Detection of chemical intrusion compounds in water distribution networks by quality sensors data mining

Eva Carolline M. Oliveira¹, Bruno Melo Brentan², Renato F. Dantas³, Letícia dos Santos Macedo⁴, Edevar Luizotto Junior⁵, Lubienska C. L. J. Ribeiro⁶

¹,³,⁶ School of Technology, University of Campinas, Limeira, Brazil
² CRAN, Université de Lorraine, Vandoeuvre-les-Nancy, France
¹,⁴ Institute for Technological Research, Sao Paulo, Brazil
⁵ School of Civil Engineering, University of Campinas, Campinas, Brazil
¹evaoliveira@ipt.br

ABSTRACT

Several consequences can arise from unexpected terrorist actions such as an attack by chemical contamination into the drinking water distribution system. To investigate possible threats of chemical contamination in the system to detect early events and actions can help water distribution network operators making assertive decisions. This work presents a model capable of event detection parathion contamination from the monitoring of quality sensors in the network. The methodology considers the chemical reactions of the parathion pesticide when present in water into the problem of detection, using the program EPANET-MSX. The data generated from the hydraulic model and the drinking water quality served to create a pattern recognition system using the artificial neural network NARX in order to, estimate the behavior of chlorine in the network, while a change point identification methodology is used to identify significant changes caused by the contaminant inserted in the network. An estimated error analysis was performed and simulated attacks were identified.

Keywords: Fault detection, EPANET-MSX, NARX.

1 Introduction

The water distribution system (WDS) can be seen as a potential place for actions of terrorist attacks. Population coverage is one of the attractions for actions of this nature. Among the dangerous used substances, chemical contamination is one of the threats in which system is exposed and the problems became even worst because systems management companies are still unprepared to deal with these emergency situations. The absence of an emergency protocol can jeopardize an entire community.[1]

In order to test and evaluate mitigation decisions to protect public health, researchers are currently focus on studies of emergency events of contamination in WDS [2-3]. The safety of water distribution systems has become a relevant issue especially in countries that have already been targets of terrorist attacks. The United States is an example, especially after the attack on September 11, 2001.

To identify possible scenarios of both intentional and accidental contamination is a complex task. Demand patterns change daily, making difficult to predict the behavior of the system [4]-[5]. Moreover, the features of the injected contaminant are unknown and several chemical products can be used. Thus, methods for physically assuring the system are fundamental, as well as the forms of
continuous or periodic monitoring intrusions in several vulnerable points in the water distribution system.

Monitoring methods are needed to secure the system against intrusion into the vulnerable points of the WDS. Sensors are instruments that allow the monitoring of water quality in order to detect and measure changes in water parameters. Changes observed in these parameters can indicate intrusion of a possible contaminant in the network [6]-[9].

In order to identify the contamination sources, a systematic procedure to test and evaluate contamination identification methods (SI methods) was presented [9]. It is based on the analysis of sensor data, and thus collaborate with researchers and professionals in the area in the evaluation and planning of actions in case of contamination.

Researchers deal with the problem of detection of potential contaminants in the WDS using Artificial Neural Networks (ANNs) to estimate relationships between water quality parameters in the distribution network. Several event simulations were performed and an event detection algorithm was applied to detect anomalous behavior of water quality parameters [10].

The variety of chemical contaminants that can be entered into a water distribution network is enormous. Study the location of sensors using a methodology that considers the chemical reactions of the possible contaminants, allows analyzing the real changes that occur in the quality parameters of drinking water. Information concerning parameters such pH, chlorine and alkalinity of water, for example, helps to prevent possible chemical attacks. Contamination event is closely connecting the type of contaminant and its concentration in the system.

Pesticides such as organophosphates have potential to be used as a contaminant in a water distribution network due to its easy access and high toxicity to human beings. Consumption of organophosphate may result in a malfunction of the nervous system or even death if consumed in large quantities. When in contact with chlorine, the most commonly used disinfectant for drinking water, the organophosphate is oxidized and forms its corresponding oxon which is usually more toxic than the original contaminant [11]. This work used as a model contaminant in the simulations the organophosphate parathion (PA).

This work proposes to develop a computational tool that detects the intrusion of chemical components in the water distribution system network, able to reduce the empiricism in decision making and reduce the number of victims in case of attacks in potable water systems.

2 Methods

This paper presents the development of a hydraulic and water quality model and for a hypothetical network capable of detecting chemical contaminations. Two situations are simulated in the network. An absent period of contamination and another with simulated chemical attacks at specific times. For this purpose, methodology was comprised of three stages presented below.

2.1 Water quality simulation

The normal and abnormal scenarios are generated from the EPANET-MSX program, widely used to simulate water quality models [12]. Considered an extension of EPANET, it takes into account the reactions and transport of multiple species interacting in the mass flow and in the tube walls. Programs developed with the use of EPANET-MSX allow an extensive process diagnosis with safety and prevent deterioration of water quality at all points of the network, either by internal and external factors such as: low speeds and stagnation points of the flow, exposure to external agents at leakage points or by sabotage such as terrorist attacks [12]-[13].
The results of studies to identify contamination events are better evaluated when the chemical degradation of the contaminant in the water is taken into account, since it can suffer reactions of difficult prediction. The incorporation of chemical details into the model improves the ability to evaluate the potential damages of a contamination event and can help in the early detection process, as demonstrated in the work of Ohar et al. [11].

Based on the work of Ohar et al. [11] and Schwartz et al. [14], this research used a pesticide from the group of organophosphates as a model contaminant, the PA. All simulations of contamination events were carried out with PA. The water quality data is generated using the MSX input file that defined the individual water quality species of interest and the reaction expressions that govern their dynamics. The stoichiometric reactions, rate and equilibrium coefficients used in the degradation of PA considered in the input data of the EPANET-MSX are presented in Table 1.

<table>
<thead>
<tr>
<th>Reaction Stoichiometry</th>
<th>Rate/Equilibrium Coefficient at 25 ºC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5\cdot HOCl + PA \xrightarrow{K_{HOCl,PA}} PAO + 5\cdot H^+ + 5\cdot Cl^- + SO_2^-)</td>
<td>(K_{HOCl,PA} = 2.2(\pm 0.53) \cdot 10^6 \text{M}^{-1}\text{h}^{-1})</td>
</tr>
<tr>
<td>(PA \xrightleftharpoons{K_{PA,PA}} PAH)</td>
<td>(K_{PA,PA} = K_{N,PA} + K_{E,PA}\cdot [OH^-])</td>
</tr>
<tr>
<td></td>
<td>(K_{N,PA} = 2.66 \cdot 10^{-4} \text{h}^{-1})</td>
</tr>
<tr>
<td></td>
<td>(K_{E,PA} = 4.3 \text{M}^{-1}\text{h}^{-1})</td>
</tr>
<tr>
<td>(PAO \xrightarrow{K_{PAO,PA}} PAH)</td>
<td>(K_{PAO,PA} = K_{N,PAO} + K_{E,PAO}\cdot [OH^-])</td>
</tr>
<tr>
<td></td>
<td>(K_{N,PAO} = 2.1 \cdot 10^{-4} \text{h}^{-1})</td>
</tr>
<tr>
<td></td>
<td>(K_{E,PAO} = 46.1 \text{M}^{-1}\text{h}^{-1})</td>
</tr>
<tr>
<td>(PA + OCl \xrightarrow{K_{OCl,PA}} PAH)</td>
<td>(K_{OCl,PA} = 37(\pm 10) \text{M}^{-1}\text{h}^{-1})</td>
</tr>
<tr>
<td>(PAO + OCl \xrightarrow{K_{OCl,PAO}} PAH)</td>
<td>(K_{OCl,PAO} = 48(\pm 10) \text{M}^{-1}\text{h}^{-1})</td>
</tr>
<tr>
<td>(HOCl \rightarrow H^+ + OCl^-)</td>
<td>(pK_a = 7.5)</td>
</tr>
</tbody>
</table>

The total free chlorine is the parameter more affected by contact with PA is total free chlorine [14]. The PA introduces in the network reacts with the chlorine in the water and oxidizes in paraxon (PAO). This byproduct is considered to be more toxic than the parent compound (PA). PAO, on the other hand, undergoes hydrolysis and generates PAH, a less toxic byproduct. Thus, the chlorine concentration change is used as an indicator of contaminant intrusion in the network. Small changes in alkalinity and pH are monitored to sustain this indication in order to reduce the likelihood of false positive alarms. Therefore, the analyzes of results of this paper are made based on the changes in chlorine concentration during the simulations.

### 2.2 Water quality estimation using NARX network

The estimation of the quality parameters in the water distribution network can be performed accurately from ANN approaches. The second stage of this work uses the generated database to train an ANN. This training consists of a learning process where the free parameters of the ANN is adjusted to well perform outputs signal [15]. Its function is to modify ANN synaptic weights in an orderly manner in order to detect the chemical contamination in the water distribution network, in the case of this work.
The Non-linear Auto-Regressive with Exogenous Inputs - NARX model is adopted in this work to estimate the behavior of chlorine in the network, while a change point identification methodology is used to identify significant changes caused by the contaminant inserted in the network. Some studies have used the NARX model to predict and control water distribution networks satisfactorily [16]-[17]. The model consists mainly of two modules, one composed of Multilayer Perceptron (MLP) which has a strong non-linear mapping ability, and the other is composed of learning and training methods for Multilayer Perceptron [16].

A crosses model is performed, where each monitoring point has a neural network for it. The NARX are trained using the hour of the day and the day of the week as external inputs and delay’s chlorine (adopted with 12 steps of time) to model the chlorine concentration. The architecture of the NARX is compound by 12 neurons in a hidden layer, a delay of 20 and an intermediate layer. The comparison between estimated time-series and the measurements allows a statistical process to generate warning flags for a possible attack.

### 2.3 Water quality fault detection

Since the NARX networks is trained using the generated data for normal conditions, the analysis of the residual signal (obtained by the difference between the estimated and the measured signals) can be conducted to identify the faults in the quality system.

To ensure the detection of attack, a statistical control process to identify abrupt changes in time series is applied in error estimation. In this study, the standard deviation is used as statistical control variable, since the mean of this series tends to zero, as an imposition of the tuning process [17]. In a contamination event, the error increases significantly, although the mean value can be zero. On the other hand, the standard deviation increases and is different from the condition of non-attack, favoring the control of this statistical parameter.

### 3 Case Study

The methodology described is applied in an urban network [18] as shown in Figure 1. Network topology consists of 15 nodes, 23 links and 3 sources of supply. The data generation for this study is done for a period of 8760 hours (equivalent to period of a year). EPANET and EPANET-MSX are used to design hypothetical chemical attacks. Both are integrated into programming environment Matlab. Period-extensive simulations of the hydraulic behavior and water quality within the network are carried out.

The model of water quality considered in simulations was based on piped water supplies adopted in Brazil. The chlorine used to network chlorination was added on three reservoirs each 12 hours of simulation, with concentration of 1.5 mg/L, according by the Water Quality Ordinance [19]. The experiment lasted 8760 hours with reports of quality every 15 minutes.

The first six months of simulation is carried out under normal conditions, without anomalies. During the last six months, three chemical attacks occurred in random periods. The contaminant is added constantly at different nodes at concentration of 12.4 mg/L. This rate represents PA limit of solubility. Concentration values above this threshold could result in contaminant visibility in water, that is irrelevant to this research. Table 2 presents simulated attacks occurring in the period of one year.
Nodes 12 and 15 are chosen as monitoring points of the chlorine concentration during the experiment. When analyzing the flow network, the water flows through all other nodes reaches nodes 12 and 15. Thus, it can identify the contamination of most of the attack scenarios and assume the role of network sensors. In Figure 1 is identified all reservoirs and nodes cited. Figure 2 represents the simulation performed for the monitored nodes. It is seen chlorine behavior in the network under normal conditions and on chemical contamination scenario. As expected, the decay of chlorine concentration is abrupt in contamination conditions. This reinforces its role as an external intrusion indicator. In the patterns identification step, a water quality database was build. The neural network NARX is trained with these data set. In order to evaluate forecast quality, curve of estimated data is overlapped real data curve in Figure 3.
The results match to standard deviation control of the series of errors for each forecasting of contamination event. Three simulated attacks are identified by error analysis as observed in Figure 4, which we can observe a very small estimated error.

4 Conclusions

Detection of chemical contamination events in water distribution networks is favored by the chemical information specific to the contaminant in question. In this way it is possible to evaluate the potentiality of the damages. Thus, the detection system that considers the chemical impact becomes more agile and more accurate than most systems that use time-series-oriented signal processing.

The model proposed in this work was able to detect an anomaly from hydraulic monitoring and water quality. The high prediction resulting from the NARX network ensured the accuracy of the identification of simulated chemical attacks.

The results of this research can contribute significantly to increase the analysis of adverse system risks, assist network operators in decision-making and reduce the number of victims in the event of chemical contamination in drinking water systems, especially dealing with terrorist threats.
Figure 4. Controlled error for monitored chlorine concentration in node 12 and 15 for concatenated non-attack and attack dataset and detail of an identified attack.

5 Acknowledge

This study is supported by the IPT, Technological Research Institute of the State of São Paulo, Brazil. The authors are grateful for the support, which is primordial for the development of the work.

6 References


