

Evaluation of Non-Uniqueness in Contaminant Source Characterization Based on Sensors with Event Detection Methods

Jitendra Kumar[1], E. M. Zechman[1], E. D. Brill[1], G. Mahinthakumar[1], S. Ranjithan[1], J. Uber[2]
[1]North Carolina State University
Raleigh, North Carolina
{jkumar, emzechma, brill, gmkumar, ranji}@ncsu.edu
[2]University of Cincinnati
Cinicinnati, OH
jim.uber@uc.edu

Abstract

Due to the present state of sensor technology, during a water distribution contamination event, sensors may be able to detect only the presence of a contaminant and not necessarily the complete concentration profile. Some sensors trigger a detection based on a specified threshold concentration of observation, yielding a binary detection/no-detection signal. Event detection can also be based on observed concentrations of water quality parameters, such as pH and chlorine, which are routinely measured. These concentration observations are then processed through event detection algorithms to yield a detection/no-detection signal. These event detection techniques filter the measured concentrations at sensors to produce a discrete signal. When using this filtered information to characterize the contamination source, the certainty of identifying a unique solution is likely reduced, i.e., a set of widely different source characteristics may provide a match for the sensor observations. The authors previously presented an evolutionary algorithm-based procedure for source characterization and for assessing non-uniqueness by generating a set of maximally different alternatives. The procedure is extended here to characterize a contaminant source and any non-uniqueness arising by using sensor information processed through different event detection methods.

Introduction

In the wake of recent terrorist attacks, there has been a growing concern over the safety and security of water distribution networks. A pollutant introduced into the water distribution system will spread through the system and can adversely impact public health. Development of an accurate contaminant warning system is critical for initiating a response as information about the location, character and the extent of the pollutant is required before regulatory measures can be taken. USEPA (2005) has emphasized the need of a robust and comprehensive drinking water monitoring program that would provide early indication of an attack and minimize the public health consequences. It also helps specify the desired characteristics of an Early Warning System. The goal of an Early Warning Monitoring System is to reliably identify low probability/high impact contamination events (chemical, microbial, radioactive) in source water or distribution systems in time to allow an effective local response that reduces or avoids entirely the adverse impacts that may results from the event (ILSI, 1999).

Several research studies focusing on the detection of a contaminant source in a network during a contamination event have been reported in literature. Source detection has been approached as an inverse problem by various researchers. For example, Waanders et al. (2003) used a direct sequential technique to solve a small-scale optimization problem with a standard successive quadratic programming approach. Laird et al. (2005b) reported an origin-tracking algorithm based on a Lagrangian non-linear programming approach. Laird et al. (2005a) proposed a mixed-integer programming based method for the identification of unique injection scenarios. Guan et al. (2006) used a simulation-optimization approach using a reduced gradient method to minimize the sum of squares of prediction errors at the monitoring locations.

The reported investigations of procedures for source identification assume that contaminant-specific probes exist to continuously monitor for those contaminants in the distribution system. While this assumption enables the investigation of source identification procedures, it is highly unlikely to find in reality such probes. Also, this assumption implies that one knows a priori the types of contaminants that would be used in an intentional attack. Even when contaminant-specific probes are used, the detection may not always yield a continuous concentration value. Depending on the degree of sophistication or resolution of the sensor, the sensor output could range from a threshold-based binary response (e.g., contaminant is detected, or not detected, above a pre-specified contaminant level) to a multilevel discrete response (e.g., low, medium or high concentration of contaminant). Realistically, the current technology is more suitable to support the detection of the presence of a contaminant in the network via changes in commonly monitored water quality parameters, such as pH, conductance, and chlorine. Presence of unusual contamination can be deduced by detecting anomalies in the levels of these parameters. Based on rules to identify abnormal variations in commonly monitored water quality parameter measurements, event detection algorithms filter these observations to detect potential contamination in a water distribution network. Thus, the source identification procedures must be modified to include the filtered observations when characterizing the contaminant source. The potential reduction in information due to the filtering process makes the solution to the inverse problem more challenging. An objective of the work presented in this paper is to investigate the solution to the source identification problem based on filtered sensor information.

In addition, the information filtering through any event detection algorithm increases the degree of non-uniqueness in the solution to the inverse problem. Non-uniqueness arises when a set of different pollution source characteristics may be identified to give similar prediction errors (Zechman et al. 2006). Non-uniqueness in a system is related to the amount of data available to identify the source. When the available information is insufficient to identify a unique solution, determining a single solution may provide misleading assessment about the contamination event. One approach to address non-uniqueness is through the identification of a set of solutions representing different source characteristics that all lead to similar observations at the sensors. To generate alternatives systematically, a search is conducted for alternative source characterizations that are maximally different in their source characteristics and perform similarly well in predicting concentration profiles at the sensors. The degree of difference among the alternative solutions represents the range of possible source characterizations that fit the observed concentration profiles and helps quantify the uniqueness of an identified source. Zechman and Ranjithan (2004) proposed an evolutionary algorithm-based search procedure for generating alternative solutions. In another study, Zechman et al. (2006) addressed the non-uniqueness in the water distribution system contaminant source problem using this method. Another objective of this paper is to extend this procedure for addressing non-uniqueness in the source identification problem when the sensor information is filtered using an event detection algorithm.

Event Detection Algorithms

Time series of routine observations of pH and chlorine concentration, as well as of other water quality parameters, are available at sensors in a water distribution network. Event detection algorithms are employed for detecting anomalies in the observations that may indicate a contamination event. For a dynamic phenomenon as in a water distribution system, a contamination event is associated with any significant change in behavior over time. Data mining techniques can be applied to identify the time points at which the changes occur, thus triggering the presence of potential contamination events. Such a problem is called change-point detection problem in standard statistics literature. Guralnik and Srivastava (1999) proposed such an approach for contamination event detection based on a maximum likelihood estimation method applied to time series data. Klise and McKenna (2006) used a multivariate algorithm for the detection of anomalous water quality in a water distribution system. Multivariate distance between the current and expected water quality measurement is first calculated, where the expected water quality is defined as the nearest neighbor in multivariate space using all previous observations. If the distance from the expected water quality is within a threshold, the water quality is considered to be normal, and it is included in the moving window to predict the next water quality measure. If the distance exceeds the threshold, the water quality is considered anomalous and that measure is not used in determining the next value for expected water quality.

Another set of event detection procedures is associated with the degree of resolution of the sensors that are designed to detect specific contaminants. At the lowest degree of resolution, the sensor provides a binary response (e.g., detect or not detect) depending on whether the contaminant level in the water is above or below a pre-specified threshold level. At a higher degree of resolution, a sensor provides a discrete response (e.g., low, medium, high) depending on pre-specified threshold levels corresponding to each discrete level.

Source Identification Problem Description

An optimization model can be employed for detecting the contaminant source and its characteristics given the sensor information that is filtered through an event detection algorithm. The following set of equations, Eqns. 1-3, represents a mathematical description of the water distribution contamination source determination problem for a scenario where the observations are based on a set of low-resolution sensors. Depending on a pre-specified threshold concentration, the event detection algorithm generates a binary response. This model can be easily modified (by updating Eqns. 2 and 3) for other instances of low-resolution sensors with multiple discrete responses. Without loss of generality, this model is defined for a single contaminant source event.

$$\textit{Minimize } E = \max\left|BS_k(t) - BS^*{}_k(t)\right|, \forall k \in N_s, \forall t \in \Theta_s \tag{1}$$

Subject to

$$BS_k(t), BS_k^*(t) = 0 \;\; \textit{if} \;\; C_k(t,t_0,C_0), C_k^*(t,t_0,C_0) \le T_C \tag{2}$$

$$BS_k(t), BS_k^*(t) = 1 \;\; \textit{if} \;\; C_k(t,t_0,C_0), C_k^*(t,t_0,C_0) > T_C \tag{3}$$

where $BS_k^*(t)$ and $BS_k(t)$ are observed and simulated binary signals, respectively, at the sensor k at any time t. N_s is the number of sensor nodes in the network. Θ_s is the total time for which network is being simulated. $C_k(t,t_0,C_0)$ and $C_k^*(t,t_0,C_0)$ are simulated and observed concentration which is a function of time t, start time of the contamination event t_0, and the initial mass loading concentration C_0 of the contaminant, respectively, of the contaminant being monitored at any time t, and T_c is the threshold concentration level for detection of a contamination event. A contamination alarm is triggered when the concentration of the pollutant exceeds this threshold. The objective function, Eqn. 1, seeks to minimize an aggregation of the difference between $BS_k^*(t)$ and $BS_k(t)$, the observed and simulated binary signal, respectively, from the sensor at time t at sensor node k. Eqns. 2 and 3 convert the concentration observations at the low-resolution sensors into the detection/no-detection binary signal depending on a threshold concentration value for the particular contaminant.

Solution Approach

The inverse problem of source identification in water distribution system can be solved using a simulation-optimization method. Evolutionary algorithms (EAs) (Holland, 1975) are a class of heuristic methods that provide a global search mechanism to identify efficiently near-optimal solutions for large non-linear optimization problems. Dandy et al. (1996) and Savic and Walters (1997) used genetic algorithms for the optimal design of water distribution systems. Vitkosvsky et al. (2000), Lingireddy and Ormsbee (2002) solved the inverse problem of calibration of water distribution networks using evolutionary algorithms. In the present study, an EA-based procedure is employed for the source identification problem.

Evolution strategies (ES) (Schwefel, 1995) is an EA-based search method that is being explored as a solution approach for the source determination problem. ES searches using a population of solutions. The first generation starts with a random population of μ individuals. A set of λ new solutions are created in each generation using random probabilistic mutation. From the combined array of parent and offspring solutions (denoted as $(\mu+\lambda)$ selection) or from the set of offspring alone ((μ,λ) selection), the set of μ individuals are selected for the next generation. Zechman et al. (2006) used an ES-based search approach to solve the source identification problem in water distribution systems. In that study, the inverse problem was defined in terms of continuous concentration observations assuming the availability of perfect sensors for the specific contaminant. We extend that approach to solve the inverse problem with filtered information resulting from low-resolution sensors.

A Niched Co-Evolution Strategies (NCES) (Zechman and Ranjithan, 2004) extends an ES to generate a set of alternative solutions. NCES is based on concept of cooperative co-evolution to evolve a set of subpopulations to identify maximally different alternative solutions. The set of subpopulations are guided toward regions in the solution space that are distant from other subpopulations in the search for alternative solutions. The subpopulations share information and cooperate while co-evolving in the different regions of the solution space. Selection of fit solutions depends on both the prediction error of a solution and its distance from other solutions in the decision space. The details of the algorithmic steps can be found in Zechman and Ranjithan (2004). This approach was adapted by Zechman et al. (2006) to address the non-unique issue present when solving the source identification problem in water distribution systems. This approach is extended here to study the increased non-uniqueness arising from using filtered sensor information when resolving the source characteristics.

Illustrative Case Study

The application of evolutionary strategies coupled with event detection algorithms for detecting the source location is demonstrated through an illustrative case study. The problem of non-uniqueness due to limited information is addressed using NCES. An example network available as a part of standard EPANET distribution is used in this study. The network consists of 97 nodes including two sources, three tanks and 117 pipes, and the water distribution system is simulated using EPANET.

To generate a set of synthetic observations for an illustrative hypothetical contamination event, a non-reactive contaminant source is introduced into the network at a node and twelve sensors are placed in the network to observe the consequent concentration profiles. The hydraulics in the network is simulated hourly over a 24-hour time period. The hydraulics is assumed to be at steady state within each hour of the simulation. For each hourly hydraulic condition, the contaminant transport is simulated in 5-minute intervals, and the concentration values at the sensors are observed at the end of each 5-minute increment. These concentration observations are then processed through event detection algorithm to trigger a binary signal to trigger the presence of contamination based on a fixed threshold value. Numerous instances of contamination event with varying threshold rules for generating the discrete signals are considered in the study.

Results

Ongoing investigations addresses various aspects of the inverse problem of source identification including use of binary/discrete detection signals from the event detection algorithms, and various degrees of complexity of contamination events with single and multi-species contaminants. An ES-based search procedure is studied to evaluate the complexities associated with the source characterization problem as well as to assess the challenges posed to the solution algorithm. Alternative solutions are generated from the filtered information from event detection algorithms using NCES. Again, the degree of enhancement in the non-uniqueness issue resulting from data filtering through the event detection algorithms are evaluated, the complexities it introduces in water security management is studied. Different problem representation and algorithmic parameters are explored to verify the robustness of the proposed search methods. Many different problems instances are being explored with varying levels of sensor information filtering to understand the degree of difficulties in contamination event characterization using sensors with limited capabilities.

References

Dandy, G. C., Simpson, A. R., and Murphy, L. J. (1996). "An Improved Genetic Algorithm for Pipe Network Optimization" *Water Resources Research,* 32(2), pp. 449-458.

Guan, Jiabao., Aral, Mustafa M., Maslia, Morris L., Grayman, Walter M., (2006), "Identification of contaminant source in water distribution systems using simulation-optimization method: Case study", Journal of Water Res. Planning and Management, 132(4), 2006.

Guralnik, Valery., Srivastava, Jaideep., (1999), "Event detection from time series data", Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, United States, Pages: 33 – 42, 1999.

Holland, J. H. (1992). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press, Cambridge, MA

ILSI (1999), "Early warning monitoring to detect hazardous events in water supplies", International Life Sciences Institute-Risk Science Institute.

Klise, Katherine A., McKenna, Sean A., (2006), "Multivariate applications for detecting anomalous water quality", Proceedings of 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, OH, August 27-30, 2006.

Laird, Carl D., Biegler, Lorenz T., van Boleman Waanders, Bart G., (2005a), "A Mixed-integer approach for obtaining unique solutions in source inversion of drinking water networks", World Water Congress 2005, 173(47).

Laird, Carl D., Biegler, Lorenz T., van Boleman Waanders, Bart G., Barlett, Roscoe A., (2005b), "Contamination source determination for water networks", *Journal of Water Res. Planning and Management*, Vol. 131(2), 2005

Lingireddy, S. and Ormsbee, L. (2002). "Hydraulic Network Calibration Using Genetic Optimization." *Civil Engineering and Environmental Systems,* 19(1)

Savic, D. A. and Walters, G. A. (1997). "Genetic Algorithms for Least-Cost Design of Water Distribution Networks" *Journal of Water Resources Planning and Management,* 123(2), pp. 67-77.

Schwefel, H.-P. (1995) Evolution and Optimum Seeking, Wiley & Sons, New York

USEPA (2005), "Technologies and techniques for early warning systems to monitor and evaluate drinking water quality: A state-of-art review."

van Bloemen Waanders, B. G., Bartlett, R. A., Bigler, L. T., and Laird, C. D. (2003) "Nonlinear Programming Strategies for Source Detection of Municipal Water Networks." *Proceedings of the ASCE World Water and Environmental Congress*, Philadelphia, PA, June 23-26

Vitkosvsky, J. P., Simpson, A. R., and Lambert, M. F. (2000). "Leak Detection and Calibration using Transients and Genetic Algorithms" *Journal of Water Resources Planning and Management,* 126(4), pp. 262-265.

Zechman, Emily M., and Ranjithan, S., (2004), "An Evolutionary Algorithm to Generate Alternatives (EAGA) for Engineering Optimization Problems." *Engineering Optimization,* 36(5), pp. 539-553.

Zechman, Emily M., Brill, E. Downey., Mahinthakumar, G., Ranjithan, S., Uber, James., (2006), "Addressing non-uniqueness in a water distribution contamination source identification problem", Proceedings of 8th Annual Water Distr. Sys. Anal. Symposium, Cincinnati, OH, August 27-30, 2006.