Water Research 110 (2017) 180–191

Contents lists available at ScienceDirect

Water Research

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

Integrating physically based simulators with Event Detection Systems: Multi-site detection approach

Mashor Housh*, Ziv Ohar

Faculty of Management, Department of Natural Resources and Environmental Management, University of Haifa, Haifa, Israel

ARTICLE INFO

Article history:
Received 2 June 2016
Received in revised form
8 November 2016
Accepted 3 December 2016
Available online 8 December 2016

Keywords:
Water distribution systems
Model-based EDS
Multi-site Event Detection
Water security
Spatial contamination simulation
Water quality

ABSTRACT

The Fault Detection (FD) Problem in control theory concerns of monitoring a system to identify when a fault has occurred. Two approaches can be distinguished for the FD: Signal processing based FD and Model-based FD. The former concerns of developing algorithms to directly infer faults from sensors' readings, while the latter uses a simulation model of the real-system to analyze the discrepancy between sensors' readings and expected values from the simulation model. Most contamination Event Detection Systems (EDSs) for water distribution systems have followed the signal processing based FD, which relies on analyzing the signals from monitoring stations independently of each other, rather than evaluating all stations simultaneously within an integrated network.

In this study, we show that a model-based EDS which utilizes a physically based water quality and hydraulics simulation models, can outperform the signal processing based EDS. We also show that the model-based EDS can facilitate the development of a Multi-Site EDS (MSEDS), which analyzes the data from all the monitoring stations simultaneously within an integrated network. The advantage of the joint analysis in the MSEDS is expressed by increased detection accuracy (higher true positive alarms and fewer false alarms) and shorter detection time.

1. Introduction

Nowadays, there is a growing concern related to securing critical infrastructure such as water infrastructure in which possible contamination could affect a large number of consumers. Therefore, developing tools for early contamination events detection became a necessity for securing these systems. Events such as the water supply poisoning in Scotland (Gavriel et al., 1998), the contamination events in Japan (Yokoyama, 2007) and the recent Elk River contamination spill (Manuel, 2014) further highlight this point.

Most of the studies (as manifested in the literature review) focused on the development of single-site, signal processing based Event Detection Systems (EDSs), in which the reading from a single monitoring stations are analyzed to infer normal and abnormal conditions of the system. Moreover, most of these signal processing based, single-site EDSs are based only on quality parameters which are measured at the monitored site and exclude other factors such as the water source, operational hydraulic changes, tank levels, and longitudinal and radial mixing which can result in very high quality parameters' variability. This high, but normal, background variability reduces the likelihood of the signal processing based EDS to detect a real event. Noteworthy that some signal processing based EDS (e.g. Canary) does take into account operational data signals, such as tank levels, pressures and pumps status. Nevertheless, this data is analyzed using signal processing tools without a physically based model which links the different variables.

Criticism over the effectiveness of the single-site approach (Water Research Foundation, 2014; Liu et al., 2016) motivates the development of multi-site EDSs. Water Research Foundation (2014) found that EDSs which use sensor data from multiple sites could reduce false positives and negatives rates and overcome some of the drawbacks of the single-site EDSs.

While the single-site, signal processing based approach was demonstrated to be effective in several studies in the literature, nearly all EDSs (except Liu et al., 2015) were evaluated against artificial simulated contamination events. These artificial events were generated by adding "spike-like noise" to background water quality data obtained from water utilities during normal operation conditions (Klise and McKenna, 2006; McKenna et al., 2008). As such, these simulated events contain neither the specific chemical reactions nor the hydraulics operation conditions in water network. This type of events' generating approach is over simplifying the

* Corresponding author.
E-mail address: mhoush@univ.haifa.ac.il (M. Housh).

http://dx.doi.org/10.1016/j.watres.2016.12.003
0043-1354/© 2016 Elsevier Ltd. All rights reserved.

contamination event problem by neglecting the correlation between the multiple sensors' signals, the hydraulics conditions of the network, and the chemical reactions between the different quality parameters.

The first objective of this paper is to develop a framework which is able to simulate events that preserve the temporal and the spatial correlation between the signals of the multiple sensors. This is achieved by integration chemically and physically (i.e. hydraulics) simulator to enable the simulation of the contamination event propagation inside the water distribution system and thus preserve the temporal and the spatial correlation between the signals of the multiple sensors.

The second objective of the study is to show how the performance of the single-site, signal processing based EDS is significantly reduced when considering the hydraulic conditions of the water network. To cope with this reduction in performance, we propose a model-based EDS which integrates a hydraulic simulator within the EDS framework, and thus explain a large amount of the variability in the signals by the physical and the chemical interactions instead of trying to explain the whole variability by signal processing methods.

The third objective of the study is to develop a multi-site EDS (MSEDS), which detects the contamination events by simultaneously analyzing the data from multiple stations as well as the hydraulic operation and the chemical interactions in the water distribution system. Thus, unlike the single station approach all the characteristics of the network e.g. pressure, flow, velocities, etc., are utilized to improve detection performance.

2. Literature review

The EDSs research area can be divided into two subfields, the first is experimental and conceptual and the second is the development of new algorithms for contamination events detection.

On the experimental and the conceptual side of the EDSs, for example, Byer and Carlson (2005) and Hall et al. (2007) performed contaminant event detection experiments in Water Distribution Systems (WDSs) to show how different water quality parameters diverge from "normal" behavior as a result of different pollutants. This change can trigger events detection and thus provide a mechanism for an early warning system. Yang et al. (2008) showed that free chlorine residual behavior could be used for alerting on the presence of a contaminant. Helbling and VanBriesen (2009) developed and simulated a model for predicting the behavior of chlorine within the WDS following a microbial contamination event. USEPA (2013a) tested different water quality parameter response for four large municipalities. Schwartz et al. (2014) simulated organophosphate pesticides contaminants behavior in a WDS. In addition to rapid decrease in free chlorine, Schwartz et al. (2014) also included the predicted changes in alkalinity and pH. Panguluri et al. (2009) have conducted an evaluation for sensors technology in monitoring systems of WDSs to identify technologies which are capable of detecting anomalous changes in water quality due to contamination events. All of the above studies motivated the surrogate approach, in which instead of trying to identify the presence of the pollutant, we try to detect their footprint in the signals of classical water parameters which are regularly monitored in the WDS.

On the algorithmic research of EDSs, Klise and McKenna (2006) used multivariate Euclidean distance to classify the sensors' measurement as normal or events. Whereas McKenna et al. (2007) used binomial event discriminator method for the same purpose. McKenna et al. (2008) used a linear prediction filters to predict the water quality at a future time step and calculated a residual between predicted and observed which is then classified as normal operations or contamination events. Murray et al. (2011) demonstrated a methodology to detect E. coli contamination using surrogate parameters of turbidity, conductivity, and pH using Bayesian Belief Networks (BBNs). Hou et al. (2013a) described an EDS methodology based on integration of wavelet analysis and RBF neural network. Compared to time series increments, the proposed algorithm resulted in higher detection probability and lower false alarm rates. Hou et al. (2013b) developed a methodology based on three interconnected stages of autoregressive prediction, assigning probabilities based on predication's error, and inferring events based on Dempster–Shafer evidence theory. Perelman et al. (2012) and Arad et al. (2013) used surrogate approaches for EDS by utilizing Artificial Neural Networks (ANNs) for detecting possible outliers in several water quality parameters. Housh and Ostfeld (2015) extended the work of Arad et al. (2013) by developing a methodology for integrating single alarms through the use of a logit model.

Oliker and Ostfeld (2014a) developed a weighted Support Vector Machine (SVM) model for event detection. Oliker and Ostfeld (2014b) used a machine learning model for water quality event detection based on a multivariate analysis combined with an unsupervised minimum volume ellipsoid scheme for events classification. Liu et al. (2014) used laboratory data of contamination experiments for the development of event detection algorithm. Results showed contaminant detection capabilities after 9 min from the injection time with contaminate concentration of 0.01 mg/L. Using the laboratory results reported in Liu et al. (2014, 2015) improved the reported detection capabilities by suggesting an EDS methodology based on a non-dominated sorting genetic algorithm. The model was capable of detecting contaminants intrusions at a concentration of 0.008 mg/L after only 1 min of injection. Liu et al. (2016) tested conventional methods on reconstructed real contamination incidents and concluded that some single-site EDSs are more suitable for detecting sudden "spike-like" events which are different from the pattern of real-contamination events, and thus, these systems will fail in a real-life application.

In the industry, several commercial EDSs were developed by commercial companies and are available in the market (such as Bluebox™, GuardianBlue™ and Aquarius™). Some of these products are often coupled with hardware systems and monitoring equipment sold by the companies. In addition, the USEPA provides a freeware CANARY (USEPA, 2012) for event detection purposes (Note: CANARY is not commercial product). To date, all of the commercial EDSs and CANARY have the same basic leading principle of recognizing anomalies in the water quality data after applying machine-learning algorithms on available historical data from sensors' signals.

The previously described efforts focused on the development of a single-site EDS. MSEDSs which are based on multiple stations is gaining attention in recent studies. Nearly all the papers discussed earlier, highlighted the potential of multi-site approach to enhance the performance of the EDS. Water Research Foundation (2014) discussed the limitation of the single-site approach and proposed a multi-site EDS system which is comprised of two sensor stations with one possible path between the two. Autocorrelation between the stations was used to infer the event from the simultaneous analysis of the signals from the two stations. Koch and McKenna (2011) simulated a spatial contamination event (a plume) and developed a spatial analysis tool which uses a space-time clustering on the measurement to define a group of sensors that were expected to detect the contaminant. Yang and Boccelli (2014) integrated the binary signals from independent single-site EDSs using a backtracking algorithm and cross-referencing to evaluate events' probability. Yang and Boccelli (2016) used simulated water quality models of nicotine and KCN for generating spatial data of pH, free chlorine and conductivity which are used as surrogate quality parameters in the EDS. Oliker and Ostfeld (2015) suggested a spatial