Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

A dynamic thresholds scheme for contaminant event detection in water distribution systems

Jonathan Arad [a], Mashor Housh [b], Lina Perelman [a], Avi Ostfeld [a,*]

[a] Faculty of Civil and Environmental Engineering, Technion – IIT, Haifa 32000, Israel
[b] Faculty of Civil and Environmental Engineering, University of Illinois, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 4 September 2012
Received in revised form
22 November 2012
Accepted 8 January 2013
Available online 23 January 2013

Keywords:
Water distribution systems
Water quality
Water security
Event detection
Dynamic thresholds
Bayesian analysis

ABSTRACT

In this study, a dynamic thresholds scheme is developed and demonstrated for contamination event detection in water distribution systems. The developed methodology is based on a recently published article of the authors (Perelman et al., 2012). Event detection in water supply systems is aimed at disclosing abnormal hydraulic or water quality events by exploring the time series behavior of routine hydraulic (e.g., flow, pressure) and water quality measurements (e.g., residual chlorine, pH, turbidity). While event detection raises alerts to the possibility of an event occurrence, it does not relate to origins, thus an event may be hydraulically-driven, as a consequence of problems like sudden leakages or pump/pipe malfunctions. Most events, however, are related to deliberate, accidental, or natural contamination intrusions. The developed methodology herein is based on off-line and on-line stages. During the off-line stage, a genetic algorithm (GA) is utilized for tuning five decision variables: positive and negative filters, positive and negative dynamic thresholds, and window size. During the on-line stage, a recursively Bayes' rule is invoked, employing the five decision variables, for real time on-line event detection. Using the same database, the proposed methodology is compared to Perelman et al. (2012), showing considerably improved detection ability. Metadata and the computer code are provided as Supplementary material.

1. Introduction

Incidents of contaminant injections into water distribution systems emphasize the need for contamination event detection systems. Events like the poisoning of drinking water in Scotland (Gavriel et al., 1998) and the water supply system poisoning in Japan (Yokoyama, 2007) highlight how intentional sabotage remains a major concern to public health (WHO, 2004; Greenfield et al., 2002). Accurate alerts in real-time are required with respect to response strategies, operation optimization, and overall system efficiency (Kroll, 2006; Jain and McLean, 2006; Hasan et al., 2004).

Contamination event detection methods are aimed at identifying abnormal system behavior. Identification of most contaminants typically requires grab sampling followed by laboratory analysis (ASCE, 2004). Since on-line instrumentation capable of measuring and detecting all possible contaminants does not exist, the presence of pollutants can be inferred only through surrogate measurements. Such measurements can distinguish irregularities in monitored water quality, hydraulic parameters, and their interplay during normal operating conditions. This approach rests on the premise that a contaminant injected into a water distribution system (WDS), whether deliberately, accidentally, or

* Corresponding author. Tel.: +972 4 8292782; fax: +972 4 8228898.
E-mail address: ostfeld@tx.technion.ac.il (A. Ostfeld).
0043-1354/$ – see front matter
http://dx.doi.org/10.1016/j.watres.2013.01.017

naturally, will affect at least one of the on-line monitored hydraulic and/or water quality parameters and the interactions between them. Based on this notion, the analysis of continuously monitored data can be utilized to give an indication of contaminants' presence in a water distribution system.

The US government generated research in this field by establishing the Environmental Technology Verification program in 1995 to investigate how changes in water quality parameters can be detected by real-time sensors. Extensive research is being performed by academics, governmental agencies, private companies, and in close collaborations between them. These projects range from theoretical to applied research on commercially-developed hardware and software. Klise and McKenna (2006) and McKenna et al. (2008) suggested three approaches to water quality detection based on a comparison between predicted and measured values that evaluate each technique through receiver operating characteristic (ROC) curves. The US Environmental Protection Agency (EPA) has conducted experiments on more than 30 contaminants (e.g., pesticides, insecticides, metals, bacteria), which may be employed in intentional acts of water contamination (U.S. EPA, 2005a; U.S. EPA, 2005b). Hall et al. (2007) provided results from experiments that tested the response of commercial water quality sensors of different designs and technologies to chemical and biological loads. These tests suggested that free chlorine, total organic carbon (TOC), oxidation reduction potential (ORP), conductivity, and chloride were the most reactive parameters to the majority of contaminants. Yang et al. (2009) tested 11 contaminants at different concentrations using a pilot-scale pipe system. The suggested adaptive transformation of sensory measurements reduced background noise and enhanced contaminant signals. This allowed for contaminant detection and further classification based on chlorine kinetics.

Today, water utilities have access to a variety of private and public event detection hardware and software. Hach (GuardianBlue™) produces commercial hardware and software solutions for event detection. Edthofer et al. (2010) provides event detection software combined with management and data validation tools. The CANARY software (Hart et al., 2007, 2010; Murray et al., 2010), developed at Sandia National Laboratories in collaboration with the EPA National Homeland Security Research Center, is a freely available tool for detecting contamination events. It provides both off-line and on-line analysis capabilities for detecting anomalies in regularly-monitored water quality and hydraulic parameters that indicate possible contamination events.

The development of multi-parameter water quality models and their calibration is challenging due to the large amount of information and number of parameters involved (e.g., Chungsying et al., 1994; Rauch et al., 1999). Thus, their application in the context of contamination event detection is complicated.

Data-driven models such as artificial neural networks (ANNs) have been successfully used for water quality assessment that focuses on modeling and prediction. This includes prediction of residual chlorine, the temporal variation of substrate, biomass concentrations, and residual chlorine (e.g., D'Souza and Kumar, 2009, Gibbs et al., 2006; Rodriguez et al., 2002; Rodriguez and Serodes, 1999). Arad et al. (2011) utilized decision trees for evaluating event detection procedures. Later, Perelman et al. (2012) utilized ANN to improve the performance of the estimation model. To detect possible quality threats in water distribution systems, Perelman et al. (2012) integrated ANN models with Bayesian sequential analysis to estimate the probability of such events. Closely related methodologies were also utilized to detect leaks and bursts in water distribution systems (Romano et al., 2010a, 2010b).

The method described below attempts to improve the study of Perelman et al. (2012) by applying adaptive updating dynamic thresholds, which are used to classify sensory observations. Dynamic, as opposed to fixed, thresholds are characterized by a sliding window size, and positive and negative dynamic thresholds that are computed as a function of the error standard deviation of the sliding window. In addition, dynamic thresholds are characterized by and positive and negative filters, which discriminate possible outliers in the measurements.

In this method, optimal parameter values are obtained using the commonly applied genetic algorithm (GA) (Holland, 1975; Goldberg, 1989). The following sections provide details on the suggested method and its application to contamination event detection in water distribution systems.

2. Methodology

The general event detection framework relies on continuously transmitted hydraulic and water quality data from the supervisory control and data acquisition (SCADA) system.

A previously published event detection method (Perelman et al., 2012) consists of four main modules for each incoming observation from the SCADA system: (1) data analysis – the interplay between water quality parameters are examined using ANNs, (2) identification of outliers – calculated residuals are classified as normal or abnormal observations using fixed thresholds for each parameter, (3) identification of events – based on consequent classification of errors, events are identified by updating Bayesian probability, and (4) synchronized decision – information from multiple quality indicators is fused to provide a unified decision support about a contamination event. Steps 1 and 2 are initially trained using available data collected from the water utility. Steps 1–4 are then repeatedly executed in real-time for each new observation. This study suggests improving stage 2 by identifying outliers using an updating dynamic threshold instead of a fixed one. This would upgrade all subsequent stages/events identification and fused decisions. A summary of the overall event detection methodology is presented in Fig. 1 and described in following sections.

2.1. *Error calculation*

The input parameters attained from the SCADA system include hydraulic (pressure) and water quality (e.g., residual chlorines, pH) time series data.

Artificial neural networks (ANNs) are utilized to estimate the relationships between water quality parameters during normal operation. The development of ANN does not require