International Conference on Hydroinformatics

2014

Decision-Making Frameworks For Using Sensor Data And Evolutionary Algorithms To Flush A Contaminated Water Distribution System

M. Ehsan Shafiee

Emily Zechman Berglund

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

DECISION-MAKING FRAMEWORKS FOR USING SENSOR DATA AND EVOLUTIONARY ALGORITHMS TO FLUSH A CONTAMINATED WATER DISTRIBUTION SYSTEM

M. EHSAN SHAFIEE (1), EMILY ZECHMAN BERGLUND (2)
(1): Postdoctoral Researcher, Dept. of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC, 27695, USA, E-mail: mshafie@ncsu.edu
(2): Assistant Professor, Dept. of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC 27695, USA. E-mail: emily_berglund@ncsu.edu

ABSTRACT

In the event that a contaminant enters a water distribution system, opening hydrants to flush contaminated water can protect consumers from becoming exposed. Strategies for operating hydrants can be developed to specify the selection of hydrants and the timing of operations to maintain a minimum water quality for every demand nodes in the network or maximize the amount of contaminant that is removed from the network. As an event unfolds, however, sensor data may be the only information that is available to indicate the location and timing of the contaminant source, and ultimately, hydrant strategies must be selected in a highly uncertain environment. The decision-making framework for making real-time decisions to select hydrant strategies relies on computational and sensor technologies, including the accuracy and precision of sensor data; the timeliness of data availability (e.g., streaming data or data that is collected manually); and computational capabilities to execute search simulation-optimization frameworks in real-time. This research will explore a decision-making framework to provide a library of response options that can be selected based on sensor data as an event unfolds. The library of hydrant strategies is developed *a priori* using a simulation-optimization framework. Potential sources are classified based on the order of sensors that are activated, and hydrant strategies are identified to maximize average performance for events within each class through the application of a genetic algorithm framework. The decision-making frameworks are applied and compared for a set of events that are simulated for two networks: the virtual city of Mesopolis and the town of Cary.

INTRODUCTION

Water distribution networks are vulnerable infrastructures to chemical contaminants [1] and bacterial outbreaks [2]. Public health is threatened when a contaminant propagates in the water network and reaches a segment of the population. To protect consumers, water utility managers can harden water networks by installing sensors, develop preparedness plans to implement response actions as contamination events unfold, and develop models to evaluate the effectiveness of response actions to prevent or reduce public health consequences [3]. Response actions prescribe decisions that can be selected to manage contamination plume movement in the network. For example, utility managers may open hydrants and manipulate valves to confine contaminant or alert consumers about an event to reduce consumption and exposure.

Flushing contaminated water by opening hydrants is typically easy to implement and is one of the least expensive techniques to maintain water quality in a network [4]. To remove contaminated water from a distribution system, a hydrant strategy can be identified to specify the timing and location for opening hydrants. Hydrants should be selected in proximity of the contaminant plume to improve the effectiveness of hydrants in removing the contaminant. A model of the water network and the contaminant source can be used to test hydrant flushing strategies and ensure that effective strategies are selected. The location of the contaminant plume can be discerned with information about the source of the contaminant, which is not typically available during an event. Warnings from water quality sensors may be the only information that is available to indicate the movement of the contaminant plume. Using sensor information to determine the contaminant source can lead to uncertainty in the source location, timing, and loading, due to limitations in the number of sensors and amount of data [5]. Uncertainty in locating the contaminant source leads to difficulties in identifying hydrants for flushing in a real-time decision-making approach. Therefore, new methodologies are needed to provide guidance and assist decision-makers in the selection of hydrant strategies in a highly uncertain environment.

A new simulation-optimization approach is developed in this research to identify a set of hydrant strategies that can be used as guidance for real-time management of a water distribution contamination event. The approach follows a set of steps: a Monte Carlo Simulation method generates a set of contamination events with diverse characteristics; a classification method is applied to group contamination events based on the order of activated sensors; and a population-based algorithm is applied to identify hydrant strategies for each class of event. Two optimization models are presented here to represent the problem of hydrant strategy identification, based on the total mass of contaminant that is removed from the system and the concentration of contaminant at each node. Finally, a decision tree is constructed to provide response actions for sensor activations. The approach is implemented and demonstrated for two case studies. The city of Mesopolis is a virtual city of 150,000 residents, and the Town of Cary, North Carolina, provides water to 150,000 residents.

DEVELOPING HYDRANT STRATEGIES FOR WATER CONTAMINATION EVENTS

Hydrant flushing is used to maintain high water quality in water networks during normal and emergency conditions [6]. Methodologies have been developed to identify hydrant strategies using water quality sensor data. Approaches have been developed that could identify hydrant strategies in a real-time manner for a water contamination event. For example, a source identification method [8] can be applied to use sensor data to characterize a water event, and an optimization algorithm can be applied to identify a hydrant strategy for source characteristics [7]. The approaches available in the literature are limited, however, in real application. The source identification approach is not able to precisely characterize a water event due to the complexity of the water network. In addition, both source identification and hydrant strategy identification approaches consume significant computational time and can create a delay in responding, which can lead to undesirable public health consequences.

This research develops a methodology that uses imprecise sensor information to facilitate the decision-making process during a real-time water event. The methodological components are (1) a Monte Carlo simulation approach is used to generate contamination events; (2) events are classified based on the order of sensors that are activated; (3) two optimization problems are formulated to represent the problem of hydrant strategy identification; and (4) a NGA-based search is used to identify a hydrant strategy for each class of contamination events. Each optimization problem is solved to produce a hydrant strategy for each event class. Steps are described as follows.

Monte Carlo Simulation approach to generate contamination events

The Monte Carlo Simulation approach is a computational method that randomly samples from a given domain of variables. A water contamination event is characterized using a set of variables – type of contaminant (bacterial or chemical), entry point of contaminant, contaminant load, start time of injection, and duration of injection [9]. A probability distribution is assigned for each variable, based on the hydraulics of the water network and the type of contaminant.

Classification of contamination events using sensor information

Each contamination event that is generated using Monte Carlo Simulation is simulated using a water distribution system model EPANET [10]. An ensemble of sensors is modeled to detect a contaminant, and each contamination event is grouped in a class based on the activated sensors and the order of activation. In this study, the first two sensors that are activated are used to classify events. The total number of event classes that can be generated using two sensors out of a total of *m* sensors is equal to $\mathrm{m} + \binom{\mathrm{m}}{2}!$

Optimization models: maximum mass and reliability

Identification of a hydrant strategy for one class of events is formalized as an optimization model. Two optimization models are described here.

The first model is the maximum mass model. The objective statement maximizes the total mass of contaminant that is removed from the network, represented by Eqns. 1-5. The model should be solved separately for each event class.

$$\underset{\boldsymbol{H},\boldsymbol{T},\boldsymbol{D}}{\text{Maximize}}\; c = \frac{\sum_{i=1}^{i=N} \begin{cases} f_{EPANET,e_i}(\boldsymbol{H},\boldsymbol{T},\boldsymbol{D}), & p_{v_k} \geq p_{min}\, psi, \forall v_k \in \boldsymbol{V}, k=1,\dots,n_v \\ 0, & \text{otherwise} \end{cases}}{N} \tag{1}$$

$$e_i \in \boldsymbol{E}, i = 1, \dots, N \tag{2}$$

$$h_j \in \boldsymbol{Hyd}, h_j \in \boldsymbol{H}, j = 1, \dots, n \tag{3}$$

$$0 \leq t_j \leq T_{max}, t_j \in \boldsymbol{T}, j = 1, \dots, n \tag{4}$$

$$1 \leq d_j \leq D_{max}, d_j \in \boldsymbol{D}, j = 1, \dots, n, \tag{5}$$

where c is the aggregated amount of contaminant that is flushed from the network over all hydrant flows and is calculated as the average for a set of N contamination events. The contamination events, e_i, should be selected from the list $\boldsymbol{E}$, which is the list of all events in one class. The maximum value for N is the size of the list $\boldsymbol{E}$. The solution to the optimization model represents a hydrant strategy and includes $\boldsymbol{H}$, $\boldsymbol{T}$, and $\boldsymbol{D}$, which are lists of hydrants, delays, and durations respectively. The list of hydrants, $\boldsymbol{H} = \{h_1, h_2, \dots, h_n\}$, is a list of integers, which represents the indices of nodes. Each value for h_j is selected from a limited list of nodes in the network that are hydrants, $\boldsymbol{Hyd}$. For each hydrant that is selected, t_j and d_j are the decision variables that represent the time at which the hydrant should be opened, measured as the delay after the sensor is activated, and the duration over which the hydrant should remain open, respectively. The number of hydrants that are used to create a hydrant strategy, *n*, is specified by the user *a priori*. The hydrant strategy is used as input for the function evaluation, f_{EPANET,e_i}, which is the hydraulic and water quality model for simulating a contamination event, e_i. The hydraulic simulator calculates both the amount of contaminant removed for each event, and the water pressure, p_{v_k}, at a terminal node, v_k. Terminal nodes are identified as those nodes with a non-zero water demand. If water pressure drops below p_{min} pounds per square inch, which is the minimum pressure that must be maintained in the network for firefighting emergencies, at any terminal node, the performance of the hydrant strategy is penalized, and the amount of contaminant that is removed is set to zero for the event, e_i.

The second formulation of the hydrant strategy problem is the reliability model (Eqns. 6-10). The model should be solved to minimize the concentration of contaminant that is present at every node in the water network during the simulated time. The reliability model should be solved for each event class separately.

$$\underset{\boldsymbol{H},\boldsymbol{T},\boldsymbol{D}}{\text{Minimize}}\; c_n = \frac{\sum_{i=1}^{i=N}\sum_{t=1}^{t=T}\sum_{j=1}^{j=n_v}\begin{cases} f_{EPANET,e_i}(\boldsymbol{H},\boldsymbol{T},\boldsymbol{D}), & p_{v_k} \geq p_{min}\,psi, \forall v_k \in \boldsymbol{V}, k=1,\dots,n_v \\ +\infty, & \text{otherwise} \end{cases}}{N \times T} \tag{6}$$

$$e_i \in \boldsymbol{E}, i = 1, \dots, N \tag{7}$$

$$h_j \in \boldsymbol{Hyd}, h_j \in \boldsymbol{H}, j = 1, \dots, n \tag{8}$$

$$0 \leq t_j \leq T_{max}, t_j \in \boldsymbol{T}, j = 1, \dots, n \tag{9}$$

$$1 \leq d_j \leq D_{max}, d_j \in \boldsymbol{D}, j = 1, \dots, n \tag{10}$$

where c_n is the amount of contaminant that is present at the demand nodes in the network during the simulation time T and is calculated as the average for a set of N contamination events. Since the $f_{EPANET,e_i}(\boldsymbol{H},\boldsymbol{T},\boldsymbol{D})$ is not the function of t and j, EPANET is only executed once for each e_i and the contamination concentration is integrated at each demand node over simulation time subsequently.

Noisy Genetic algorithm to identify a hydrant strategy

The optimization models represented by Eqns. 1-5 and Eqns. 6-10 are nonlinear problems, due to the characteristics of hydraulic calculations for a looped network. The genetic algorithm [11] is a population-based search algorithm that has been applied successfully for a range of complex water management problems, including water distribution management problems [12]. To use a genetic algorithm-based approach, the value of *N* in Eqns. 1 and 6 may be set equal to the number of events in a class, and the genetic algorithm can be used to maximize the performance over all events in a class. This approach, however, is computationally impractical, because this would require a high number of simulation evaluations to evaluate each solution, or hydrant strategy, and a genetic algorithm uses several thousands of solution evaluations to converge to a final solution. NGA is developed and applied here to identify a solution that can perform well for all events in a class, by sampling a representative set of events for each solution evaluation. NGA [13] follows the algorithmic steps of a genetic algorithm, with the exception of the fitness function evaluation. NGA evaluates the objective function based on a number of realizations of the uncertain variables for each solution and can be implemented by using a Monte Carlo sampling mechanism as part of a solution's evaluation. The average of the sampled fitness values is assigned as the fitness of a solution. In this study, the list of contamination events is sampled by randomly selecting N contamination events from the list. For each selected event, the performance of a hydrant strategy is evaluated using an EPANET simulation, and the average across several events is assigned as the fitness. The same amount of contaminant is introduced for each event, and each event is given equal weight in calculating the fitness function.

CASE STUDIES: A VIRTUAL WATER NETWORK AND A REAL-WORLD NETWORK

The simulation-optimization framework is applied to identify hydrant strategies for Mesopolis, a virtual city. The Mesopolis dataset was developed as a case study for research in threat management for urban infrastructure. Mesopolis is simulated with diverse land uses comprised of residential, commercial, and industrial areas, and within the city limits, there is a naval base, an airport, and a university (Fig. 1a). Water is withdrawn at an intake located south of the city from a river that runs north through the center of Mesopolis. A branched pipe delivers raw water to two water treatment plants (WTP), located on opposite sides of the river. The West WTP supplies water to the older sections of Mesopolis, located on the western side of the river. The East WTP distributes water to the eastern section and, during peak demand periods, to a large portion of the central and western districts. The network is modeled as a skeletonized water network with one reservoir, 1588 nodes (706 of these are terminal nodes), 2058 pipes, 13 tanks, and 65 pumps. Four demand patterns are applied for different nodes based on residential, commercial, industrial, and naval land uses. Three sensors are placed in the network to detect contaminant based on insight about flow directions and hydraulic zones that govern contaminant transport in Mesopolis.

The Cary water network is a realistic pipe system that is larger than Mesopolis, even though it serves a similar number of consumers (Fig. 1b). The water network has three reservoirs, nine tanks, 26,986 nodes, 28,331 pipes, and 20 pumps. The north reservoir withdraws water from a lake and provides more than 95% of water to the city after it is treated at a water treatment plant. Demand nodes are initialized with 13 diverse demand patterns. The water pipe is modeled using a high resolution, and end-use connections are modeled. As a result, the model is computationally more expensive than the Mesopolis water network. Thirteen nodes are used as the sensor locations for Cary water network. The Cary water network provides a realistic case study to demonstrate the performance of the hydrant strategy approach to protect public health.

The simulation-optimization model will be applied for both water networks to develop three types of decision trees for three types of management approaches: risky, risk-averse, and adaptive. A decision-maker may take a risky management approach and choose between strategies to implement a hydrant strategy immediately or wait to receive additional information from water quality sensors to refine management actions. A risk-averse approach implements a hydrant strategy that is designed after the warning from the first sensor is received. Risk-averse decision trees are designed for a decision-maker who prefers to respond to an event quickly. Finally, an adaptive management strategy implements hydrant strategies in response to each sensor activation. The simulation-optimization approach is applied for event classes to create decision trees that can be used during an event to respond to information from sensors.

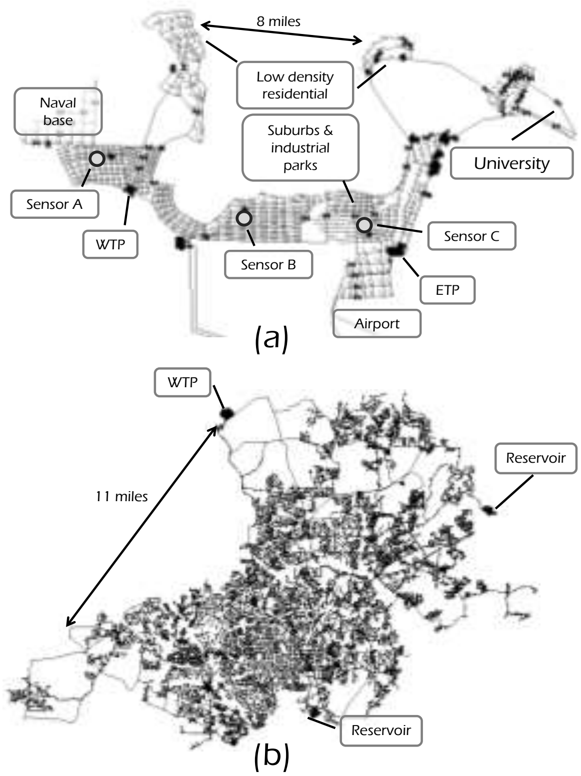

Figure 1. (a) Mesopolis water distribution network, land uses, and sensor network. (b) Cary water distribution network.

REFERENCES

[1] Greick P. “Water and terrorism" *Water Policy*, 8(6), 481-503, (2006).

[2] Hrudey S. and Hrudey E. “*Safe Drinking Water. Lessons from Recent Outbreaks in Affluent Nations*”, International Water Association, (2004).

[3] United State Environmental Protection Agency “*Response protocol toolbox: Planning for and responding to drinking water contamination threats and incidents overview and application*" EPA report, Environmental Protection Agency, Washington, D.C., EPA report 817-03-007, (2003).

[4] Ellison D., Duranceau, S. A. S. D. G. and McCoy, R. “*Investigation of pipe cleaning methods*" AWWA Research Foundation, Denver, (2003).

[5] Zechman, E. M. and Ranjithan, S. R. “Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system" *Water Resources Planning and Management*, 135(5), 334-343 (2009).

[6] Friedman, M., Kirmeyer, G. J., and Antoun, E. “Developing and implementing a distribution system flushing program." *American Water Works Association Journal*, 94(7), 48, (2002).

[7] Laird, C., Biegler, L., van Bloemen, B., and Bartlett, R. “Contamination source determination for water networks." Journal of Water Resources Planning and Management, 131(2), 125-134, (2005).

[8] Guan, J., Aral, M., Maslia, M., and Grayman, W. “Identification of contaminant sources in water distribution systems using simulation optimization method: Case study." Journal of Water Resources Planning and Management, 132(4), 252-262 (2006).

[9] Rasekh, A. and Brumbelow, K. “Probabilistic Analysis and Optimization to Characterize Critical Water Distribution System Contamination Scenarios", *Journal of Water Resources Planning and Management*, 139 (2), 191-199, (2013).

[10] Rossman, L. “*EPANET user's manual*" U.S. Environmental Protection Agency Risk Reduction Engineering Lab, (2000).

[11] Goldberg, D. E. “Genetic Algorithms in Search, Optimization, and Machine Learning”, Addison-Wesley, (1989).

[12] Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B.,Ostfeld, A., Singh, A., and Zechman, E. “State of the art for genetic algorithms and beyond in water resources planning and management", *Journal of Water Resources Planning and Management*, 136(4), 412-432, (2010).

[13] Smalley, B., Minsker, B., and Goldberg, D. “Risk-based in situ bioremediation design using a noisy genetic algorithm", Water Resources Research, 36(10), 2043- 2052, (2000).