \boldsymbol{q}(k)^\top\right]^\top = f_h\left(\boldsymbol{d}(k), \boldsymbol{h}(k-1), \boldsymbol{u}(k), \boldsymbol{v}(k); \mathcal{G}\right), \forall k \in \mathcal{K} \\
& p_{\min} \leq \boldsymbol{p}(k) \leq p_{\max}, \ \forall k \in \mathcal{K} \\
& \boldsymbol{u}_{\min} \leq \boldsymbol{u}(k) \leq \boldsymbol{u}_{\max}, \ \forall k \in \mathcal{K},
\end{aligned}$$

Problem 1 is a highly complex multi-objective optimization problem. Due to this high complexity, some simplifications need to be made to solve it. In the following sections a simplified version of Problem 1 is defined and then a solution methodology using heuristic optimization algorithms is given.

PROBLEM SIMPLIFICATION

The nonlinear functions in Problem 1 require the knowledge of water demands at each time step and the initial conditions in the network in order to be calculated. The following assumption is therefore imposed:

Assumption 4.1 *Water demands $\boldsymbol{d}(k)$ in the network are known for $k \in \mathcal{K}$ and change only at the defined discrete time steps. The initial head-state conditions of the network $\boldsymbol{h}(0)$ are known.*

The valve settings $\boldsymbol{v}(k)$ are considered as binary variables. Depending on the number of valves $n_v = |\mathcal{L}_v|$, the number of combinations of this input for each time step is 2^{n_v}. Considering that valve settings can change at every time step, then for the valve control sequence $\boldsymbol{V}(\bar{k_d})$ there are $2^{n_v \times \bar{k_d}}$ possible combinations. This search space is large even for heuristic algorithms to handle and in the case of large networks a solution may never be found. The problem can be simplified by reducing the decision variables search space using the following assumption:

Assumption 4.2 *Only one control input is applied to the system, such that the valve control vector $\boldsymbol{v} = \boldsymbol{v}(1) = \boldsymbol{v}(2) = \cdots = \boldsymbol{v}(\bar{k_d})$ and the pump set-point vector $\boldsymbol{u} = \boldsymbol{u}(1) = \boldsymbol{u}(2) = \cdots = \boldsymbol{u}(\bar{k_d})$.*

The pressure set-points $\boldsymbol{u}$ are continuous variables, which are bounded due to physical and actuator limitations. Problem 1 can be further simplified by discretization of the pressure set-points using the following assumption:

Assumption 4.3 *Pressure set-points are discrete and predefined for each node $i \in \mathcal{N}_p$, such that $u_i \in \mathcal{U}$, where the set of pressure settings $\mathcal{U} = \{u_{i,\min}, \cdots, u_{i,\max}\}$ has finite elements.*

The objective function of Problem 1 can be simplified by defining a single objective which is the linear combination of impact $I(k)$ and number of valve actions $\Delta V(k)$. In order to perform this simplification, the two quantities must first be normalized and then given appropriate weights. Let $I_{max}(k)$ denote the impact upper bound at time step k which is equal to the total water consumed in the network until time k, the number of valves n_v be the maximum number of valve actions when Assumption 4.2 holds and $\beta \in [0, 1]$ be an appropriately chosen weight factor. The single-objective function cost will then be given by:

$$J(k) = (\beta/I_{max}(k))\, I(k) + ((1-\beta)/n_v)\, \mathit{\Delta} V(k). \tag{6}$$

The conversion of this multi-objective problem into a single-objective problem using the weighted sum of two criteria, has the advantage of reduced computational complexity and the disadvantage of losing some solutions which may be optimal or near-optimal, and would appear on the 2D Pareto frontier. The selection of parameter β is important, in order to extract the best solution considering the defined optimality criteria. In this work β was selected by the authors so that the impact objective receives a higher priority. The shape of the Pareto frontier, depends on the network topology and the feasibility of hydraulic solutions when the algorithm tries to construct a path from the contamination source to the sensor location. Networks with looped topology may allow a larger number of valves to be closed, thus providing more freedom in constructing paths which will result in smaller detection times and less impact. An example of the Pareto frontier created for a specific case of the Hanoi network is given in Appendix S3 of Supplemental Data.

We then define a simplified version of Problem 1 as follows:

Problem 2 *Given a water network with a set of sensor nodes $\mathcal{N}_s$ and a set of nodes suspect of contamination $\mathcal{N}_c$ at time $k=0$, find the valve control vector $\boldsymbol{v}$ and the pressure set-point vector $\boldsymbol{u}$ such that:*

$$\begin{array}{ll}
\underset{\boldsymbol{v},\boldsymbol{u}}{\operatorname{argmin}} & J(\bar{k}_d) = \left(\beta/I_{max}(\bar{k}_d)\right) I(\bar{k}_d) + ((1-\beta)/n_v)\, \mathit{\Delta} V(\bar{k}_d). \\
 & I(\bar{k}_d) = \max\limits_i \left\{ f_{imp}\left(k_d^i, \boldsymbol{D}(\bar{k}_d); \mathcal{N}_c^i(k)\right)\right\}, \forall i \in \mathcal{N}_c \\
 & k_d^i = \min\left\{k \in \mathcal{K} : \mathcal{N}_s \cap \mathcal{N}_c^i(k) \neq \emptyset\right\}, \forall i \in \mathcal{N}_c \\
s.t. & \mathit{\Delta} V(\bar{k}_d) = \left|\boldsymbol{v}(1) - \boldsymbol{v}(0)\right|_1, \\
 & \left[\boldsymbol{p}(k)^\top \; \boldsymbol{q}(k)^\top\right]^\top = f_h\left(\boldsymbol{d}(k), \boldsymbol{h}(k-1), \boldsymbol{u}, \boldsymbol{v}; \mathcal{G}\right), \forall k \in \mathcal{K} \\
 & p_{\min} \leq \boldsymbol{p}(k) \leq p_{\max}, \; \forall k \in \mathcal{K} \\
 & \boldsymbol{u} \in \mathcal{U}
\end{array} .$$

SOLUTION USING EVOLUTIONARY ALGORITHMS

In this work we propose a solution methodology which uses a Genetic Algorithm (GA) provided by MATLAB® to select the appropriate inputs $\boldsymbol{v}$ and $\boldsymbol{u}$ which minimize the objective function of Problem 2. Additionally, it uses a hydraulic simulator to implement the hydraulic solver of (1) and contaminant trace function of (2) for a given network. The solution methodology is described below.

Implementation of nonlinear functions

The EPANET water distribution network simulator libraries (Rossman 2000) are used with the EPANET-MATLAB Toolkit (Eliades et al. 2016) to implement the hydraulic solver of (1) and contaminant trace function of (2) for a given network. The contaminant trace function $f_{tr}(\cdot)$ provided by EPANET, returns the set of nodes that are affected by the contaminant, as well as the percentage of contaminated water arriving at each node. A threshold c_{thr} is defined to indicate the minimum percentage of contaminant concentration that sensors can detect.

Decision variables

The GA decision variables are the vector of valve input $\boldsymbol{v}$ and the vector of pressure set-points $\boldsymbol{u}$. These are given to the GA as a single vector $\left[\boldsymbol{v}^\top \boldsymbol{u}^\top\right]^\top$ and later used appropriately in the hydraulic simulation.

Fitness function

The GA evaluates each input using an appropriately selected fitness function. In this work, the fitness function represents the objective function of Problem 2. In addition, it includes terms that penalize solutions that do not lie in the search space defined by the problem constraints. This is a way to impose these constraints which also assists the GA to find a feasible solution.

The first term of the fitness function is the pressure constraint penalty which penalizes pressure deviations outside the pressure constraints. Given the inputs $\boldsymbol{v}$ and $\boldsymbol{u}$, a hydraulic simulation in EPANET calculates the pressures $\boldsymbol{h}(k), k \in \mathcal{K}$ in the network as in (1). The pressure penalty is then defined as:

$$P = \max_{k \in \mathcal{K}} \left(\max_{i \in \mathcal{N}} \left(p_{min} - p_i(k), 0\right) \right) + \max_{k \in \mathcal{K}} \left(\max_{i \in \mathcal{N}} \left(p_i(k) - p_{\max}, 0\right) \right) \tag{7}$$

After calculating the hydraulics of the network, the EPANET contaminant trace function calculates the set of contaminated nodes in the network $\mathcal{N}_c(k), \forall k \in \mathcal{K}$ as in (2). The contaminant detection time k_d^i for each node $i \in \mathcal{N}_c$ is then calculated as follows:

$$k_d^i = \begin{cases} \min\left\{k \in \mathcal{K} : \mathcal{N}_c^i(k) \cap \mathcal{N}_s \neq \emptyset\right\}, & \mathcal{N}_c^i(k) \cap \mathcal{N}_s \neq \emptyset \\ \bar{k_d} + 1\,, & \mathcal{N}_c^i(k) \cap \mathcal{N}_s = \emptyset \end{cases} \tag{8}$$

Using the previous results, the contamination impact $I(\bar{k_d})$ is calculated as follows:

$$I(\bar{k_d}) = \max_i \left\{ f_{imp}\left(k_d^i, \boldsymbol{D}(\bar{k_d}), \mathcal{N}_c^i(\bar{k_d})\right) \right\}, \forall i \in \mathcal{N}_c \tag{9}$$

The cost $J_{GA}(\bar{k_d})$ of the GA fitness function is then given by:

$$J_{GA}(\bar{k_d}) = \alpha P + \left(\beta / I_{max}(\bar{k_d})\right) I(\bar{k_d}) + \left((1-\beta)/n_v\right) \mathit{\Delta} V(\bar{k_d}). \tag{10}$$

where $\alpha >> \beta$ is the weight factor for the pressure constraint term.

The ‘pressure constraint penalty’ can be incorporated into a custom GA selection function, thus removing it from the cost function $J_{GA}(\bar{k_d})$. This has the added benefit of having one less weight factor to tune (α). However, the selection function should be constructed in such way that maintains the property of pointing the search direction towards solutions that do not violate pressure constraints by quantifying how ”far” the current solution is from a feasible solution. This is why the selection function should have as input, not only the current solution cost, but also the “pressure constraint penalty” as defined in (7). An indicative custom selection function example in MATLAB, which uses both the fitness function cost and the pressure constraint cost, is given in Appendix S4 of Supplemental Data.

EXAMPLE APPLICATIONS

Improving contamination monitoring example

The benchmark network "Hanoi" is used to demonstrate a solution to Problem 2 using the proposed methodology described in the previous section. The network is simulated for 24 hours, with a hydraulic time step of $\Delta t = 30$ minutes, using realistic water demand patterns. The maximum detection time step is set to $\bar{k_d} = 48$ time steps, which is equivalent to 24 hours. The network model used is available in Supplemental Data.

In the example of Fig. 2, we demonstrate a case in which a previously unmonitored node is able to be monitored using the proposed scheme. The sensor is placed at node $\mathcal{N}_s = \{30\}$ and the node suspect of contamination is $\mathcal{N}_c = \{18\}$, illustrated with blue and red colors respectively. The contamination spread is indicated with red color. Note that in the figure only an instance of the simulation is shown, as the hydraulic state of the network changes at each time step. Fig. 2 (Left) shows the maximum contaminant spread in the default PCD scheme where all the pipes are open. In this scenario, the contaminant never reaches the sensor. The application of the proposed ACD methodology on this scenario, yields a valve control vector $\boldsymbol{v}$ which closes the three pipes illustrated in Fig. 2 (Right). By closing these pipes, a flow path between the contamination source and sensor is created and sustained long enough for the contaminant to reach the sensor. The pressure control input $\boldsymbol{u}$ changes the head of the pressure controlled node from 100 (m) to 190 (m) in order to avoid negative pressures.

An example on the same network is given in Fig. 3, which illustrates a case where the proposed algorithm reduces the monitoring time of a contaminated node. The sensor node is selected to be $\mathcal{N}_s = \{30\}$ and the node suspect of contamination is $\mathcal{N}_c = \{1\}$. Fig. 3 (Left) shows that in the PCD scheme, where all pipes are open, the contaminant reaches the sensor at $k_d = 20$ time steps, or 10 hours. The application of the proposed methodology on this scenario, yields a control vector $\boldsymbol{v}$ which indicates that pipes "28" and "30" highlighted in Fig. 3 (Right) should be closed. Using this configuration the detection time is reduced to $k_d = 18$ (time steps), which is equivalent to 9 hours. The ACD scheme thus managed to reduce the contaminant detection time by one hour by closing two valves.

Multiple suspected nodes example

The ability of the algorithm to handle multiple suspected contamination nodes is demonstrated in this section with two examples on the "Hanoi" network, as described in the previous section. In the example of Fig. 4, we demonstrate the same two instances of the Hanoi network as in Fig. 2 and Fig. 3, with the difference that in these example a group of nodes is suspect of contamination instead of only one node. In the first instance (Fig. 4 left), the sensor is placed at node $\mathcal{N}_s = \{30\}$, the nodes suspect of contamination is the set $\mathcal{N}_c = \{15, 16, 17, 18, 27\}$ (orange color) and the actual contamination occurs at node $\{18\}$ (black circle). In the second instance (Fig. 4 right), the sensor is placed at node $\mathcal{N}_s = \{30\}$, the nodes suspect of contamination is the set $\mathcal{N}_c = \{1, 2, 3, 20\}$ (orange color) and the actual contamination occurs at node $\{1\}$ (black circle). The contamination spread is indicated with red color. As illustrated in Fig. 4, the algorithm calculates a flow path to the sensor which includes all suspected nodes, thus monitoring them all. Notice that the solutions differ from the examples in Fig. 2 and 3, as the algorithm tries to monitor all the suspected nodes. The contamination impact calculated is the maximum possible impact that would have occurred if any one of the suspected nodes was the source of contamination.

Transport Network example

In this case study we examine the benefits of using the ACD scheme in a network with optimally placed sensors. We compare the contamination impact and detection times in the default PCD case and when using the ACD scheme, for all possible contamination scenarios in the network, i.e. for every possible contamination source (nodes). The "Hanoi" network is used, which is a representative of a transport network with large pipes and demands. The "Hanoi" network model is available in Appendix S1 of Supplemental Data.

Contaminant sensors were optimally placed using the *S-Place* toolkit (Eliades et al. 2014). The *S-Place* toolkit places the sensors by minimizing the impact of all possible contaminations. Impact in *S-Place* is defined as the volume of contaminated water consumed in m^3 until the contamination is detected, similar to the definition in this work. The sets of 1,2 and 3 sensors are placed using exhaustive search, for a simulation period lasting 24 hours, deliberately chosen to match the maximum detection time $\bar{k}_d$. The toolkit also has the option to consider demand and parameter uncertainty during the sensor placement process, however it was not used because we assume known model and demands in this work.

Exhaustive search simulations were performed, where the following parameters were varied:

1. The number of sensors. Three placement scenarios were considered: 1, 2 and 3 optimally placed sensors, with sensors placed at nodes $\mathcal{N}_s = \{27\}$, $\mathcal{N}_s = \{11, 27\}$ and $\mathcal{N}_s = \{11, 21, 27\}$ respectively.
2. Contamination source. Contamination originating from all the nodes in the network (only one contamination node at each simulation).
3. Contamination detection scheme. Passive (PCD) or Active (ACD) contamination detection scheme is applied at each simulation.

The proposed ACD scheme manipulates valves and the reservoir head in order to drive the contamination to a sensor, while satisfying hydraulic constraints. It assumes that the contamination node is known *a priori*. The objective is to minimize the contamination impact and valve control actions. The constraints when the ACD scheme is applied are defined as follows:

1. Pressure constraints for all nodes are set to $p_{\min} = 20\,(m)$ and $p_{\max} = 150\,(m)$.
2. Concentration at sensors must be greater than $c_{thr} = 7\%$ for detection.

For the selection of minimum pressure, $p_{\min} = 20\,(m)$ was used because it is a common pressure specification of water utilities during fire flow (Ghorbanian et al. 2016). For the maximum pressure, it is considered that the Hanoi network is a transport network with large pipes which can withstand pressures up to $150\,(m)$, as the demands are DMA entrances.

The threshold c_{thr} is the minimum percentage of contaminant concentration that a sensor can detect. The maximum percentage is 100% and corresponds to the maximum concentration that same sensor can detect. The "contaminant" can be any water quality parameter of interest. In general, the choice of the threshold does not affect the operation of the algorithm and is problem-specific. Details about the reasoning behind the selection of the threshold in this work can be found in Appendix S5 of Supplemental Data.

As a technical note, the procedure of confirming the existence of a contaminant does not need additional information when specialized sensors which detect specific types of contaminants are used. However, indicator-parameter sensors (chlorine concentration, conductivity, Oxidation Reduction Potential (ORP), etc.) are commonly used in WDS and will only indicate a variation in a specific water quality parameter. An appropriate threshold is then needed, which will be calculated by an event detection algorithm in order to confirm the presence of a contaminant. When using the ACD methodology, an event detection algorithm should be able to anticipate the change in water quality due to the alteration of water flows. This is why a model-based event detection algorithm is recommended to be used with the ACD methodology which is able to incorporate into the detection logic the hydraulic changes made by the ACD methodology.

Simulation results include the Figures 5 to 7, that compare the default PCD scheme (Left) and the ACD scheme (Right) for each sensor placement case. The contamination source nodes that have been monitored are highlighted in yellow and the detection time and contamination impact are shown above and below each node respectively. Moreover, tables that compare the simulation results on this network for each sensor placement case are available in Appendix S1 of Supplemental Data. The results of this case study are discussed in the section "Discussion of Results".

District Metered Area example

In this case study, a realistic network from a large water utility in Cyprus is used. The "CY-DMA" network represents a District Metered Area of which, as opposed to the "Hanoi" network, the pipes are of smaller diameter and demands are residential consumers. Additionally it has a more complex structure, thus demonstrating the ability of the proposed methodology to perform in more realistic complex situations. The "CY-DMA" network model is available in Appendix S2 of Supplemental Data.

The scenario variations and constraints are the same as described in section "Transport Network example", with the exception of the sensor locations (selected using the *S-Place* toolkit) and the pressure constraints. Three placement scenarios are considered: 1, 2 and 3 optimally placed sensors, with sensors placed at nodes $\mathcal{N}_s = \{26\}$, $\mathcal{N}_s = \{9, 28\}$ and $\mathcal{N}_s = \{9, 28, 36\}$ respectively. Due to this being a network with residential consumers, in order to protect household piping from high pressures the maximum pressure constraint was reduced to the recommended value of $p_{\max} = 80\,(m)$ (Ghorbanian et al. 2016).

The results include the Figures 8 to 10, that compare the PCD scheme (Left) and the ACD scheme (Right) for each sensor placement case. Moreover, tables that compare the simulation results on this network for each sensor placement case are available in Appendix S2 of Supplemental Data. Discussion of the results is given in the following section.

Discussion of Results

In this section the results of "Transport Network example" and "District Metered Area example" case studies are discussed and evaluated. Specific metrics are defined and given in Table 1 to help in this process. The following metrics are defined:

- Cov: Network coverage, defined as the percentage of contamination source nodes monitored by the installed sensors in all scenarios.
- $\tilde{k}_d$: The median of detection times in the scenarios when a source node is classified as being monitored.
- $\tilde{I}$: The median contamination impact in all scenarios, defined as the contaminated water consumed.
- $\tilde{VC}$: The median number of valves closed in the scenarios when a source node is classified as being monitored when using the ACD methodology.
- $\tilde{H}$: The median head at the pressure control node in the scenarios when a source node

is classified as being monitored when using the ACD methodology.

The results on "Transport Network example" indicate a significant increase of network coverage, especially in the case of one sensor, as seen in Table 1. A notable outcome is that network coverage when using the ACD scheme with one sensor (Fig. 5 Right) is better than the case of three optimally placed sensors in the PCD scheme (Fig. 7 Left). This implies that if a water utility has information about a contamination *a priori*, e.g. from customer complaints, it is possible to confirm the existence of a contaminant using fewer sensors. Note that the sensors in this case study were optimally placed for the PCD scheme, meaning that the ACD performance can still be improved with a dedicated sensor placement methodology. In terms of simulation time, each simulation needs approximately 3 minutes to run on a personal computer with Intel Core i5-2400 CPU at 3.10GHz.

In general, a significant reduction of contamination impact is observed when the ACD scheme is used, as observed by the median Impact in Table 1. Even when both methodologies detect the contamination, significant decrease in impact is observed as seen from Figures 5 – 7. Note that the median impact is reduced by approximately ten times. Reduced impact is also accompanied by reduced detection time. All these benefits come at the cost of valve actions and increased pressure in the network, even though the hydraulic constraints are satisfied. The median number of control inputs for the Hanoi network are two valve actions per simulation, while the median pressure set-point ranges from 105 to 120(m) depending on the scenario.

The results on "District Metered Area example" again indicate an increase in coverage, especially in the case of 1 and 2 sensors, as seen in Table 1. The contamination impact when using the ACD scheme is reduced by approximately a factor of eight, similar to the transport network of "Hanoi". Valve actions are increased, with the median number of valve actions ranging from six to eight, due to this being a larger network in terms of nodes. The pressure set-points are significantly lower in this example as it is a network with residential consumers. In Appendix S6 of Supplemental Data, the distribution of pressure set-points and number of valve actions for each network-sensor case are graphically shown.

SENSITIVITY ANALYSIS

In order to assess the sensitivity of the algorithm to the detection threshold c_{thr}, additional simulations are performed on the "Hanoi" and "CY-DMA" networks, for the case where two sensors are installed, and this threshold is varied between the values of 4%, 7% and 10%. The results from these simulations indicate that the coverage is not affected by the change in threshold, while the impact is increased only marginally when the threshold increases. A table with the comparison of different thresholds is included in Appendix S5 of Supplemental Data.

In order to assess the performance of ACD methodology when there is uncertainty in water demands and model of the network, sensitivity analysis has been performed. Specifically, for each solution calculated by the ACD algorithm for a given contamination scenario (i.e. for a given network, source and sensor nodes), 1000 additional Monte Carlo Simulations (MCS) were performed where the demands and roughness coefficients of pipes were randomly varied between $\pm 2, 4, 6, 8$ and 10% of their estimated value, following a uniform distribution. The results from these simulations are given analytically in Appendix S7 of Supplemental Data and aggregated metrics for each uncertainty case are given in Table 2. The metrics used are: 1) The percentage of simulations where monitoring of the contaminant was achieved using the control settings calculated by the ACD algorithm. 2) The Mean Percentage Deviation of the contamination Impact mean calculated in the MCS, related to the Impact calculated by the ACD algorithm, defined as follows:

$$I_{MPMD} = \frac{1}{n_n} \sum_{i=1}^{n_n} 100 \left(I_i^{ACD} - \mu \left(I_i^{MCS} \right) / I_i^{ACD} \right) \quad (11)$$

where $\mu \left(I_i^{MCS} \right)$ is the mean Impact from the MCS for contamination scenario i and I_i^{ACD} is the Impact calculated by the ACD algorithm for scenario i. 3) The Mean Percentage Standard Deviation of the Impact calculated by the MCS for each scenario, related to the mean Impact of each scenario, defined as:

$$I_{MPSD} = \frac{1}{n_n} \sum_{i=1}^{n_n} 100 \left(\sigma \left(I_i^{MCS} \right) / \mu \left(I_i^{MCS} \right) \right) \tag{12}$$

where $\sigma \left(I_i^{MCS} \right)$ is the standard deviation of Impact from the MCS of scenario i.

The results show that *the existence* of the solution provided by the ACD algorithm is robust to demand and model uncertainty, as the suspected nodes are monitored in at least 99.6% of the varied scenarios. The variation of the contamination impact under these uncertainties, as shown in Table 2, increases analogously to the increase of uncertainty. It is observed that the Impact variation may be network dependent, as it is of smaller magnitude in the Hanoi network which has less nodes compared to the CY-DMA network.

CONCLUSIONS AND FUTURE WORK

In this work we address the problem of water contamination detection in Water Distribution Networks (WDN) by applying a methodology of Active Contamination Detection (ACD), which manipulates the system inputs in a way that assists the contamination detection process. A generalized formulation of this problem is provided. Hydraulic constraints are incorporated into the problem, thus providing solutions that maintain the system in operation. A maximum detection time limit takes into account the scenario when manual sampling is preferable to be performed. Due to the complexity of this problem, a simplified version of this problem is solved using evolutionary algorithms. Simulations are performed on networks where sensors are optimally placed for Passive Contamination Detection (PCD). The results show that it is possible to significantly reduce contamination impact, which in this work is defined as the contaminated water consumed, by manipulating the system valves and pressure set-points. Moreover, the system coverage is improved and monitoring time is decreased.

The intended use of the proposed methodology is to provide an additional tool at the disposal of water managers that can be used in conjunction with conventional water quality monitoring. A water utility may have a few specialized contamination sensors within a system (which may be of high cost), in parallel to a larger number of general water quality sensors (e.g. chlorine, ORP etc). A suspicion of a contamination may then arise due to abnormal readings of a general water quality sensor and confirmed by driving the contamination to a specialized sensor using the ACD methodology. When using general water quality sensors to identify abnormalities in water quality, measurements from more than one sensor and knowledge of water flows in the network can be combined using a suitable contamination detection methodology (Eliades and Polycarpou 2012) to pinpoint a suspected contamination location. This location may not necessarily have a sensor installed.

Another use example of the proposed methodology is the case of a sensor fault which has as a result a substantial reduction in coverage. The sensor-fault can be obvious, such as the sensor giving a constant zero reading, thus the water utility will know not to rely on the this sensor. The ACD methodology can then be used to provide additional coverage with less sensors, as seen from the simulation results in this work. The impact of the contamination when using ACD is lower than when using PCD with the same number of sensors because the methodology tries to minimize this metric.

The ACD methodology does not require operable valves at all the pipes of the network, as it is able to work given a set of existent valves in the network. The calculated solution depends on the number of valves, location of valves and topology of the network. Current water networks may lack the Industrial Control System (ICS) infrastructure to remotely control valves and apply timely emergency response. However, the development trend of modern water distribution systems is towards remote sensing and control. If a water utility wants to employ such a system specifically for the use of the ACD methodology, optimal valve placement oriented towards increasing the efficiency of ACD can significantly reduce the valves needed and provide sufficient coverage.

When using the proposed methodology, there is a trade-off between increased monitoring capability and system resilience. A way to mitigate this problem is to include in the optimization procedure a metric of system resilience and take that into account when choosing a solution. Additionally, this methodology requires the knowledge of a set of nodes which are suspect of contamination in order to be applied. Sensitivity analysis with respect to water demand and model parameter variations, show that the existence of a solution to the ACD problem is robust to these uncertainties, whereas the impact value is more sensitive. In future work, these uncertainties can be taken into account in the Problem Formulation in order for the methodology to provide a more robust solution. This work could also benefit from a study which assesses the performance of this methodology on networks with different characteristics, because the existence of a solution that does not violate pressure constraints greatly depends on the networks topology. Other extensions of this work include further investigation of the weight β in the optimization cost function, modifications to redirect flows to flushing locations when the presence of a contaminant is confirmed and modification to focus on creating isolated paths between locations suspect of contamination and sensor nodes, in order to improve the localization of a contamination source.

DATA AVAILABILITY

The following data, models, or code generated or used during the study are available in a repository or online. [https://doi.org/10.5281/zenodo.2566001]

ACKNOWLEDGMENTS

This research is partially funded by the European Union Horizon 2020 programme under Grant Agreement No. 739551 (KIOS CoE), by the Interreg V-A Greece-Cyprus 2014-2020 programme, cofinanced by the European Union (ERDF) and National Funds of Greece and Cyprus under project SmartWater2020, by the Cyprus Research Promotion foundation programme Restart2010-2020 under Grant Agreement Enterprises/0916/0023 (WaterAnalytics), and supported by the United States - Binational Science Foundation (BSF) and by the Technion Funds for Security research.

REFERENCES

Campbell, S., Horton, K., and Nikoukhah, R. (2002). "Auxiliary signal design for rapid multi-model identification using optimization." *Automatica*, 38(8), 1313–1325.

Campbell, S. L. and Nikoukhah, R. (2004). *Auxiliary Signal Design for Failure Detection*, Vol. 11. Princeton University Press.

Dorini, G., Jonkergouw, P., Kapelan, Z., and Savic, D. (2010). "SLOTS: Effective algorithm for sensor placement in water distribution systems." *Journal of Water Resources Planning and Management*, 136(6), 620–628.

Eliades, D., Kyriakou, M., and Polycarpou, M. (2014). "Sensor placement in water distribution systems using the S-PLACE toolkit." *In Proc. of 12th International Conference on Computing and Control for the Water Industry, CCWI2013*, 602–611.

Eliades, D. G., Kyriakou, M., Vrachimis, S., and Polycarpou, M. M. (2016). "EPANET-MATLAB Toolkit: An open-source software for interfacing EPANET with MATLAB." *Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI)*, The Netherlands, 8.

Eliades, D. G. and Polycarpou, M. M. (2007). "Multi-objective optimization of water quality sensor placement in drinking water distribution networks." *In Proc. of European Control Conference (ECC)*, IEEE, 1626–1633.

Eliades, D. G. and Polycarpou, M. M. (2010). "A fault diagnosis and security framework for water systems." *IEEE Transactions on Control Systems Technology*, 18(6), 1254–1265.

Eliades, D. G. and Polycarpou, M. M. (2012). "Water Contamination Impact Evaluation and Source-Area Isolation Using Decision Trees." *Journal of Water Resources Planning and Management*, 138(5), 562–570.

Eliades, D. G., Polycarpou, M. M., and Charalambous, B. (2011). "A security-oriented manual quality sampling methodology for water systems." *Water Resources Management*, 25(4), 1219–1228.

Ghorbanian, V., Karney, B., and Guo, Y. (2016). "Pressure Standards in Water Distribution Systems: Reflection on Current Practice with Consideration of Some Unresolved Issues."

Journal of Water Resources Planning and Management, 142(8), 04016023.

Gorchev, H. G. and Ozolins, G. (1984). *WHO guidelines for drinking-water quality*, Vol. 38.

Grayman, W. M., Murray, R., Savic, D. A., and Farmani, R. (2016). "Redesign of Water Distribution Systems for Passive Containment of Contamination." *Journal - American Water Works Association*, 108(7), E381–E391.

Hart, W. E. and Murray, R. (2010). "Review of sensor placement strategies for contamination warning systems in drinking water distribution systems." *Journal of Water Resources Planning and Management*, 136(6), 611–619.

Hood, C. and Ji, C. (1997). "Proactive network-fault detection." *IEEE Transactions on Reliability*, 46(3), 333–341.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks." *Journal of Water Resources Planning and Management*, 124(4), 192–198.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., and Faloutsos, C. (2008). "Efficient sensor placement optimization for securing large water distribution networks." *Journal of Water Resources Planning and Management*, 134(6), 516–526.

Lansey, K. and Mays, L. W. (1999). *Hydraulic Design Handbook*. McGraw-Hill Education.

Ostfeld, A. and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *Journal of Water Resources Planning and Management*, 130(5), 377–385.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J.-P., Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S.-T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y., and Walski, T. (2008). "The battle of the water sensor networks (bwsn): A design challenge for engineers and algorithms." *Journal of Water Resources Planning and Management*, 134(6), 556–568.

Rossman, L. (2000). “EPANET 2: Users Manual.” *Report No. EPA/600/R-00/057*, US Environmental Protection Agency. Office of Research and Development. National Risk Management Research Laboratory.

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., Ostfeld, A., Eliades, D. G., Aghashahi, M., Sundararajan, R., Pourahmadi, M., Banks, M. K., Brentan, B. M., Campbell, E., Lima, G., Manzi, D., Ayala-Cabrera, D., Herrera, M., Montalvo, I., Izquierdo, J., Luvizotto, E., Chandy, S. E., Rasekh, A., Barker, Z. A., Campbell, B., Shafiee, M. E., Giacomoni, M., Gatsis, N., Taha, A., Abokifa, A. A., Haddad, K., Lo, C. S., Biswas, P., Pasha, M. F. K., Kc, B., Somasundaram, S. L., Housh, M., and Ohar, Z. (2018). “Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks.” *Journal of Water Resources Planning and Management*, 144(8), 04018048.

Ung, H., Piller, O., Gilbert, D., and Mortazavi, I. (2017). “Accurate and optimal sensor placement for source identification of water distribution networks.” *Journal of Water Resources Planning and Management*, 143(8), 04017032.

Zhao, Y., Schwartz, R., Salomons, E., Ostfeld, A., and Poor, H. V. (2016). “New formulation and optimization methods for water sensor placement.” *Environmental Modelling & Software*, 76, 128–136.

List of Tables

	Cov (%)		$\tilde{k}_d$ (hours)		$\tilde{I}$ (m^3)		$\tilde{VC}$	$\tilde{H}$ (m)
	PCD	ACD	PCD	ACD	PCD	ACD	ACD	ACD
Hanoi								
1 Sensor	40.6	75	6.5	2.5	11398	1642	2	120
2 Sensors	65.6	78.1	5.5	2.5	4962	1289	2	105
3 Sensors	68.8	81.3	4	2	3807	956	2	115
CY-DMA								
1 Sensor	40.7	74.7	4	1.5	91	3	8	40
2 Sensors	61.5	79.1	4	1	24	3	6	40
3 Sensors	76.9	80.2	4	1	16	2	6	40

Cov:	Percentage of nodes monitored by sensors.
$\tilde{k}_d$:	Median contamination detection time in hours.
$\tilde{I}$:	Median contamination impact in m^3.
$\tilde{VC}$:	Median number of valves closed by the ACD scheme.
$\tilde{H}$:	Median head at pressure control nodes in meters.

TABLE 1. Result metrics from simulation scenarios using Passive (PCD) and Active (ACD) contamination detection schemes on Hanoi and CY-DMA networks with 1 to 3 sensors.

	Hanoi			CY-DMA		
Uncertainty (%)	Solved (%)	I-MPMD (%)	I-MPSD (%)	Solved (%)	I-MPSD (%)	I-MPSD (%)
2	100.00	3.95	5.35	100.00	11.41	13.10
4	100.00	6.22	7.61	100.00	17.83	18.80
6	100.00	7.65	8.51	100.00	22.23	21.79
8	99.92	9.18	8.83	100.00	26.22	23.48
10	99.61	9.68	9.19	100.00	29.73	24.40

Uncertainty:	Variation of combined water demand and roughness in the MCS as a percentage of their known values
Solved:	Percentage of MCS where monitoring of the contaminant was achieved using the settings calculated by the ACD algorithm
I-MPMD	Mean Percentage Deviation of the mean Impact from the MCS from the Impact calculated by the ACD algorithm
I-MPSD	Mean Percentage Standard Deviation of the Impact from the MCS

TABLE 2. Sensitivity analysis results of the ACD methodology applied to the Hanoi and CY-DMA networks

List of Figures

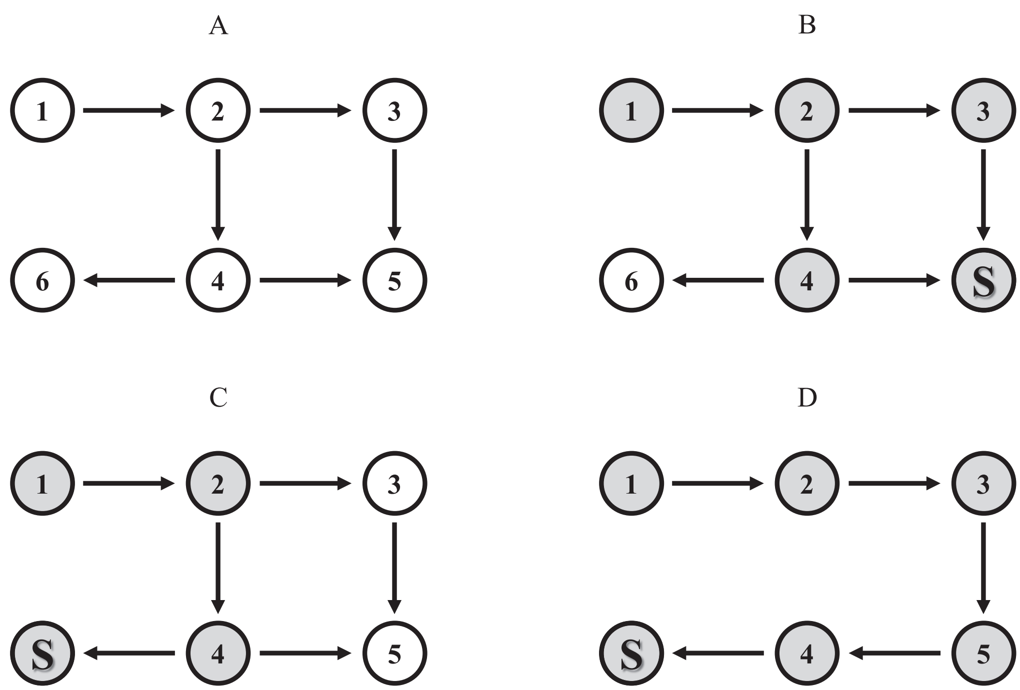

FIG. 1. Six node system. Arrows represent flow directions, S represents fixed sensor location and gray marks covered nodes.

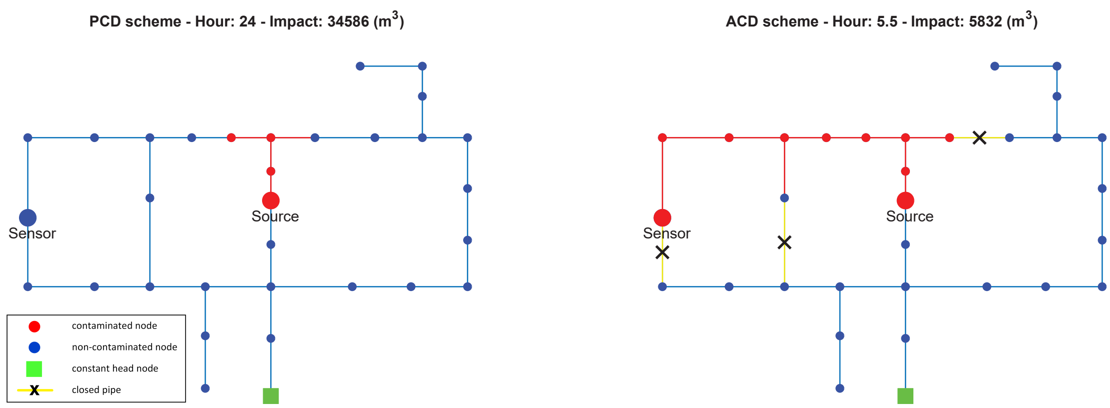

FIG. 2. Example of monitoring a previously unmonitored node using the proposed methodology on the Hanoi network.

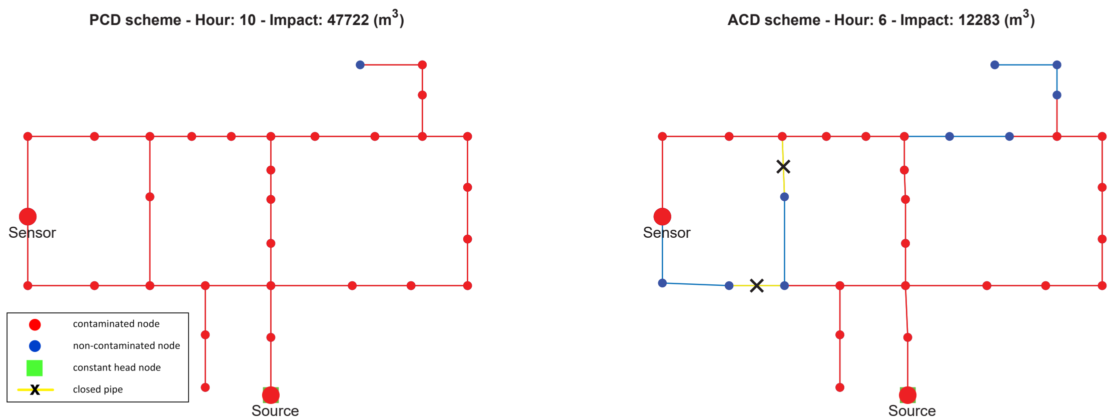

FIG. 3. Example of monitoring time reduction using the proposed methodology on the Hanoi network.

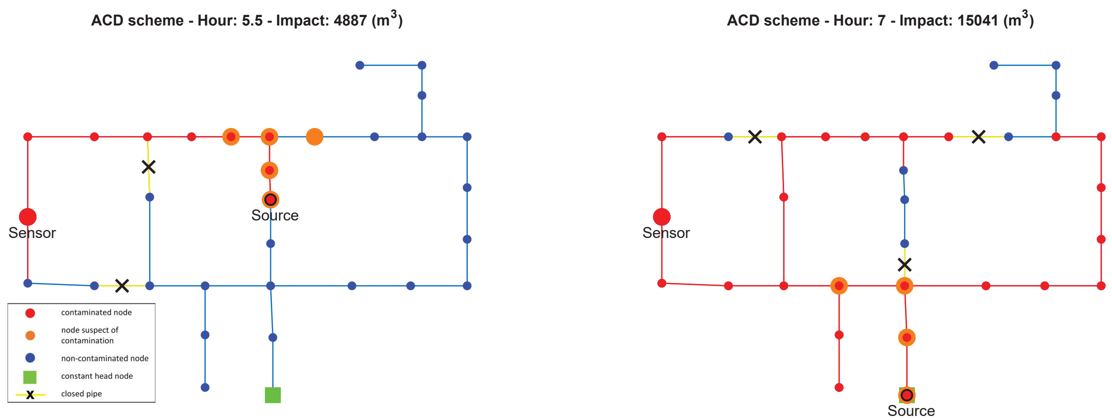

FIG. 4. The examples of Fig. 2 (left) and Fig. 3 (right) with multiple nodes being suspect of contamination and the solution using the proposed ACD methodology.

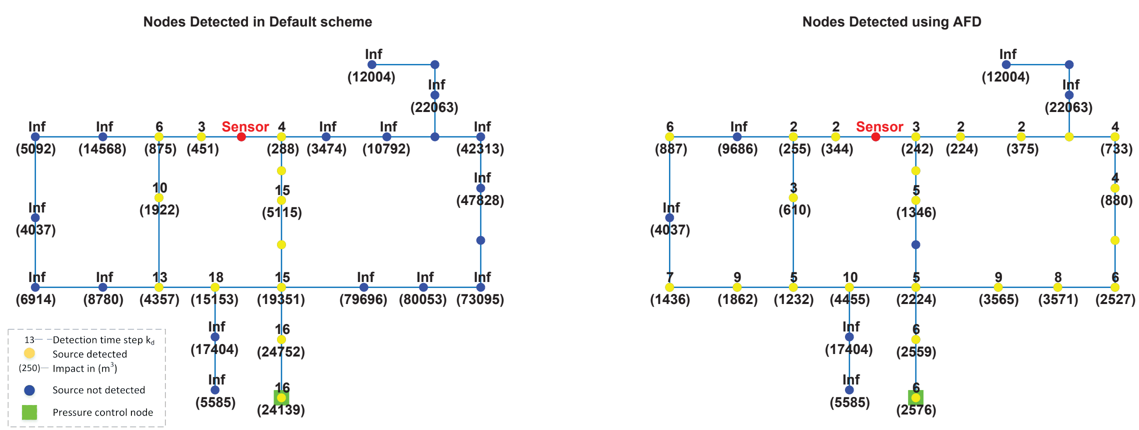

FIG. 5. Hanoi network with 1 optimally placed sensor. Left: Contaminant nodes and detection time (impact) for default PCD case. Right: Contaminant nodes and detection time (impact) with ACD scheme.

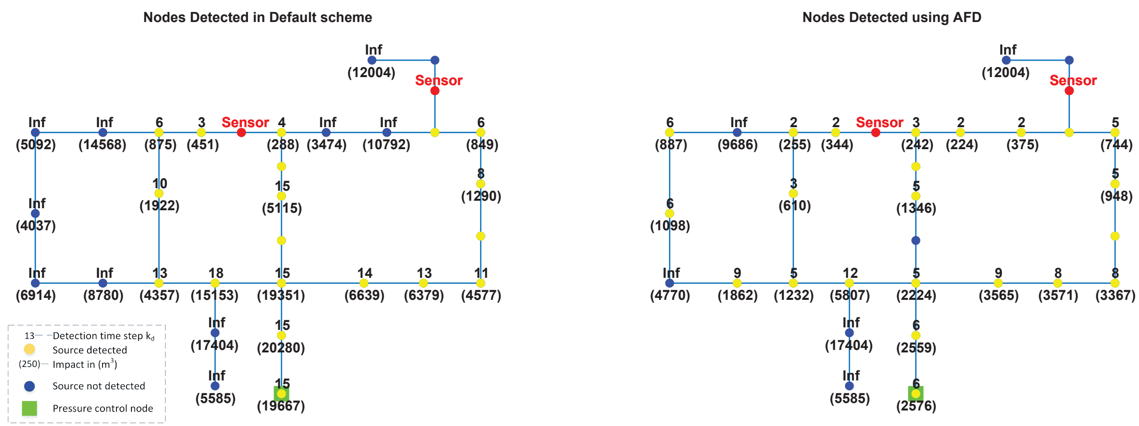

FIG. 6. Hanoi network with 2 optimally placed sensors. Left: Contaminant nodes and detection time (impact) for default PCD case. Right: Contaminant nodes and detection time (impact) with ACD scheme.

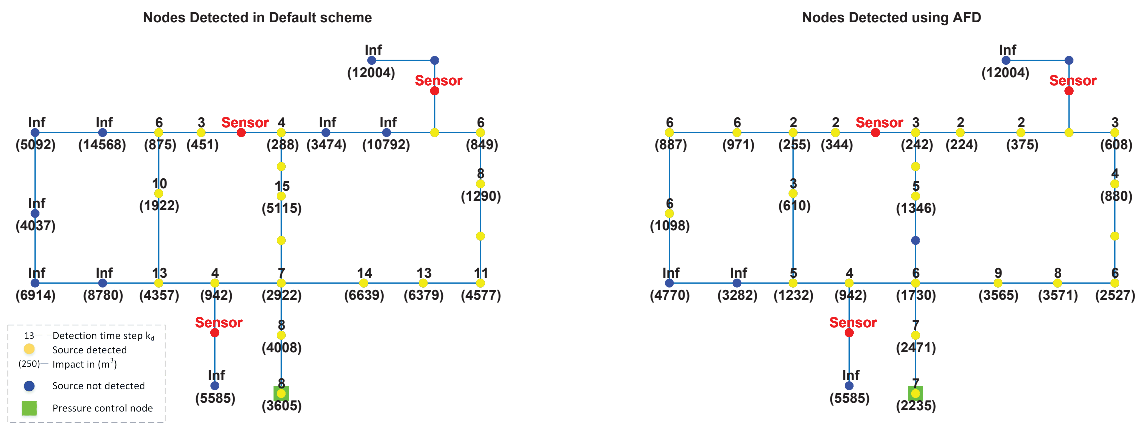

FIG. 7. Hanoi network with 3 optimally placed sensors. Left: Contaminant nodes and detection time (impact) for default PCD case. Right: Contaminant nodes and detection time (impact) with ACD scheme.

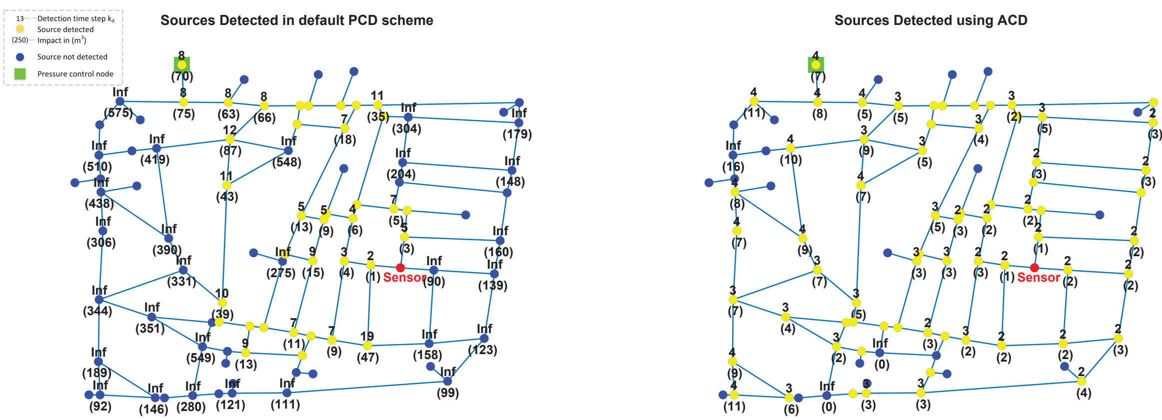

FIG. 8. CY-DMA network with 1 optimally placed sensor. Left: Contaminant nodes and detection time (impact) for default PCD case. Right: Contaminant nodes and detection time (impact) with ACD scheme.

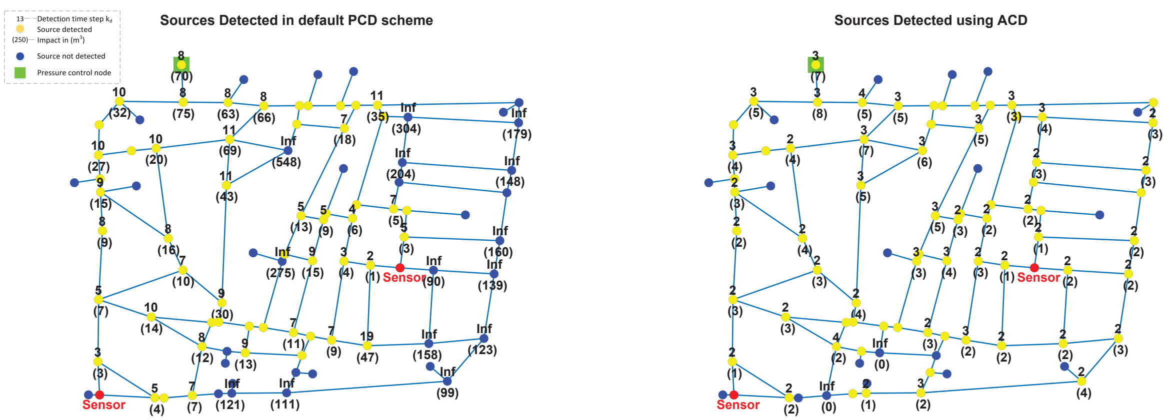

FIG. 9. CY-DMA network with 2 optimally placed sensors. Left: Contaminant nodes and detection time (impact) for default PCD case. Right: Contaminant nodes and detection time (impact) with ACD scheme.

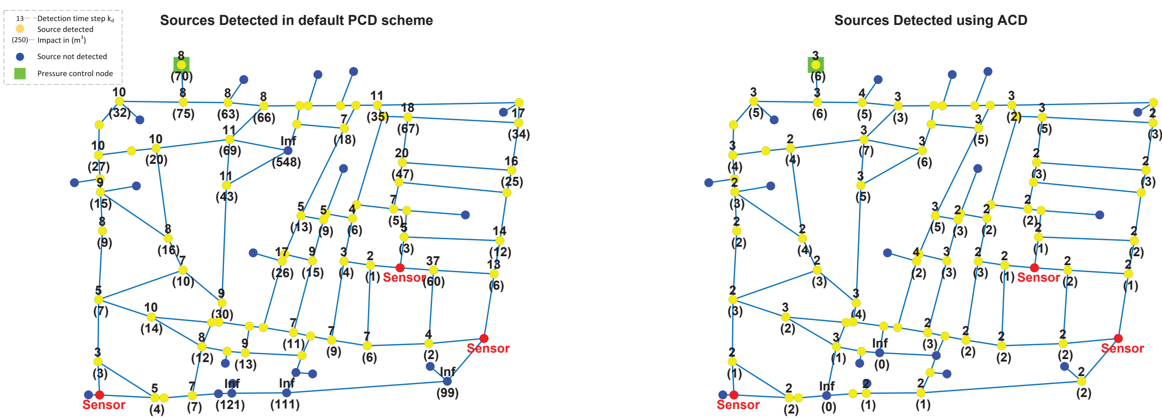

FIG. 10. CY-DMA network with 3 optimally placed sensors. Left: Contaminant nodes and detection time (impact) for default PCD case. Right: Contaminant nodes and detection time (impact) with ACD scheme.