CHALMERS
UNIVERSITY OF TECHNOLOGY

Effective placement of sensors for efficient early warning system in water distribution network

Master's thesis in Master Program Infrastructure and Environmental Engineering

MASHREKI ISLAM SAMI

Department of Architecture and Civil Engineering
Division of Water Environment Technology
CHALMERS UNIVERSITY OF TECHNOLOGY
Master's Thesis ACEX60
Gothenburg, Sweden 2018

MASTER’S THESIS REPORT NO. ACEX60-18-4

Effective placement of sensors for efficient early warning system in water distribution network

MASHREKI ISLAM SAMI

Department of Architecture and Civil Engineering

Division of Water Environment Technology

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2018

Effective placement of sensors for efficient early warning system in water distribution network

Master's Thesis in Master Program Infrastructure and Environmental Engineering

MASHREKI ISLAM SAMI

Master's thesis report no: ACEX60-18-4

Department of Architecture and Civil Engineering

Division of Water Environment Technology

Chalmers University of Technology

SE-412 96 Göteborg

Sweden

Telephone + 46 (0)31-772 1000

Department of Architecture and Civil Engineering

Division of Water Environment Technology

Effective placement of sensors for efficient early warning system in water distribution network
MASHREKI ISLAM SAMI
Department of Architecture and Civil Engineering
Chalmers University of Technology
Gothenburg, Sweden 2018

ABSTRACT

Water security and monitoring of quality of supplied water in the distribution network has become one of the important issues in recent times especially with rising concern of deliberate contamination as an act of terrorism. Effective early warning system is the future in water distribution network (WDN) and this report contributes with significant knowledge on sensor placement localization in this regard. In this research study a simple WDN model was developed with 31 nodes which includes 3 outflow nodes and one inflow node. Contamination events at each node was simulated using EPANET. As previous studies experimented with optimization techniques and algorithms, this report focuses on simple technique for identifying critical node junctions for contamination events. All the nodes were ranked according to the chemical concentration received after intrusion at a particular node. A simple sorting matrix was then used for identifying the critical node for sensor placement and intrusion points in the WDN. To justify the sensor placement and their response activities, a pilot scale WDN was constructed to test water quality changes using pH sensors for acidic intrusion scenarios. Experiments with 27 combinations of sensor activities, outflows and intrusion points were carried out where results show similar pattern of hydrodynamics and quality changes like the EPANET simulations. Comprehensive analysis and comparison study show the usability and incapability of EPANET in tracking intrusion mass with respect to time. In EPANET the time pattern settings are dependent on user inputs which changes the response outputs. Hence time-variant calculations for source tracking becomes critical. Nevertheless, EPANET is very convenient to monitor flow parameters and concetration zones using the contour feature which can be additional support for optimization methods using different algorithms or Graph theory. Phenomena for chemical mass transport such as advective flow and dispersion and their effect on sensor response were key findings in this research. Intruded chemical flow pattern helped to understand how to predict the location of intrusion using sensor response pattern.

Keywords: water distribution network, sensor placement, optimization, early warning system, scaled distribution network, source tracking

Effective placement of sensors for efficient early warning system in water distribution network
MASHREKI ISLAM SAMI
Department of Architecture and Civil Engineering
Chalmers University of Technology
Gothenburg, Sweden 2018

SAMMANFATTNING

Vattensäkerhet och övervakning av dricksvatten i distributionsnätet har med tiden blivit en allt viktigare fråga särskilt som medvetenheten om terrorism och avsiktlig förorening av vattnet ökat. Effektiva system för tidig varning i distributionsnät kommer att bli vanligare i framtiden och denna rapport bidrar med ökad kunskap om var sensorer lämpligen bör placeras. I denna forskningsstudie har en pilotskalemodell av ett distributionssystem innehållande 31 noder, med 3 utloppsnoder och en inflödesnod, byggts upp. Föroreningshändelser vid varje nod simulerades med hjälp av EPANET. I andra rapporter har optimisering av sensorers placering utförts med hjälp av matematiskt avancerade algoritmer medan denna rapport enbart fokuserar på enkla tekniker för att identifiera kritiska noder för föroreningshändelser. Alla noder rankades utifrån uppmätt kemikaliehalt utifrån injicerad mängd kemikalier i specifika noder. En förenklad sorteringsmatris användes för att identifiera kritiska noder för placering av sensorer samt inträngningspunkter i distributionssystemet. För att verifiera placeringen av sensorer konstruerades ett ledningsnät i pilotskala där undersökningar av vattenkvalitetsförändringar genomfördes med hjälp av pH-sensorer, vilka detekterade värden från olika syrainträngnings-scenarier. Experiment med 27 olika kombinationer av sensorplaceringar, utflöden och injiceringspositioner genomfördes där resultaten i stort liknar det hydrauliska mönstret från EPANET-simuleringarna. En omfattande analys- och jämförelsestudie visar både på EPANETs användbarhet och oförmåga för spårning av mängden inträngd förorening i förhållande till tiden. I EPANET-modellen beror resultatet av tidsåtgången för ett spårämne att förflytta sig en viss sträcka helt av inställningen som användaren gör i programmet. Därför blir tidsvariationsberäkningar för källspårning kritisk. Trots allt är EPANET ett mycket bra verktyg för att övervaka flödesparametrar och koncentrationsvariationer, vilken kan bli ytterligare ett stöd för optimering med olika algoritmer eller grafteori. Fenomenen av kemisk masstransport som advektivt flöde och flödesdispersion samt deras effekt på sensorns signal var viktiga resultat i denna studie. Mönstret av hur det kemiska ämnet injekterats in i systemet bidrog till förståelsen av hur man kan förutsäga platsen där ämnet trängt in i systemet, med hjälp av signalerna ut från sensorerna.

Nyckelord: distributionsnät för vatten, sensorplacering, optimering, tidigt varningssystem, nedskalat ledningsnät, källspårning

ACKNOWLEDGEMENT

It gives me immense pleasure to acknowledge my sincere thanks and gratitude to my supervisor **Thomas Pettersson**; *Associate Professor* in the *Department of Architecture and Civil Engineering* at *Chalmers University of Technology*. This thesis would not have been possible without his enormous help and continuous support in all the ups and downs during the research works. It is an honor to work under his supervision and produce this thesis.

I would love to give my heartful thanks to the persons who were there just in the time of need; **Victor Vinas** and **Nashita Moona,** *Ph.D. students* in the *Department of Architecture and Civil Engineering* at *Chalmers University of Technology*. Their guidance and technical support helped me to overcome some of the major obstacles in this study and organize the data in this report to what it is today. The discussion sessions were motivating which kept me focused.

My sincere gratitude to **Lars-Ove Sörman** for all the efforts in building the laboratory scale water distribution network. His vast experience and technical expertise have shone light to this research study. I am lucky enough to learn one or two skills while setting up the network from Lars-Ove.

I am greatly indebted to my family members who always believed in me and supported throughout my course of study. I would also love to give my warm thanks to all my inspiring friends who kept motivating and held me high in the last two years, especially during my worst days. I am nothing without these relationships in my life.

Last but not least, I am most grateful to *Swedish Institute (SI)* for granting me the prestigious *Swedish Institute Study Scholarship*. Last two years have been an amazing experience with world-class research opportunity, fascinating events and workshops, and cultural exchanges which highly influenced my leadership towards sustainable developments. Plus, I cannot thank enough for all the travels to amazing destinations in Europe. I hope I can uphold *Swedish Institute's* legacy of building the Network for Future Global Leaders (NFGL).

Mashreki Islam Sami
Gothenburg, September 2018

Table of Contents

List of Figures

List of Tables

1

INTRODUCTION

Water is one of the most important inorganic compounds and a common substance found in earth where almost all living organisms are dependent on it for their biological processes. Even though 2/3rd of the earth's lithosphere is covered with water, there remains concerning issues regarding drinking water including uncertain events of contaminations, human intervened pollution, and scarcity. As a result, United Nations (UN) established the 17 Sustainable Development Goals (SDGs) where the 6th goal focuses on clean water and sanitation. In the recent report of secretary-general of UN, sustainable management and ensuring availability of water is prioritized (Progress towards the Sustainable Development Goals, 2017).

Clean potable water has always been major issues in the world mostly in the developing countries. Water distribution is one of the key infrastructural development that ensure safe drinking water to consumer. However, the huge networks of pipeline to distribute water are under threats of contamination events or leakage generating low pressure flow and pollutant intrusion. Moreover, by the time collecting and testing water sample for contamination, hundreds of people can be affected. Such event happened in 2000, an outbreak of waterborne disease epidemic in Ontario, Canada, affected 2,300 people (Hrudey, Payment, Huck, Gillham, & Hrudey, 2003) and investigations found out about polluted groundwater with primary pathogens *Escherichia coli* and *Campylobacter jejuni*. In 2007 in China, poisoning of water supply caused 71 people infected (Hu, Ren, Liu, Li, & Jie, 2017) and another severe outbreaks occurred in Finnish town where cross-connection between sewage and drinking water pipelines. Contamination in the tap water resulted 8453 residents on the town suffering waterborne gastroenteritis (Laine et al., 2011). In Unites States alone recorded 780 outbreaks associated with drinking water suppy within 36 years span (1971 to 2006) out of which 685 outbreaks caused acute gastroenteritis illness (Craun et al., 2010).

There have been many researches in the past few decades on water distribution system (WDS) with a goal to distribute safe water to consumers. Pipe breakage, leakage, contamination intrusions are some of the major concerns in the WDS until today. It is estimated that in an efficient water distribution systems (WDS), around 3 - 7 % of water is lost due to leakage whereas developing countries may lose up to 50% or more through leakage events (F. Colombo, Lee, & Karney, 2009). In addition, contamination events and their location are difficult to trace which give rise to the question, how much population is affected before the contamination is detected and the affected water pipes can be closed down and the contaminated water can be removed from the system. Hence to minimize the effect of such contamination events early; warning systems are being studied, developed, and tested for their serviceability and accuracy.

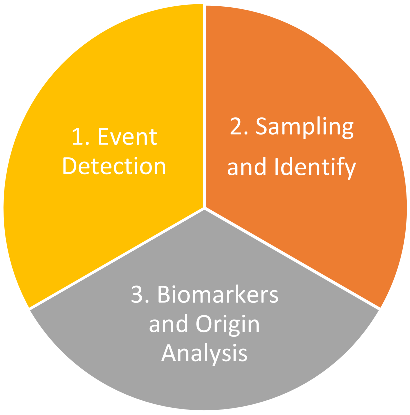
This report presents an experimental research conducted at Chalmers University of Technology to study the optimization of sensor location to effectively detect undesirable water quality and later generate an early warning signal.

1.1 Background information/ problem statement

Adoption of sensors in the WDS is an innovative and quick method for getting water quality data which has also reduced the involvement of physical sample collecting and testing. Optimal sensor placement for identifying contamination sources or pressure variation due to pipe leakage is a major issue in recent times primarily because of the cost associated with type and number of sensors. The problem arises even more when it comes to large networks where the deployment and the maintenance of the sensors requires substantial financial investment (Zeng, *et al*., 2016). Moreover, the uncertainties regarding contamination events at any place and time of the network are hardly predictable and detectable. Furthermore, the threat of chemical or biological contamination through deliberate injections can have severe consequences to consumers if the source and type of contaminant is not traceable (Ostfeld et al., 2008). To tackle these uncertainties of contamination events it is not possible to place sensors to every desirable location of a large WDS hence optimization with a smaller number of sensors is vital (Xu, small, VanBriesen, Fischbeck, & Johnson, 2010). Apart from the cost perspective, the amount of computational time and storage space required to process the data and generate early warning signal is a matter of complexity (Hu et al., 2017). For example, in a large water distribution network with contaminant intrusion in more than 10,000 nodes will generate millions of scenarios which takes nearly 170 days to reach exhaustive simulation (Krause et al., 2009).

1.2 Contamination identification processes

When dealing with contamination events within the WDS, there are three process that are considered important. The processes involve detection, identification, and analysis of the contaminant. The three processes are described further in the next page.



1. Event Detected: Using sensors and advanced computer programs, contamination is detected and traced back to source of the contamination therefore allowing to produce an early warning signal. Events may include pressure drop, leakage, intrusion of contamination or deliberate injection of contaminants. Time for this process of generating early warning signal should not be more than up to a minute.

2. Sampling and identify: This is the second process after the event is detected and location of the source of contamination is identified. This mainly involve collecting sample from the source and run quick tests to identify the type of contamination i.e. Chemical, Biological, Radiological or Nuclear (CBRN) emissions. This may take few hours before the authority can raise awareness and make necessary measures.

3. Biomarks or Origin Analysis: This step is to analyze the sample and case scenario to investigate source of contamination. For example, if *E.Coli* was found it indicates fecal contamination source and can be tracked down to nearby wastewater sources. This usually take long time to conclude and present a decision. This step allows us to make policies, measurements, and action plan for the future threatful events.

1.3 Objective of the study

The objective of this study falls in the first part of the process to identify the source of contamination and produce an early warning signal before reaching the consumers. This objective can be realize using optimal sensor placement (OSP) method and determine the most suitable placement or location in the pipe network. In addition to the sensor placement optimization, suitable mathematical models or algorithms should be used to simulate the data such that the time for detection and analysis is less, and space to store the data is minimum. To achieve the objective of generating early warning signal, it is important to have less number of sensors producing less but effective data which will require less time to simulate. The primary objective of the study will be to build a pilot scale WDN model and place sensors based on simulated data from EPANET. The data from simulations will be used to identify locations for sensors which will detect intrusion/contamination injected through specific node of the network. The later part of the study will analyze the response data from the sensors and make predictions to backtrack the source location of contamination. To develop an efficient early warning system the following criteria should be taken into consideration. I) Time required for detection, II) detection likelihood, III) size of population at risk of the contamination. Through experimentation it is likely to evaluate and analyze spread of contamination for conservative and non-conservative substances in the water. The research also opens the scope to contribute to the risk assessments and policy measures in the field of water distribution. Implementation of the knowledge from the research can improve available WDN in Gothenburg or Sweden significantly.

1.4 Hypothesis

The hypothesis developed for this research is mainly focused on the experimental study of a pilot scale WDN. The optimization scenario from the hydraulic simulation model (EPANET) and the pilot scale model can be then tested against real life model in further studies. It is also to experiment if the probabilistic approach for backtracking the pollution source can be a tool for the future of early warning system of WDS.

1.5 Methodology

This study will experiment with a small water distribution network using EPANET software to simulate the scenario of contamination events. The simulated results and the data will then be analyzed to find the best location for the sensors to be placed within the network. Once the software simulation is carried out, the pilot scaled WDS with sensors placed at optimal nodal junctions will be built and test with some chemical intrusion to analyze the response from real time scenario of contamination events. The WDS will consists of gauges to measure necessary hydraulic properties like pressure gauge, a continuous supply of water, valves to control outflow (resembling consumption), junctions to insert sensors and inject intrusions, and data logger. The basic idea is to inject some type of solvent into the water system and then monitor the time of detection from sensors.

Simple sensors for detecting undesired water quality (e.g. pH) and the set of data from all the sensors placed in the modelled WDN, we can observe the pattern of detections which may benefit us for prediction of the source of contamination.

1.6 Limitation

A full scale water distribution network is much more complex system than the experimental pilot scaled model used in this research. Therefore, some obvious limitations such as complex pipe network, variation of elevation, change of pipe properties etc. which are not considered in the scaled model. For a small pilot scale model and the research scope, the contaminant transport pattern was not analyzed as more advanced sensors and computational system are required. EPANET is the only simulation software that will be used for conducting this study

This page is intentionally left blank

2

LITERATURE REVIEW

Researchers around the world have conducted some extensive study over the past decades on the optimization of sensors placement and test effectiveness in the water distribution network. Such researches mainly focused on the real-world models and run simulations with the collected data to define the event of uncertainties like contamination infiltration or leakage in the water distribution system (WDS). It is possible to analyze data of water quality models that can be integrated with the hydraulic models to simulate the dispersion of contaminant particles and changes of water quality within the distribution network (Savic et al. 2009). The data generated can then be used in the simple optimization algorithm methods such as Genetic Algorithm or Greedy Algorithm etc. (Nicklow et al., 2010). However, experiment of a pilot scale WDS in presence of several sensors to localize a contamination event and generate an early warning system has not been very common. This chapter briefly discuss previous advancements in the field of sensors placement and optimization in WDN.

2.1 Related literatures

Since the rapid development of computer system in the last few decades, mathematical models and simulation software are being used in designing water distribution network. In the process of optimizing the sensors placement in the WDN, mathematical methods such as Linear algebra and Numerical analysis are useful for modelling water quality, minimizing sensor number and identifying location of placement. In the paper by Rossman and Boulos (1996), use of numerical analysis to determine water quality such as water age and chemical detection was discussed using two different mathematical model approaches, namely Eulerian and Lagrangian. The paper also suggests the use of these two approaches for different scenarios; for example, Eulerian allows less memory usage when it comes to modelling water age while Lagrangian method is quick to analyze. Moreover, linear algebra has contributed to the water quality modelling and sensor placement by reducing the number of nodes from the simulation which are unlikely to be contaminated during an injection event. Using linear algebra to reduce the nodal junctions to only possible contamination injection location, and then using minimum relative entropy (MRE) to compare posterior and prior information distribution, the minimum entropy distance to prior distribution is considered as location of contamination event (Propato, Sarrazy, & Tryby, 2010).

From all the optimization studies in past few years, issues regarding the water distribution network are associated with the size of the network which relates to data processing time and storage space. Network with thousand possibilities of contamination event will generate data from the sensors which may take hundreds of days to reach to conclusion (Ostfeld et al., 2008). According to the battle of the water sensor networks (BWSN); in large network of water distribution system, one of the major concerns is infinite probabilities of scenarios with varying objectives such as population affected, time of detection and locations of contamination intrusion. Hence using optimization techniques and algorithms can help reduce the processing time to generate early warning signal. (Ostfeld et al., 2008). It is also difficult to generalize a system or sensor placement method for every network as each has unique physical and hydraulic properties.

The initial problems one need to consider with the sensor placement in water distribution for contaminant detection are sensitivity of the sensors and concentration of contamination. As the sensors are device which can be defective or broken, the detection of contaminant can be false. In addition, the low concentration of the contaminant can pass through the sensors without being detected hence affecting the optimization. The false positive or false negative results generated by sensors should be supervised carefully before running simulation or experiments. In such scenario it is possible to include a mathematical expression in the simulation algorithm where the faulty sensors are not considered during the runtime or a minimum concentration limit/value is set to generate positive detection of contaminant (Krause et al., 2009).

In recent times, when dealing with optimization of the water distribution network for sensors placement, few algorithms and network theories are most efficient. Greedy algorithm (GA) and Genetic algorithms are mostly used for optimization as seen from previous research works. The book *Introduction to algorithms* (Cormen, *et al*., 2009) defines genetic algorithm as "A greedy algorithm always makes the choice that looks best at the moment. That is, it makes a locally optimal choice in the hope that this choice will lead to a globally optimal solution." It is difficult to explain how the algorithms here but for a large set of data such as contaminant concentration from nodal junctions of WDS, greedy algorithm can minimse the locations for placing sensors. Sela Perelman *et al.* (2016) optimizes a large scale WDS for sensors placement using greedy algorithm to identy the location of pipe failure.

In comparison to greedy algorithm, Genetic algorithm has performed better in optimization of sensor placement localization. Genetic algorithm is defined as a heuristic approach search derived from the principles of Darwin's theory of evolution by John Holland in the year 1960, later modified further by David E. Goldberg in 1989 (Sadeghi, *et al*., 2014). The basic process in the genetic algorithm can be seen from the Figure 1 taken from the book *Genetic Algorithm Essentials* by Kramer (2017). Due to complexity to explain, the processes and terms are not defined in detail in this paper. Genetic algorithm has also wide range of use in water resources planning and management such as in wastewater treatment to configure treatment processes, hydrologic and fluvial modelling to generate solutions for cost and detention system water-shed regions e.g. storm water detention (Nicklow et al., 2010).

Algorithm 1 Basic GENETIC ALGORITHM

```
initialize population
repeat
    repeat
        crossover
        mutation
        phenotype mapping
        fitness computation
    until population complete
    selection of parental population
until termination condition
```

Figure 1: Basic processes of genetic algorithm

In the paper by Krause *et al.* (2009); an optimal solution for the sensor placement in large water distribution network (>12000 nodes) is presented. As stated in the paper it is guaranteed to obtain 63% optimal solution with offline bound which means that the analysis is carried out with without running any algorithm such as greedy algorithm (GA) and reduced time step of 5 minutes. Moreover, with online bound where algorithms are used for processiong large number of sensors, it is possible to get approximately 95% optimal solution for the sensors location.

Dorini, *et al*. (2010) presented an interesting paper on the sensor placement issue. Sensors Local Optimal Transformation System (SLOTS) is promising methodology to perform better than the greedy algorithms (GA) methodolgies used in the BWSN; for both single-objective and multi-objective scenarios. The SLOTS methodology considers a sensor from a set of sensors and move around from one location to another when all other sensors are stationary. As the sensor move, it continuously measures performance (objective functions for single or multi) improvements until the best improvement scenario is reached. All the sensors go through such procedure until all the sensors find their location for best performance improvement. This SLOTS algorithm seems to perform better to obtain 95% optimal solution and dominates many greedy algorithms tested for sensor placement in BWSN.

It is so far understood that mathematical models and computer algorithms are more preferred for the optimization problems. An extraordinary work by Rathi, Gupta and Ormsbee (2015) - *A review of sensor placement objective metrics for contamination detection in water distribution networks* - many different methods were discussed revolving sensors placement optimizations and compared work of other researchers in the field. This paper also provides significant knowledge of different criteria or objectives that are taken into consideration when dealing with sensors placement and optimization in the water distribution network. Similar to the multi-objective study of the Sankary and Ostfeld (2017), the paper explains *Demand coverage, Time of detection, Population exposed, Extent of contamination, Volume consumed, Detection likelihood and Risk of contamination* etc. The simplification of the methodologies followed in the different study by different researchers in the paper is praiseworthy. Remarks made in the end for different strategies methodologies are issues related to the network size considerations (large or small, number of nodes etc.); modelling the contamination event (type of contamination, dynamic and static simulation etc.); and use of computer programming for optimization.

In the research conducted by Sankary and Ostfeld (2017a), brought innovative method of using both fixed and inline/mobile sensors within networks of three different sizes: small, medium and large. The method involved mobile sensors where the following assumptions were made for the sensors in the pipe network: 1) the sensors travel along the flow and with same flow velocity; 2) the sensors has no impact on the hydraulic operations in the network; 3) uninterrupted wireless signals from the pipes to surface transceiver beacons and 4) sensors are able to detect contamination continuously when the contaminant concentration is above 0.1mg/L. The data generated from the fixed and mobile sensors an augmented messy Genetic Algorithm was initiated for optimization. The results provide important insights on how cost is key factor in developing early warning system considering the multiobjectives in WDN.

Sankary and Ostfeld (2017a) has contributed significantly to the development of early warning system within the WDS. In the paper *Inline Mobile Sensors for Contaminant Early Warning Enhancement in Water Distribution Systems,* discusses the implementation of mobile wireless sensor network (MWSN) to test the effectiveness of inline sensors in the water pipes. The inline sensors are capable of monitor water quality and trasmint the data wirelessly. The research analyzes the data of both inline mobile sensors and fixed sensors for better optimization of contamination detection taking population affected criteria as well as the total system costs to operate the technology. Results show lower battery life of the mobile sesnors is key factor to affect the detection performance, hence higher number of sensors are required to cover large network to obtain effective results. Quality and efficiency tranceiver beacon used for the mobile sensor greatly depends on their costs; expensive transceiver enhance the performance of the sensors but lowers the number of use if budget is limited. Running messy Genetic Algorithm (mGA) for both mobile and fixed sensors simulation, fixed sensors is still preferable for reliable and accurate measurements when it comes to cost effectiveness. Another research by Sankary and Ostfeld (2017b) discussed operational uncertainty in developing early warning system and the performance with respect to cost of implementation,

To develop an early warning system in the WDN some objectives are focused which are important to consider for simplifying the model efficiency. The objectives which are key for water network contamination event are Time of Detection, Detection likelihood, Population affected prior to detection. According to the journal article *Scaled Multiobjective Optimization of an Intensive Early Warning System for Water Distribution System Security* (Sankary and Ostfeld, 2017c), interesting comparison between these different objectives were demonstrated. The literature also compared cost of system implementation and maintenance with respect to the all the other objectives. However, the findings of the research suggest that cost effectiveness of fixed sensors compared to mobile sensors. Performance wise, inline mobile sensors are not cost efficient when used implemented for continuous monitoring where fixed sensors are performing better with low cost of implementation and maintenance. When it comes to optimization of the network, the messy Genetic Algorithm (mGA) performed very well in small networks but large networks are still complicated to process and generate optimized solution and requires more simulation runtime and computing storage.

One of the recent studies show that it is possible to optimize and allocate source of contamination considering the detection time and duration of contamination and using backtracking algorithms to find the potential source (Ung et al., 2017). In this case, the sensors will generate only binary responses; positive or negative and create a matrix for only positive responses. The paper mainly focuses on the accuracy and the specificity percentage to find the source of contamination using the evaluation method from (Seth et al., 2016), which is shown below.

$$\text{Accuracy } (\%) = \frac{\text{Likeliness measure of the true injection node}}{\text{Highest likeliness measure over all candidate nodes}} \times 100$$

$$\text{Specificity } (\%) = \frac{\text{Number of nodes with lower likeliness than true injection node}}{\text{Total number of candidate nodes}} \times 100$$

In addition, a Monte Carlo simulation method was used to assist Greedy algorithms to sort out suitable sensor location. One of the fascinating work was done by SafeWater Project funded by European Union (EU) to develop a full-fledged contamination event detection and management system (Deuerlein, Meyer-harries, & Guth, 2017). The research is conducted by 9 different organizations contributing to different aspects of the project. The event detection system developed consists of sensors which can detect chemical, biological, radiological, and nuclear (shortly known as CBRN) contamination and can raise an alarm in almost real time. In the event management tool, another specialty of the project, allows user to contribute contamination data using the online system. The data can be immediately be reviewed, and experts can carry out actions (e.g. closing of valve, or safely channel the contamination out of the system) from the same portal.

In the recent researches on sensor location and optimization, computer algorithms and concept of parallel computing is more efficient. To help optimization and reduce the time and storage issue of large data management, cloud computing parallel to physical simulation have benefited developing EWS. The concept of parallel computing allows several different computers or components working individually but coordinated to produce/analyze one specific criteria. Since each computer/component is working individually with specific task, it becomes faster to process and accumulate the required data.

In more advanced research using cloud computing in parallel to genetic algortithm has produced significantly better results that meet the objectives of the study. MapReduce is a programming model which can be run in parallel to genetic algorithm to reduce the runtime of simulation with optimized suggestion of sensor placement (Hu, Tian, & Yao, 2015). Nevertheless, MapReduce requires enormous computing resources as the readability and writability of data undergo repetition of previous processed data to generate the next data. Due to this data process framework, MapReduce tend to present similar execution time for different cluster data size.

In the field of contamination or leakage detection in WDS, the cost of sensors plays an important role for the necessity to optimized sensor placement. This lead to the objective of reducing the number of sensors in the network without the hindering the efficiency of contamination detection. However, large water distribution network can generate huge data from thousands of sensors which becomes problematic to process quickly and the storage space required is not feasible most of the cases (Hu et al., 2017). Many algorithms have been deployed and it is observed to be efficient in small and medium scale WDS (see the Spark based algorithm). In latest researches genetic algorithm has developed and succeeded to improve the optimization of the sensor placement in the WDN.

Spark based genetic algorithm has overcome the issues of evolutionary genetic algorithm generating better and results for optimization. In comparison to MapReduce, Spark can execute given tasks considerably a lot faster due to the different data process and management of the two programs. According to Hu *et al.* 2017; how MapReduce process data by reading through previous data takes more computational resources as well time whereas Spark uses resilient distributed datasets (RDD) which is capable of storing data in the cache memory and algorithms can perform simulation efficiently. However, MapReduce or Spark are used parallel to the Genetic Algorithm contributing to the objective of reducing the data process time and optimize sensor location.

The extensive researches on sensors placement in water distribution network using mobile sensors, fixed sensors and optimizing with different algorithms has mentioned the complexcity of computational resources, time constraints to process data and cost effectiveness. To simplify the task, algorithms are focusing on graph theories and complex network theories for more stochastic approaches for studying the sensor placement issue. A study on the application of graph theory was conducted by Perelman and Ostfeld (2013) mentions how the water utilities can be benefited using the graph theory methodologies to suggest sensor placement location by analyzing available information on network topology and demand loadings.

At the international conference Hydroinformatics 2018 (HIC2018) some recent works on the safety of the water distribution networks was presented. One interesting work was on identification of critical components in the distribution network using time-dependent data (Aydin, 2018). The research explains the utilization of graph-based analysis of the network, called *entropic degree*, using flow paraments for longer period of EPANET simulation. Weighing on the flow in the links of the network, the critical location can be identified. This can benefit related authorities and decision makers to develop strategic plan to reduce the effect during the time of crisis such as natural disasters, deliberate attacks and intrusion etc.

In another research presented at the HIC 2018, *centrality matrics* are used to explain how nodes and links in a distribution network can be optimized using Complex Network Theory (CNT). In evaluating centrality matrics, important aspects of WDN considered are *closeness* and *betweenness* as these two helps the concept of short path analysis (Simone et al., 2018). Here the closeness is defined as the measurement of distance between the nodes and shortest distance relates to closeness between the nodes. The betweenness concept discussed in the paper by Simone *et al.* (2018) relates to the fraction of shortest paths/links reaching a certain node and the node having highest number of shortest path fraction is considered to have highest betweenness centrality. Another method that was taken into consideration was degree and neighborhood degree; which explains the links connected to a node and the nodal connectivity between the neighboring nodes. The combination of these centrality study allows the possibility to identify the most vulnerable location of the network and help decision makers for management purpose.

Most interesting research mentioned in the HIC 2018 conference was focused on the sensor placement using graph theory to locate critical sensors location with different scenarios taken into consideration. The Graph Spectral Techniques (GST) is suitable for sensors placement location as it does not require extensive computation resources (Nardo *et al.*, 2018). The main advantage of this technique discussed in the article is that no hydraulic parameters or flow dynamics data are required. The study also focuses on the complex network theory (CNT) to optimize the sensor locations using eigenvector centrality to identify the important nodes from the list of node ranking. To analyze the methodology, the WDN was evaluated with weighted (where graph geometric characteristics and other flow dynamics and pipe parameter were considered) and un-weighted (where only considers the connectivity of the nodes in the network) values. Comparing the two studies with genetic algorithm (GA) optimized WDN to find the similarities in locating sensors placement.

The backtracking of source contamination is adjoint to the chemical or particle transport equations. It computes the contamination transport in reverse time, beginning at the sensors, to enumerate potential nodes of upstream contamination. Use of the backtracking methods allow handling large-size networks that would require huge calculation time if performed with a forward scheme.

2.2 Conclusion

The literature study here for identifying sensor placement location has evolved from mathematical point of view to more probabilistic and graph theory. The shift mainly happened due to the unavailability of the proper data, uncertainties in the network and hydraulic activities, extensive mathematical models and varying results in optimization algorithms.

The complexicity of the mathematical models and different algorithm application has led to the development of more stochastic analysis for optimization of the sensor location placement.

Even though mathematical analysis tends to be more accurate, the probabilistic approaches can give range of possible outcomes which needs expert investigation to pinpoint strategic location of sensor placement.

Moreover, the high computational resources required to analyze a large water distribution network is also a key factor in this regard whereas graph theory methods are more simpler, easy to analyze and quick for decision-making purposes in the management sectors. Finally, the method adopted for this research study tries to focus on the recent studies for sensor placement locations with similar analysis like graph theory and finding the critical node using ranking technique. Later, the results from this literature study will be physically tested in a pilot scaled WDN model built for this master thesis project.

This page is intentionally left blank

3

MODEL & METHODOLOGY

In this chapter the planning and the process of carrying out this research study will be discussed. To simplify the methodology of the study, the research work was divided into three phases.

Phase 1: Developing pilot scaled WDN model;

Phase 2: Optimization and construction of WDN pilot model;

Phase 3: Experimental method and backtracking and source location identification.

3.1 Developing pilot scaled WDN model

In the beginning of the study a suitable simple design layout of water distribution network was developed. Understanding the available space and resources in the laboratory, the measurements and system to build the network was established. Therefore, the following network (Figure 1) was designed and then tested using software called EPANET.

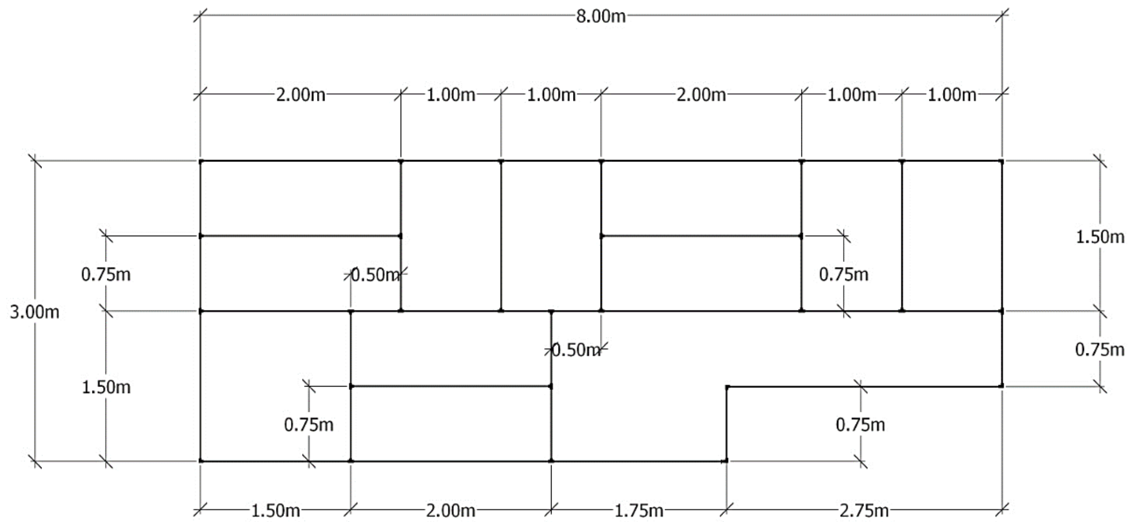


Figure 2: Design for lab scale water distribution network for experimentation.

The 50 m (approx.) pipe length in a looped pilot scaled WDN consisted of 31 nodes/junctions and 44 links/pipes with a continuous supply of water. To build the pilot scale model for the experimental study of backtracking contamination in the WDN, a laboratory facility of 16m X 6m was used. For the benefit of future calculations, the dimensioning of the components was set as multiple of 0.25 m. The scaled pipe model was built up with the dimension 8m X 3m on the concrete floor with fixed wooden supports.

As for the materials and components used in the network, PVC pressure pipe with diameter 40mm were used throughout the network (see photo of the system in Figure 5). To control the flow of water in the system, three close valves were placed to control the outflow rate representing consumption. Another close valve was placed at the junction of the inflow to control flow rate in the network. Several up-righted T-junctions that connect the pipes were installed to act as intrusion nodes of contamination and sensors locations. These T-junctions locations were determined using the data generated from simulations in EPANET, then optimized using simple sorting matrix. More regarding the placement of sensors and intrusion point will be discussed in later in this chapter.

EPANET is a common analysis tool for modelling water distribution network. EPANET is developed by the US Environmental Protection Agency (USEPA). EPANET can simulate hydraulic characteristics and water quality in network of pressured pipes (USEPA, 2000). The software is simple and has user friendly interface to maneuver through different options and settings. To setup a water distribution model in EPANET there are functions like junctions (nodes), links (pipes), pumps, reservoir and tanks etc. to be used and each function can be specified with desired parameters and inputs. It is possible to simulate the model for any desired period of time, in hours or days, and produce results in the form of graphs or tabular data sheet. A result report, for the whole process of simulation, can also be obtained in the program in the form of text file.

3.1.1 EPANET simulation

In the EPANET software, the WDN layout (Figure 2) and the hydraulic properties were set up for the whole simulation period. An important setting in the EPANET is the headloss formula. Three different head loss equations are available; Hazen-Williams, Darcy-Weisbach and Chezy-Manning. All three equations use the following formula with different resistance coefficient (A) and flow exponent (B).

$$h_L = Aq^B$$

where h_L = headloss (friction loss in pipes), q = flow rate (Volume/Time). The expressions and related information used for calculating resistance coefficient (A) and flow exponent (B) are given in Table 1 and Table 2 (Agency, 2000). In this study Darcy-Weisbach equation was used for the whole simulation period.

Table 1: Headloss formula for pipes with pressure flow (headloss is in feet and flow rate is cfs).

Formula	*Resistance Coefficient (A)*	*Flow Exponent (B)*
Hazen-Williams	$4.727\ C^{-1.852}\ d^{-4.871}\ L$	1.852
Darcy-Weisbach	$0.0252\ f(\varepsilon,d,q)d^{-5}L$	2
Chezy-Manning	$4.66\ n^{2}\ d^{-5.33}\ L$	2

Notes: C = Hazen-Williams roughness coefficient
ε = Darcy-Weisbach roughness coefficient (ft)
f = friction factor (dependent on ε, d, and q)
n = Manning roughness coefficient
d = pipe diameter (ft)
L = pipe length (ft)
q = flow rate (cfs)

Table 2: Roughness coefficients for new pipe.

Material	*Hazen-Williams C (unitless)*	*Darcy-Weisbach ε (feet x 10^{-3})*	*Manning's n (unitless)*
Cast Iron	130 – 140	0.85	0.012 - 0.015
Concrete or Concrete Lined	120 – 140	1.0 - 10	0.012 - 0.017
Galvanized Iron	120	0.5	0.015 - 0.017
Plastic	140 – 150	0.005	0.011 - 0.015
Steel	140 – 150	0.15	0.015 - 0.017
Vitrified Clay	110		0.013 - 0.015

The other hydraulic properties and settings used for the simulation are as follows:

Flow Units: Liter per second (l/s)
Specific Gravity: 1
Relative Gravity: 1
Pipe Roughness: 0.2
Relative viscosity: 1

In the initial stage of the simulation, effects of flow pattern such as day and night consumptions were taken into consideration. A simple flow pattern set using a 4 hours' demand time pattern with minimum of 0.5 l/s and maximum of 1.2 l/s. The simulation run time was set to 24 hours in the beginning to check flow of contamination within the network. Due to the size and flow of the network, the contamination spread happens only within first 30 minutes when there is outflow. However, due to certain limitations in EPANET in using time settings, the flow pattern was omitted from the analysis study. The issue and limitations of EPANET are thoroughly discussed in the section *4.3.2 Usability of EPANET software*.

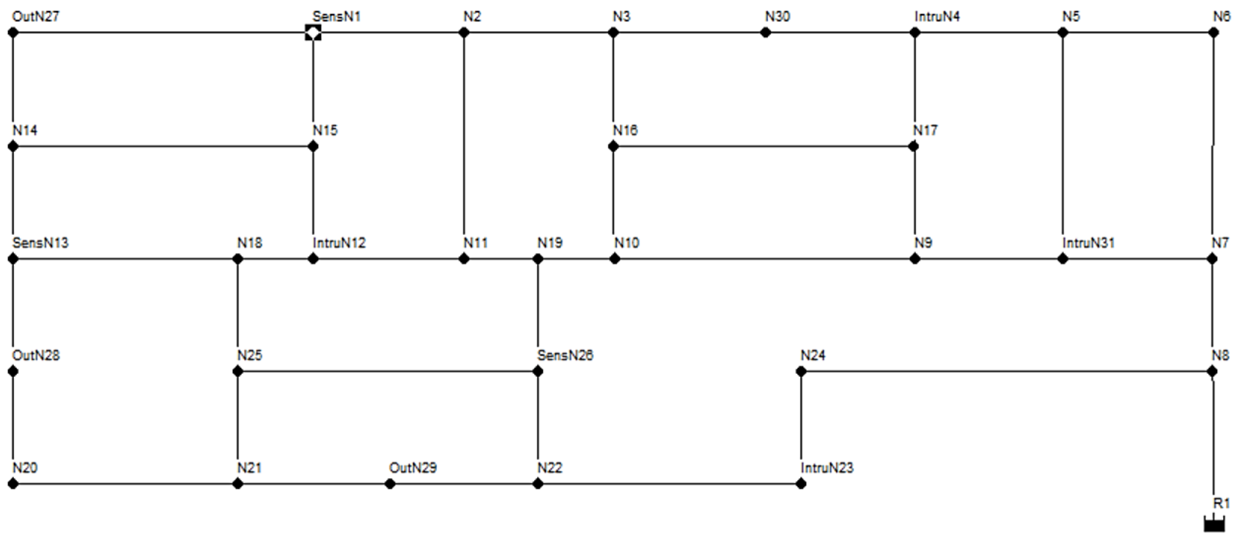


Figure 3: Lab scaled model in EPANET where the nodes and reservoir are shown. The reservoir acts as a continuous supply of water.

3.1.2 Checking performance of EPANET model

Running the simulation model, we first tested if the system was working perfectly without showing any errors such as negative pressures. Once the system was performing error free, the outflows with varying discharge rates were set and check again for errors. Since no errors were generated due to the introduction of the outflows, simple sorting matrix were used to to find sensors placement and intrusion point. Sorting matrix is simple approach to rank contaminated nodes according to the concetration received after each intrusion scenario. The **Table 3** represents all the node concetration with respect to intrusion node. The top row shows the intrusion nodes and the columns shows the concentration received by other nodes. Since the node where intrusion is taking place shows same concentrarion as injected, they were not considered when developing the ranking table shown in **Table 4**.

A reservoir with continuous flow with certain water ead was considered in EPANET as in the physical model continuous water flow from the tap was considered as inflow to the network. This allowed to eliminate most of the negative pressure errors in the pipes and nodes. Initially flow pattern was considered with varying flow for day and night but since EPANET is not dynamic and chemical transport is not time-variant, a constant inflow and demand flow was set. Moreover, no significant change in the chemical transport in the pipes was observed for different flow pattern. Therefore, the flow pattern was adjusted to constant instead of day and night variation.

3.2 Optimization and construction of WDN model in EPANET

For optimizing the sensor placement in the network (Figure 2), contamination in the all the nodes was introduced and simulated the effects in other nodes and pipes as seen in Table 3. Using Quality tool in the EPANET with 5 mg/l chlorine (with the source of contamination set as *Mass Booster*). For the reaction settings of contamination, the coefficients for pipe surface were adjusted as shown in Figure 3 below. The following settings are given in the EPANET manual and set as standards.

After generating concentration values in all the nodes corresponding to intrusion nodes, matrices of “Nodes being contaminated” vs “Intrusion Node” was created. Moreover, to investigate the validity of the contamination spread, the model was simulated using Trace tool to understand the percentage of chemical spread throughout the network from the source intrusion. The same methodology was followed to make another matrix table for the trace values where the percentage of chemical spread was considered instead of nodes being contaminated. Both the concentration and the trace matrices provided same outputs, hence only concentration data were considered to include in this report as it will be easier to relate with physical model data (pH values).

It is important to note that during the creation of the matrix table, or when running the simulation for quality or trace tool, the output data statistics were set to “Range” which takes the difference between maximum and minimum values detected by the nodes. This gives the opportunity to measure the most affected or contaminated nodes during the intrusion event at an individual node, hence we can consider them as location for sensor placement. The quality settings, reactions parameters and time properties for simulations in the EPANET are shown in the Figure 3.

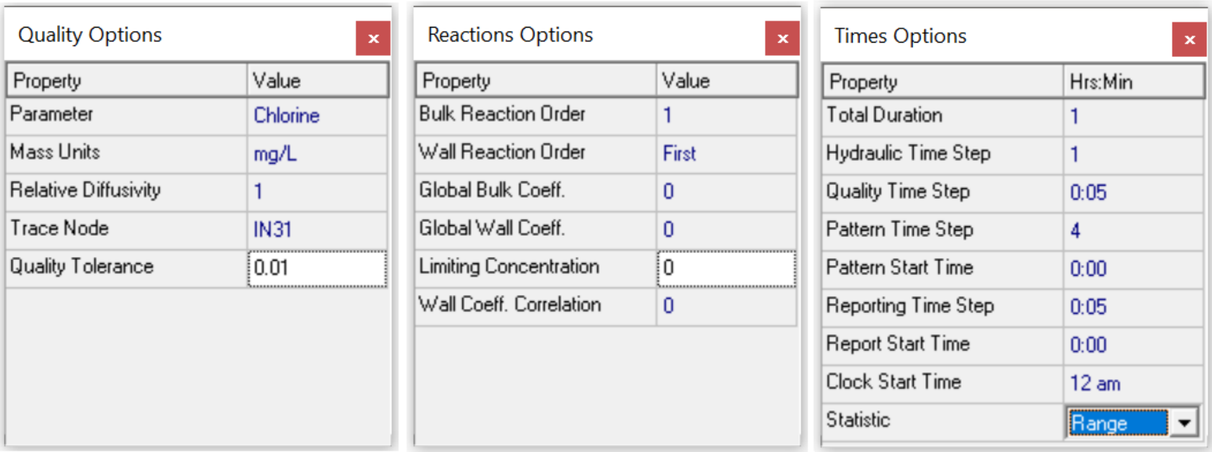


Figure 4: The quality setting for the chlorine intrusion in the EPANET model.

To process the available set of data from the 31 nodes are now processed using MATLAB coding where the concentration of was converted to Boolean values (True/False) or binary digits (0/1). To get results in binary digits, a logic was set in the MATLAB coding where the concentration values that falls in certain ranges are converted to “1” and rest are converted to “0”. The range of concentration values considered in the experiment are as follows:

1. 1 mg/l < C_x < 2 mg/l where, C_x = Concentration of chlorine
2. 2 mg/l < C_x < 3 mg/l
3. 3 mg/l < C_x < 4 mg/l
4. C_x > 4 mg/l

The generated data from the EPANET model for all the 31 nodes where used to optimize two aspects of the objective of sensors placement. 1) To understand and localize the placement of sensors, and 2) to fix certain location of intrusion nodes. The former aspect will provide information about where the sensors should be best placed, and the later will help to determine the flow of chemical through the sensor nodes so to investigate backtrack methodology.

3.2.1 Localizing sensors nodes

To localize the sensors placement, ranking of the most contaminated nodes was performed. The four sets of the concentration range mentioned above were analyzed to calculate the cumulative value for the all the sensors to find out which nodes are being affected more often. Therefore, the node that receives contamination or being contaminated more times due to individual intrusions among the 31 nodes, has the higher probabilities of detecting water quality change, hence suitable place for sensors. In addition, strategic analysis and decision was also conducted in order to define whether the placement of the sensors according the higher sensitivity values gives perfect detection. For example, N19 (Table 4) has higher rate of sensitivity when intrusion occurs in the upstream of the network; but intrusion at N23 is not detected in N19. Therefore, strategically N26 is selected as it suits the best option for intrusion in upstream and at N23. Strategic decision making was also made when comparing nodes N20 and N26 where, N20 is furthest in upstream and close to the outflow at N28. Since the major objective of the study is to back track intrusion location before reaching the consumers, sensor node placement is considered to be little far from consumption. The **Table 4** below shows the ranking of the nodes at different range of concentrations and the chosen nodes highlighted.

3.2.2 Localizing intrusion nodes

The node data from all the 31 nodes were converted to binary digits with the condition that any concentration above 1 mg/l will be True otherwise false. The resultant binaries are added to investigate which intrusion node contaminates the other nodes the most. For example, if you see the node N8 (in Figure 2), since it is connected to reservoir, whole network will contaminated resulting all nodes affected and get the highest value of 31. However, since the main concern of this study is to identify the location of the intrusion through back tracking, nodes that are further in downstream are rejected from selection. Therefore, the second best node was selected which is node N31 as is not directly connected to the reservoir but has higher spread of contamination in the network with a value of 23 (Table 3).

The remaining two placement of intrusion nodes were selected according strategic decision-making mindset and the placement of consumption nodes. Between nodes N4, N5 and N6; the node N4 seems to be of better option as it is not very much in the upstream. Also N5 and N31 has similar impact on the network being at the same alignment. The node N23 was selected to investigarte the sensor at N26 and outflow at N29, since not many nodes have any affect on the outflow node N29. With these analysis and decision making for intrusion nodes, it is assured that every chemical intrusion will pass through at least one sensor location.

Table 3:Nodes being contaminated Vs Nodes intruded with chemicals for all the nodes in the network. Data from EPANET using 5 mg/l chemical concetration.

Intrusion	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20	N21	N22	N23	N24	N25	N26	N27
Node ID	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**	**Conc.**
N1	5.03	5.04	4.95	3.51	3.51	3.02	3.51	3.51	1.48	0.21	0.21	0	0	0	0	4.83	4.83	0	0.21	0	0	0	0	0	0	0	1.47
N2	0	5.03	4.95	3.51	3.51	3.02	3.51	3.51	1.48	0.21	0.21	0	0	0	0	4.83	4.83	0	0.21	0	0	0	0	0	0	0	1.47
N3	0	0	5.03	3.66	3.66	3.15	3.66	3.66	1.36	0	0	0	0	0	0	5.04	5.04	0	0	0	0	0	0	0	0	0	1.35
N4	0	0	0	5.01	5.02	4.31	5.01	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.71
N5	0	0	0	0	5.01	4.31	5.01	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.71
N6	0	0	0	0	0	5.01	5.01	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N7	0	0	0	0	0	0	5.01	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N8	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N9	0	0	0	0	0	0	5.01	5.01	5.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.02
N10	0	0	0	0.55	0.55	0.48	4.25	4.25	4.27	5.02	0	0	0	0	0	0.76	0.76	0	0	0	0	0	0	0	0	0	4.26
N11	0	0	0	0.55	0.55	0.48	4.25	4.25	4.27	5.03	5.02	0	0	0	0	0.76	0.76	0	5.03	0	0	0	0	0	0	0	4.26
N12	0	0	0	0.55	0.55	0.48	4.25	4.25	4.27	5.03	5.04	5.03	0	0	0	0.76	0.76	0	5.03	0	0	0	0	0	0	0	4.26
N13	1.29	1.29	1.26	1.16	1.16	1	2.04	2.03	2.04	2.41	2.31	2.31	5.02	0	2.46	1.6	1.6	2.6	2.41	0	0	0.18	0.18	0.18	0.28	0.28	2.04
N14	2.66	2.66	2.59	1.85	1.85	1.59	2.01	2.01	2.02	2.38	2.38	2.39	0	2.46	5.06	2.55	2.55	0	2.38	0	0	0	0	0	0	0	2.02
N15	2.66	2.66	2.59	1.85	1.85	1.59	2.01	2.01	2.02	2.38	2.38	2.39	0	5.04	5.04	2.55	2.55	0	2.38	0	0	0	0	0	0	0	2.02
N16	0	0	0	3.66	3.66	3.15	3.66	3.66	1.36	0	0	0	0	0	0	5.02	5.04	0	0	0	0	0	0	0	0	0	1.35
N17	0	0	0	3.66	3.66	3.15	3.66	3.66	1.36	0	0	0	0	0	0	0	5.02	0	0	0	0	0	0	0	0	0	1.35
N18	0	0	0	0.51	0.51	0.44	3.96	3.96	3.97	4.68	4.49	4.49	0	0	0	0.71	0.71	5.03	4.68	0	0	0.35	0.35	0.35	0.55	0.55	3.97
N19	0	0	0	0.55	0.55	0.48	4.25	4.25	4.27	5.03	0	0	0	0	0	0.76	0.76	0	5.02	0	0	0	0	0	0	0	4.26
N20	0	0	0	0.17	0.17	0.15	1.31	2.74	1.32	1.55	0	0	0	0	0	0.23	0.23	0	1.55	5.02	5.04	2.75	2.75	3.46	4.32	4.32	1.32
N21	0	0	0	0.17	0.17	0.15	1.31	2.74	1.32	1.55	0	0	0	0	0	0.23	0.23	0	1.55	0	5.02	2.75	2.75	3.46	4.32	4.32	1.32
N22	0	0	0	0	0	0	0	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	5.01	5.02	5	0	0	0
N23	0	0	0	0	0	0	0	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.01	5	0	0	0
N24	0	0	0	0	0	0	0	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N25	0	0	0	0.2	0.2	0.17	1.53	3.2	1.54	1.81	0	0	0	0	0	0.27	0.27	0	1.81	0	0	3.21	3.21	3.2	5.02	5.04	1.54
N26	0	0	0	0.2	0.2	0.17	1.53	3.2	1.54	1.81	0	0	0	0	0	0.27	0.27	0	1.81	0	0	3.21	3.21	3.2	0	5.02	1.54
N27out1	4.8	4.8	4.7	3.35	3.35	2.88	3.44	3.44	1.41	0.2	0.2	0.11	0	0	0.24	4.6	4.6	0	0.2	0	0	0	0	0	0	0	1.4
N28out2	0.71	0.71	0.69	0.72	0.72	0.62	1.71	1.71	1.72	2.02	1.27	1.27	2.76	0.24	1.35	0.98	0.98	1.43	2.02	2.27	2.27	1.34	1.34	1.67	2.1	2.1	1.71
N29out3	0	0	0	0	0	0	0	5.01	0	0	0	0	0	1.35	0	0	0	0	0	0	0	5.02	5.02	5	0	0	0
N31	0	0	0	0	0	0	5.01	5.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.01
Total Nodes contamin-ated	6	7	8	20	21	22	25	30	20	16	10	7	2	4	5	18	19	3	15	2	3	9	10	10	6	7	23

Table 4:Ranking of nodes corresponding to the data from all the 31 intrusion nodes and range of chlorine concentration (Cx).

1< Cx <2 mg/l		2< Cx <3 mg/l		3< Cx <4 mg/l		Cx >4 mg/l		Total Number of times Contaminated
Node Number	Number of times Contaminated	Node Number	Number of times Contaminated	Node Number	Number of times Contaminated	Node Number	Number of times Contaminated	
N1	2	N1	0	N1	5	N1	5	12
N2	2	N2	0	N2	5	N2	4	11
N3	2	N3	0	N3	5	N3	3	10
N4	0	N4	0	N4	0	N4	5	5
N5	0	N5	0	N5	0	N5	4	4
N6	0	N6	0	N6	0	N6	3	3
N7	0	N7	0	N7	0	N7	2	2
N8	0	N8	0	N8	0	N8	1	1
N9	0	N9	0	N9	0	N9	4	4
N10	0	N10	0	N10	0	N10	5	5
N11	0	N11	0	N11	0	N11	7	7
N12	0	N12	0	N12	0	N12	8	8
N13	7	N13	10	N13	0	N13	1	18
N14	3	N14	14	N14	0	N14	1	18
N15	3	N15	13	N15	0	N15	2	18
N16	2	N16	0	N16	5	N16	2	9
N17	2	N17	0	N17	5	N17	1	8
N18	0	N18	0	N18	4	N18	5	9
N19	0	N19	0	N19	0	N19	6	6
N20	5	N20	3	N20	1	N20	4	13
N21	5	N21	3	N21	1	N21	3	12
N22	0	N22	0	N22	0	N22	4	4
N23	0	N23	0	N23	0	N23	3	3
N24	0	N24	0	N24	0	N24	1	1
N25	5	N25	0	N25	4	N25	2	11
N26	5	N26	0	N26	4	N26	1	10
N31	2	N31	1	N31	4	N31	3	10

3.2.3 The final layout

Therefore, using the simple methodology of sorting matrix and strategical approach to develop a stochastic model of the sensor placement for the designed water distribution network, the following layout of the final WDN (Figure 4) with sensors and intrusion location was built as scale model and tested in real time experimentation hence enabling the comparison and accuracy with EPANET simulation. Later, the second objective of back tracking source of intrusion model will be tested.

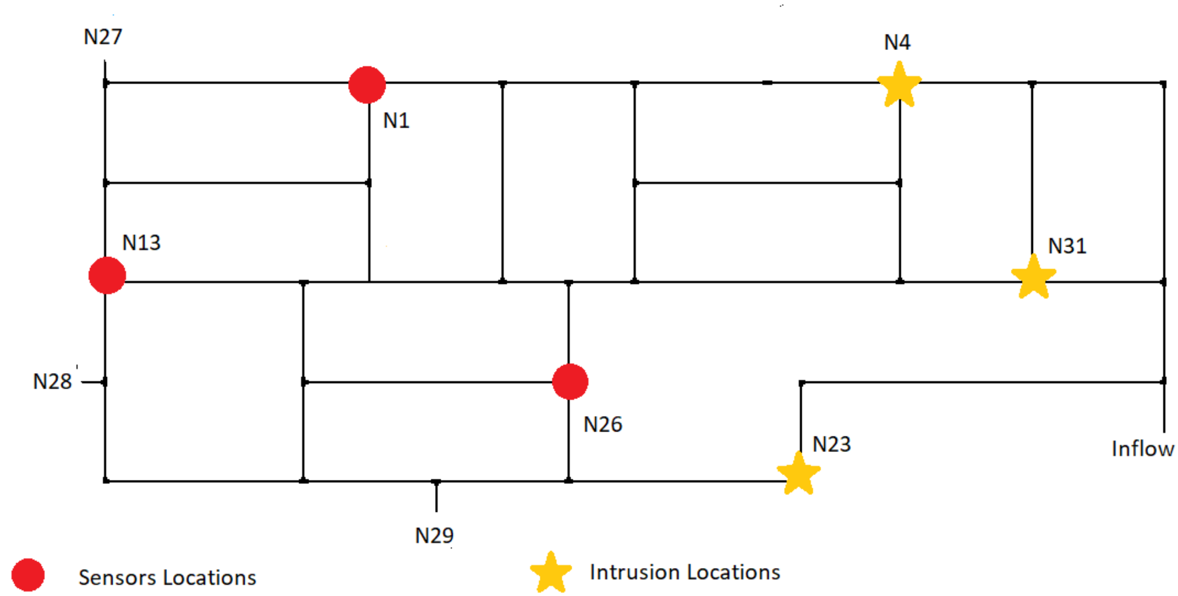


Figure 5: Final layout of sensors and intrusion points in the scale model of WDN for real time experiments.

3.3 Physical model setup

The water distribution network that was designed in EPANET was built in laboratory with the exact dimensions and setup for sensors location and intrusion points as the EPANET model. The laboratory space used was approximately 20m x 8m where a 8m x 3m WDN was installed using PVC pressure pipes, joints, and valves with the possibilities to fit sensors and chemical injections for intrusion scenario (Figure 5).

3.3.1 Materials

The piping material used for the WDN model was Polyvinyl Chloride (PVC) with the maximum working capacity of pressure of 10 MPa. The pipe outer diameter is 50 mm, inner diameter 45.2 mm and thickness of 2.4 mm. The supplied pipes were delivered in 5 meters lengths and cut according to the model dimensions. The pipe pieces were fitted with T-junctions made of same grade PVC and glued using fast adhesive. The whole PVC pipe network was supported with a wooden frame and tightly fitted with mounting clutch/clip (Figure 5).

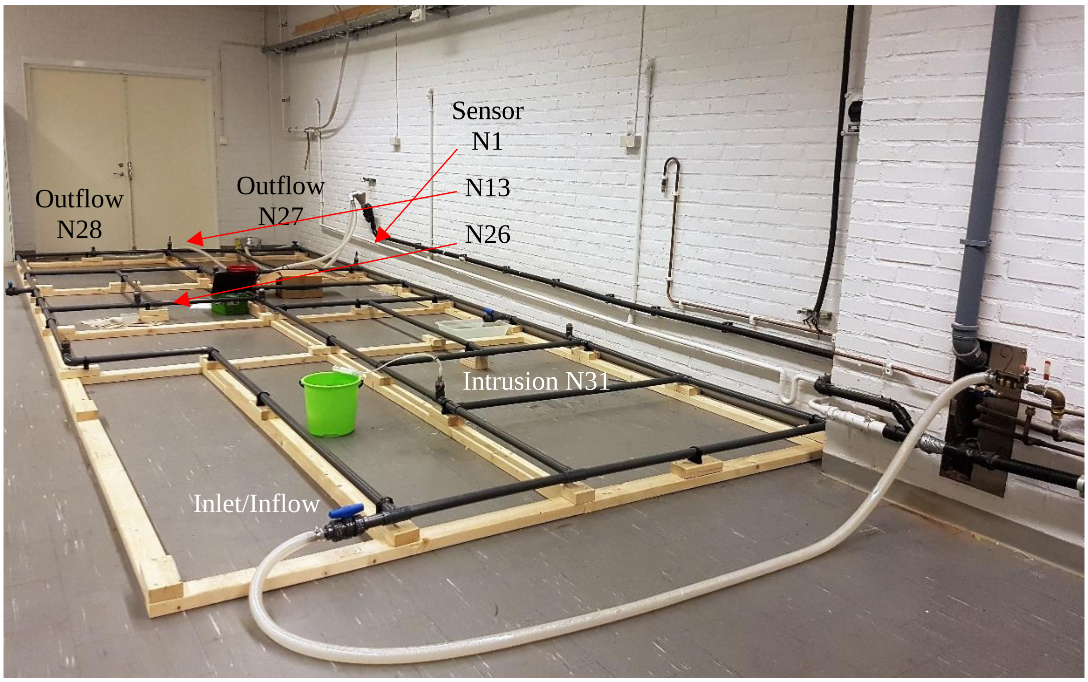


Figure 6: An overview of the laboratory setup of the WDN for the experiment Set 7 (see Table 5).

The valves used for the inflow and outflow are ball valves and flexible PVC hoses were connected to provide water inflow from the source tap and to discharge outflow water to the wastewater system. The source tap was fitted with a ball valve and a screw valve to have precise flow control (Figure 6). The flexible PVC hoses carry water going out from the network to a waste pipe elevated to approximately 1 m above the WDN pipes, therefore gravity plays important role in not clogging the basin and restrict overflow.

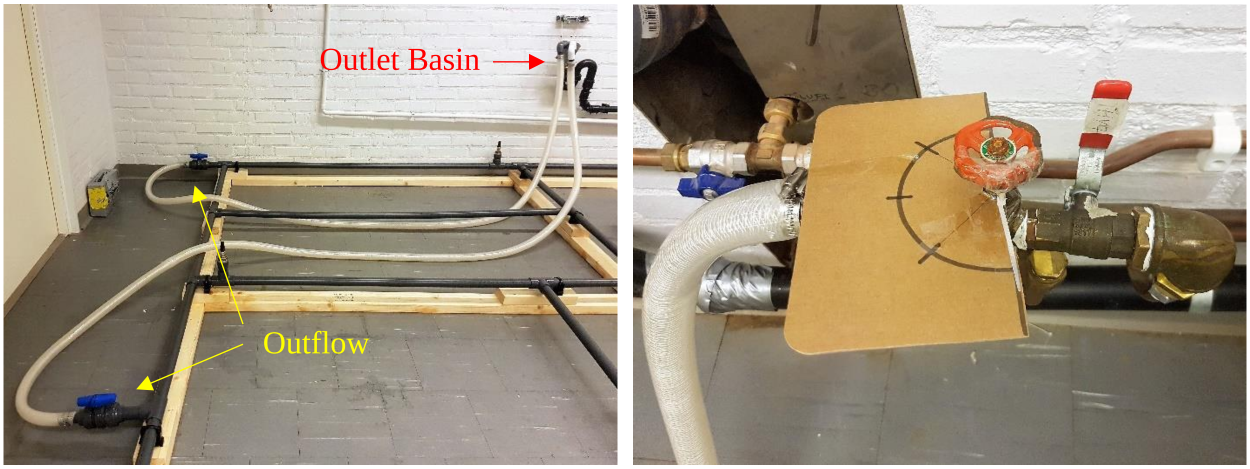


Figure 7: Outflow pipes, outlet basin and the control valves for inflow from source tap.

3.3.2 Transmitter and data logger

In the experiment two transmitters one data logger were used which were connected to the pH sensors. The transmitters are of Sensorex model TX10 (Figure 7), which measures pH and ORP (Oxidation Reduction Potential). The transmitters come with 4-20 mA current output cable and spade lug that connects the sensors. Measurement specifications for the transmitter is -2.00 to 16.00 pH, 0.01 pH resolution, +/- 0.1% accuracy. The measurement from the transmitter is logged into the HOBO 4-Channel Analog logger, a product of Onset (Figure 7). The logger is very easy to use and convenient for such research study. The logger comes with a software HOBOware (available for both Windows and Mac operating system) which helps to configure desired settings. For example, activating an alarm when the sensor reading rises above or falls below a user specified value/measurement.

Figure 8: Sensorex TX10 pH/ORP transmitter and Onset HOBO 4-Channel Analog logger.

It is also possible to set up burst logging in which the logger records data at a different interval during certain conditions. The logging rate can be adjusted from 1 second to 18 hours and type of results output such as calculating minimum, maximum, average, and standard deviation statistics can be set before the start of the logging.

3.3.3 pH sensors

For testing the intrusion events in the physical model of WDN, pH sensors were used which were perfect in response to acidic and alkaline solution. The sensors are by SI Analytics, a brand from Xylem Inc. and are adjustable with the transmitter's lug cable. Two pairs of different sensor models were used in the experiment, but both responded similarly during the tests. The pH sensors used are of model L7781 HD and L8281 HD (Figure 8). Technically the electrode L8281 HD has advanced system compare to L7781 HD in paper, however during the tests for this study, it had not been any significant variation in output of results. More detailed specification can be found in the website of SI Analytics. Two L8281 HD were placed in N1 and N26 while one L7781 HD was placed in N13.

Figure 9: Sensors used in the experiment for measuring pH, product of Xylem Inc.

3.3.4 Contamination/intrusion chemical

To simulate the scenario of contamination intrusion in the WDN, chemical solutions having acidic and basic nature were used and sensor's response were observed. The intrusion chemicals were injected into the system using simple syringe of 60 ml volume (Figure 9). The syringe can be fitted to the node of interest for intrusion. Chemicals used are Acetic acid (24%), Sodium Hydroxide (1 M), Sodium Chloride solution and commercial bleaching solution (Chlorine solution for cleaning).

Acetic Acid performed best out of the four solutions when injected in the network when water is flowing at desired flow rate. When tested for pH of the solutions, Acetic acid had pH of 2.5 (approx.), Sodium Hydroxide had pH 12 (approx.) Sodium Chloride solution had pH of 10 (approx.) and commercial bleaching solution had pH of 9.75 (approx.). Moreover, the basic solutions seemed to be diluted instantly and hence did not rise the pH value in the water, so sensors could not trace the intrusion unlike Acetic acid. Commercial bleaching solution is not suitable for such experiment as it is soapy material and can show unexpected variations in the sensors activity data. The bleaching can also stay in the system for longer period of time affecting the water quality and sensors response activity.

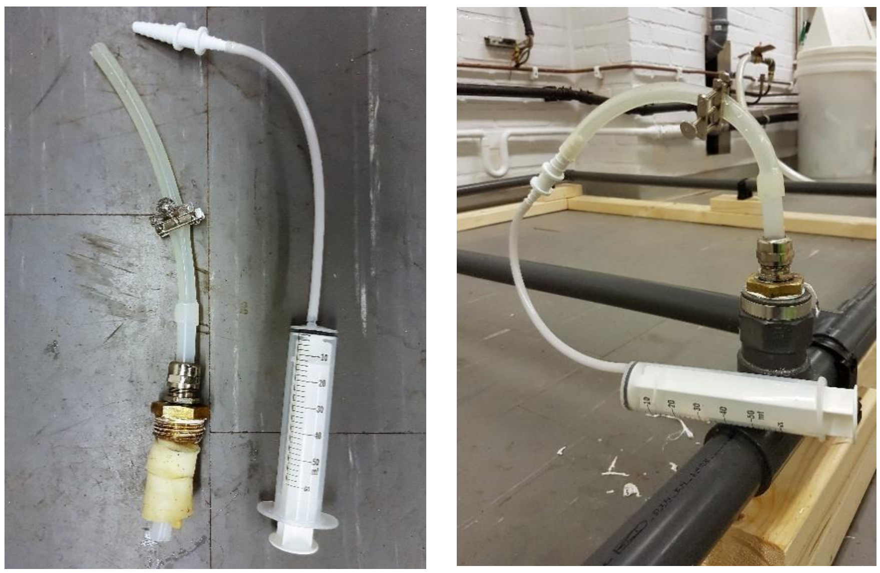


Figure 10: Syringe for injecting chemical into the system and the hardware to fit into the WDN.

3.4 Calibration of the physical model

The construction of the physical model of water distribution network is not flawless and therefore requires calibration to work according to the desired conditions. Since the WDN was constructed by hand and many improvisations were made, there were some issues that need to be discussed for future remedies. The most important issue that appeared during the initial stage of the experiment was with the pH reading. The pH sensors were responding mysteriously, such as showing over or under limit of pH scale value. It seemed that sensitive tip of the electrode was not in close contact of the flowing water. Displacing the sensors electrode vertically further down in the specific nodal junction was a primary solution (Figure 10). In such case, the (white) rubber washer that came with the sensors were removed so that the tip can reach further down into the nodal junction.

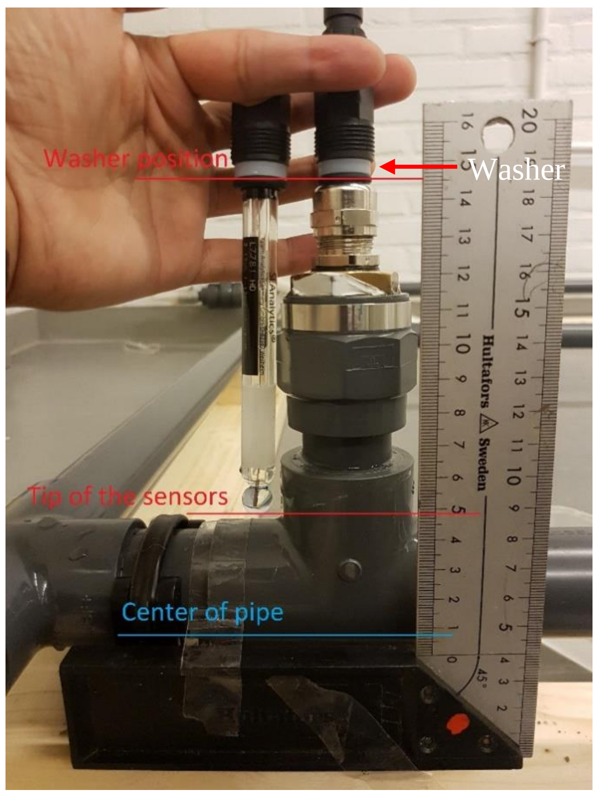


Figure 11: Positioning sensor electrode in the nodal junction of the WDN.

However, the problem of fluctuating pH values and reading over or under pH values did not disappear totally. To solve this issue all the nodal T-junctions, that had vacant space with standstill water, were filled with sponge wrapped in a plastic packet (Figure 11). This avoided stagnation of water or air trapped within the inverted T-junction which could affect the sensor reading or chemical concentration during the experiment. It is important to carefully place the sponge, so it did not affect the flow of water in the pipe. The sensors electrodes were also covered with sponge and placed in the T-junctions. The elctrodes were covered top half with sponge so that the bottom half is exposed to the water in the pipe. This covering also resisted contaminated water or air bubbles surrounding the sensors.

Figure 12: Using sponge for avoiding the stagnation of water and air in the nodal junctions.

The technique of using sponge helped to improve the sensitivity of the sensors significantly. Fine air bubbles entering the network from the source tap seemed to be trapped between the sensor electrode and the nodal junctions. The only solution for this problem was to loosen the screw at the sensor's node, to have a minimal continuous leak of water. The leak was very low, such as 1 drop of water per 5 seconds. This allowed the air bubbles that were trapped to escape and the reading from the sensors were more uniform and accurate.

To perform all testes and intrusion scenario, constant flow rate of 22 l/min was set from the source tap. The measurement was done by filling buckets several times and timing with stopwatch. The average was found to be 22 l/min, and screw valve was kept open throughout the experiment while ball valve was opened and closed to full capacity.

The flexible PVC hoses used for the outflow needed to be adjusted and fixed on the wall, so the change of elevation does not affect the flow head of water. These flexible pipes are very sensitive to elevation change and slight increase of height can stop the flow of water in the pipe to outlet basin. Another problem with the flexible hose is kinking, where bending of the hose take place and reduce the cross-sectional area resulting in pressure increment. Since we had constant flow and pressure head, due to kinking the outflow pipes were not functioning perfectly. Therefore, U-bent PVC were attached at the end of the pipes to solve the issue. (Figure 12)

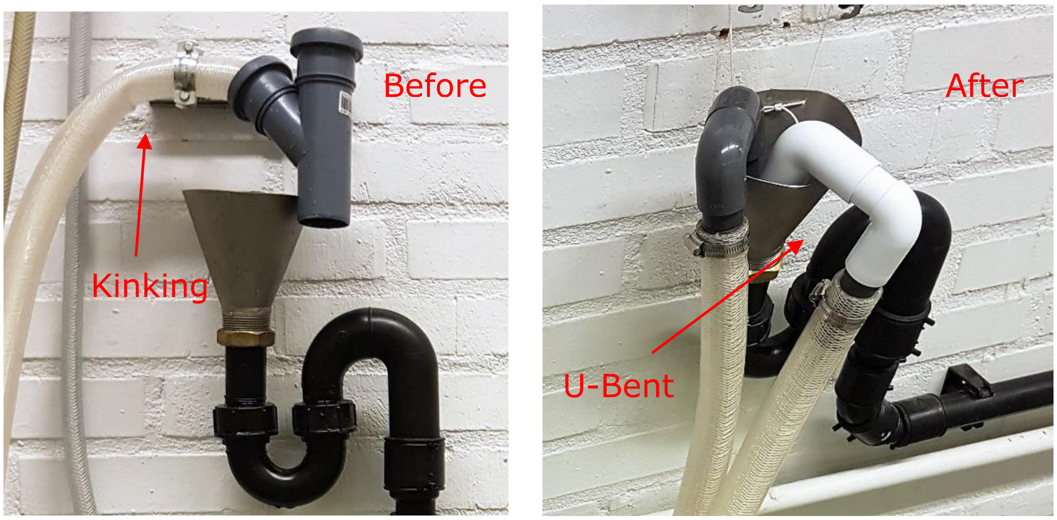


Figure 13: Kinking of flexible PCV hose (before) and attaching U-bent PVC to solve kinking (after).

The water flow in the scaled WDN system required to be running for few minutes before the experiments were started to remove the air bubbles from the source tap until a constant smooth water flow was attained. This was mandatory after changing the sensors' locations, and intrusion points from one node to another as the system needs to be depressurized before opening any nodal junctions.

3.5 Expereimental scenarios

Once all these calibrations were made and setup methods considered, the sytem is ready to perform experiments of sensors activity for intrusion scenario. For this study only two Sensorex transmitters were available, therefore, only two sensors could be used at a time. The unavailability of another transmitter for resulted in increase of operation time as all the threee sensors data could not be attained in single test. Hence there are 27 test scenario instead of only 9 (Figure 5). The experimentation with the 27 scenarios included different combinations of sensors placement, intrusion location and outflows setup. There were 3 three different intrusion points in the model and experiments done under these intrusion points were grouped to make sets. Therefore, there are sets each having 3 tests carried out with the same outflow setup but different sensor location.

3.6 Backtracking and source location identification

The backtracking methodology is one of the complicated aspects during the event of contamination in water distribution network. Important hydraulic criteria are taken into consideration when developing source location identification. Moreover, sensor's data plays an important role in the calculation of the backtracking and various of deterministic approach can be established depending on the type of sensors being used (Seth et al., 2016). Chemical sensors can detect concentration and changes in concentration from one sensor to another, and can briefly explain where the event of contamination might have happened. There are more sensors to generate alarm for changes in desired quality of water such as pH, conductivity, turbidity, dissolved oxygen etc. However, the sophisticated sensors like quality sensors can be costly and since this study is more focused on signals for water quality change and not concentration values; pH sensors were decided to be suitable.

4

RESULTS AND DISCUSSIONS

Analysis of the real time sensor data from the pilot scaled water distribution network will be discussed in this chapter. The behavior of the sensor responses were analyzed with concetration data from the specific sensor nodes in EPANET. All the sets of the 27 scenarios will be discussed and analyzed with graphical representations. The results will then be presented with the findings and limitations of the analysis as well as show how the chemical volume is transported within the network from the particular node of intrusion.

4.1 Statistics

All the 27 tests were carried out for six to eight minutes each and the pH measurements from the sensors were saved from the HOBOware software which also could generate linear graphs of the pH values vs time. However, the output data and graphs were difficult to relate especially when trying to compare with another data set. Therefore, instead of using the graphs of from the HOBOware, the pH values and time were extracted to create sets of graphs. These graphs are neat and merged with other data set to obtain single graph instead of showing three different graphs for one set of experiments. For example, Set 9 has three tests generating three different graphs (Figure 13) but these graphs were merged to one single graph of Set 9 as seen in the Figure 19.

Each test started once the pH of the flowing water reached neutral limit (7.8-8) and the intrusion chemical, in this case Acetic Acid, was injected after one minute from starting data logging. The data logger was stopped when the pH was gradually increasing towards neutrality after the intrusion and no disturbances were observed in the reading. To verify the accuracy of the sensor activity, each test was carried out twice and sometimes thrice, and then analyzed to select only one data set. Each sensor's response was cross referenced with other tests in different sets to justify accuracy for every test. In every set of combinations, response from each sensor repeated twice hence there is possibilities to achieve extract two data sets. For creating the single graph, best of the available data sets were considered.

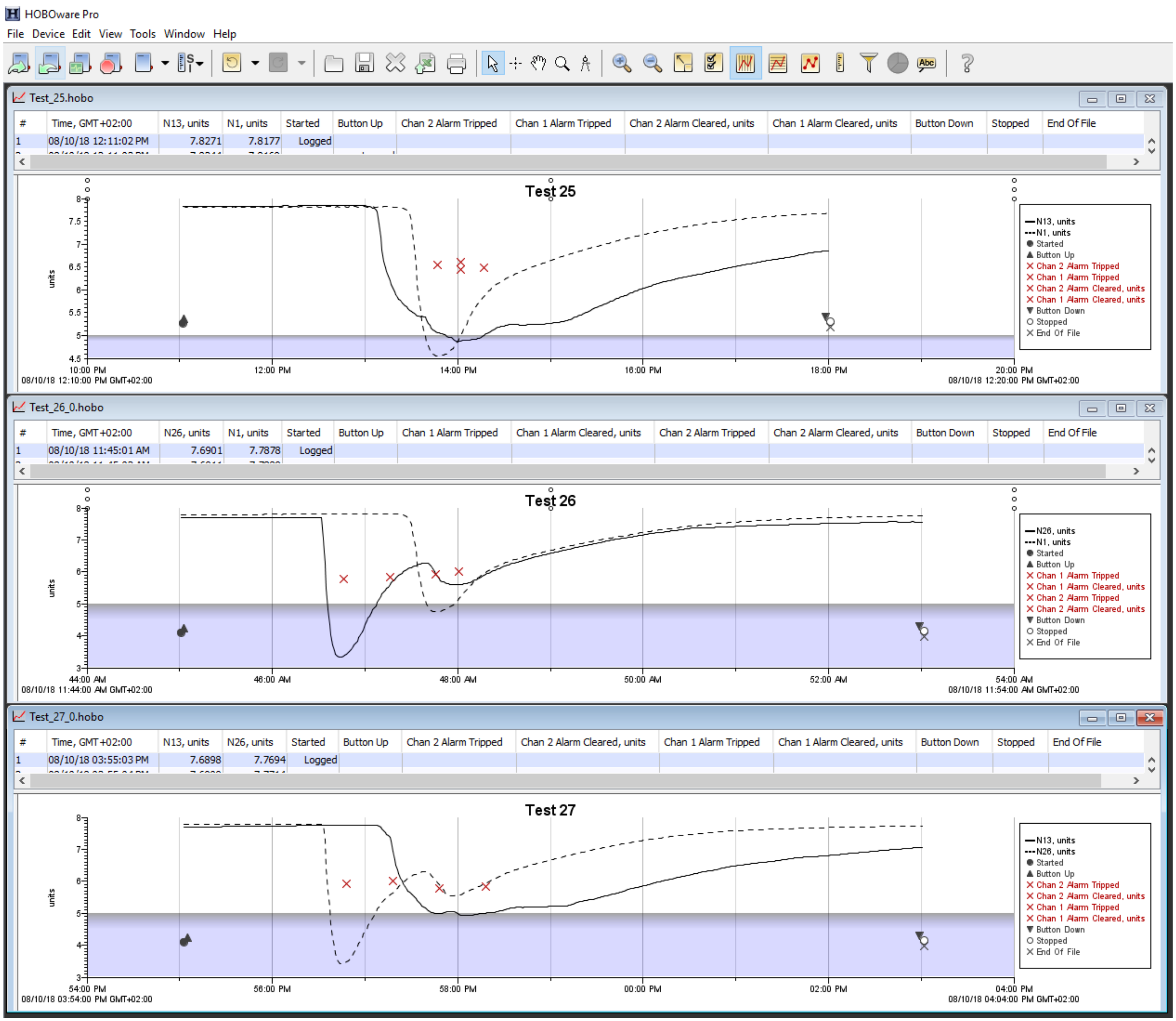


Figure 14:Sensors response, Data set and graph output from HOBOware for Tests 25, 26 & 27 of Set 9.

Table 5: Combinations of sensors placement, intrusion node location and outflow nodes for experimentation.

Intrusion N4				Intrusion N23				Intrusion N31			
Set 1	Test 1	Test 2	Test 3	**Set 4**	Test 10	Test 11	Test 12	**Set 7**	Test 19	Test 20	Test 21
Sensor N1	Active	Active	Inactive	Sensor N1	Active	Active	Inactive	Sensor N1	Active	Active	Inactive
Sensor N13	Active	Inactive	Active	Sensor N13	Active	Inactive	Active	Sensor N13	Active	Inactive	Active
Sensor N26	Inactive	Active	Active	Sensor N26	Inactive	Active	Active	Sensor N26	Inactive	Active	Active
Outflow 27	Open			Outflow 27	Open			Outflow 27	Open		
Outflow 28	Open			Outflow 28	Open			Outflow 28	Open		
Outflow 29	Close			Outflow 29	Close			Outflow 29	Close		
Constant Flow	22 LPM			Constant Flow	22 LPM			Constant Flow	22 LPM		
Set 2	Test 4	Test 5	Test 6	**Set 5**	Test 13	Test 14	Test 15	**Set 8**	Test 22	Test 23	Test 24
Sensor N1	Active	Active	Inactive	Sensor N1	Active	Active	Inactive	Sensor N1	Active	Active	Inactive
Sensor N13	Active	Inactive	Active	Sensor N13	Active	Inactive	Active	Sensor N13	Active	Inactive	Active
Sensor N26	Inactive	Active	Active	Sensor N26	Inactive	Active	Active	Sensor N26	Inactive	Active	Active
Outflow 27	Open			Outflow 27	Open			Outflow 27	Open		
Outflow 28	Close			Outflow 28	Close			Outflow 28	Close		
Outflow 29	Open			Outflow 29	Open			Outflow 29	Open		
Constant Flow	22 LPM			Constant Flow	22 LPM			Constant Flow	22 LPM		
Set 3	Test 7	Test 8	Test 9	**Set 6**	Test 16	Test 17	Test 18	**Set 9**	Test 25	Test 26	Test 27
Sensor N1	Active	Active	Inactive	Sensor N1	Active	Active	Inactive	Sensor N1	Active	Active	Inactive
Sensor N13	Active	Inactive	Active	Sensor N13	Active	Inactive	Active	Sensor N13	Active	Inactive	Active
Sensor N26	Inactive	Active	Active	Sensor N26	Inactive	Active	Active	Sensor N26	Inactive	Active	Active
Outflow 27	Close			Outflow 27	Close			Outflow 27	Close		
Outflow 28	Open			Outflow 28	Open			Outflow 28	Open		
Outflow 29	Open			Outflow 29	Open			Outflow 29	Open		
Constant Flow	22 LPM			Constant Flow	22 LPM			Constant Flow	22 LPM		

4.2 Data analysis

In this section of the chapter, the analysis of the data and graphs that were generated from the sensor activities due to intrusion of acetic acid will be discussed. As mentioned above, 9 sets of graphs were created using data from HOBOware. The graphs will be discussed, analyzed and compared with results from EPANET simulations.

4.2.1 Intrusion at node N4

Starting with the intrusion event at node N4, there are 3 sets with 9 separate tests; Set 1, Set 2 and Set 3 and all having same inflow from source tap and the graphs generated respectively (Figure 14). In Set 1, outflow node 27 and node 28 are open at full capacity and together they discharge water at 22 l/min. When intrusion is injected from node N4, it is carried by the flow to directly towards N27 without spreading into the downstream network. As a result, only sensor placed in node N1 can detect the intrusion while other sensors at N13 and N26 shows no response. Similar output is visible in the scenario of Set 2 where only N27 and N29 are open to full capacity. It seems that the intrusion chemical is flowing as laminar flow and not mixing perfectly to spread throughout the system network. From graph in Figure 14, we can see that only N1 is displaying sharp fall in pH for both Set 1 and Set 2. However, there is slight delay in response to reach neutral pH in case of Set 2. The same response behavior is visible in case of Set 3 when only N28 and N29 are open. The data from HOBOware shows that in Set 1 and Set 2, sensor at N1 detects intrusion in 27 seconds and 28 seconds respectively while in Set 3, it is slightly longer with 35 seconds (Note: intrusion was injected after 1 minute of start). Moreover, when N27 is closed, the intrusion is forced to flow through sensor placed at N13 and since some distance is covered to reach the sensor; the response comes at 1 minute 35 seconds.

The WDN model was simulated in EPANET using the hydraulic parameter similar to the pilot scale model but considering time settings for 1 minute (Figure 15). It is evident that N1 always receives high concentration of chemicals (3.5mg/L from 5mg/L injected) while other sensors receive negligible amount. However, the Set 3 in this case displays similar sensor activities at N1 and N13 even though outflow at N27 is restricted. Surprisingly, sensor at N26 does not show any changes in pH in real-time tests as it shows for EPANET simulations where it receives approximately 1mg/L chemicals.

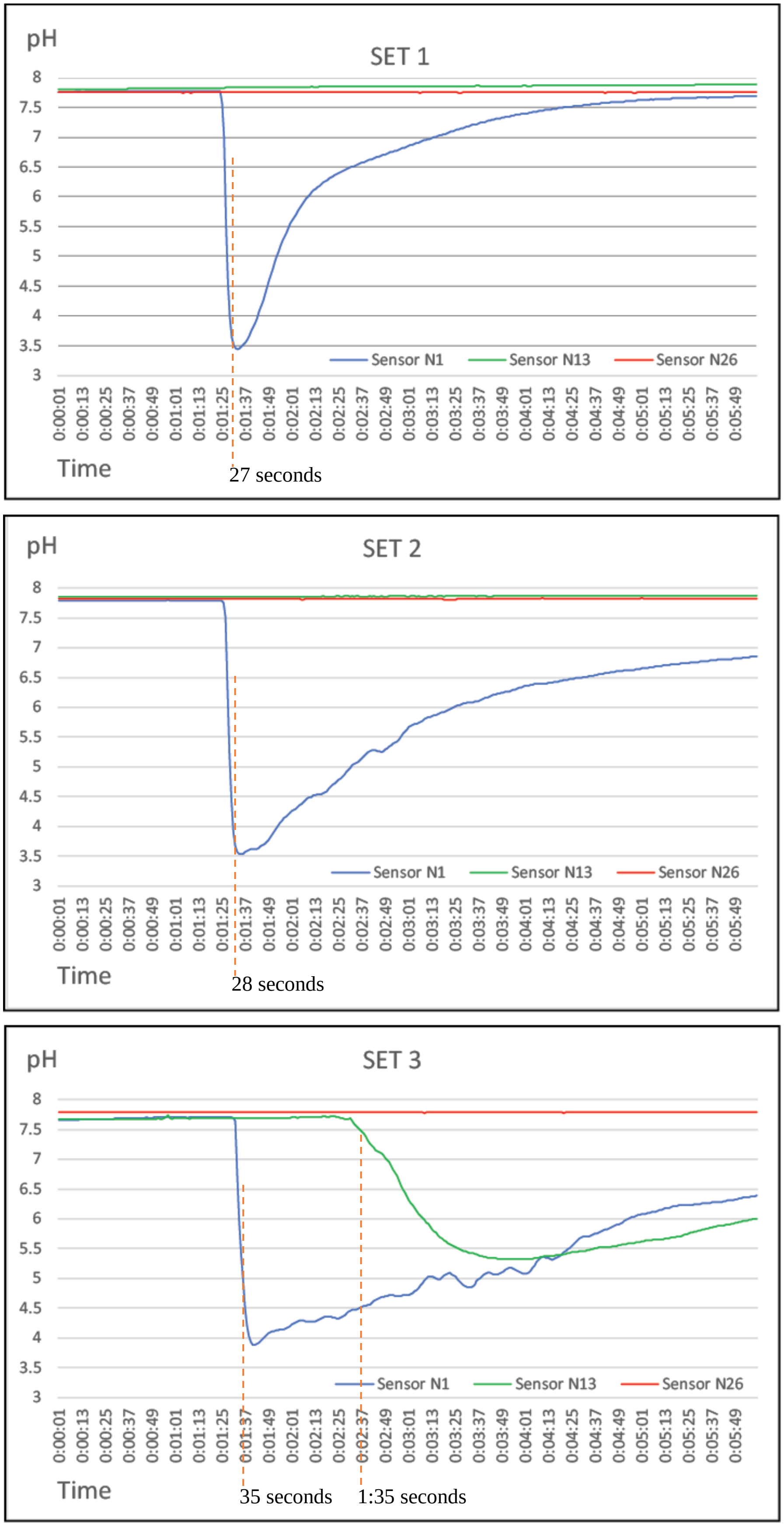
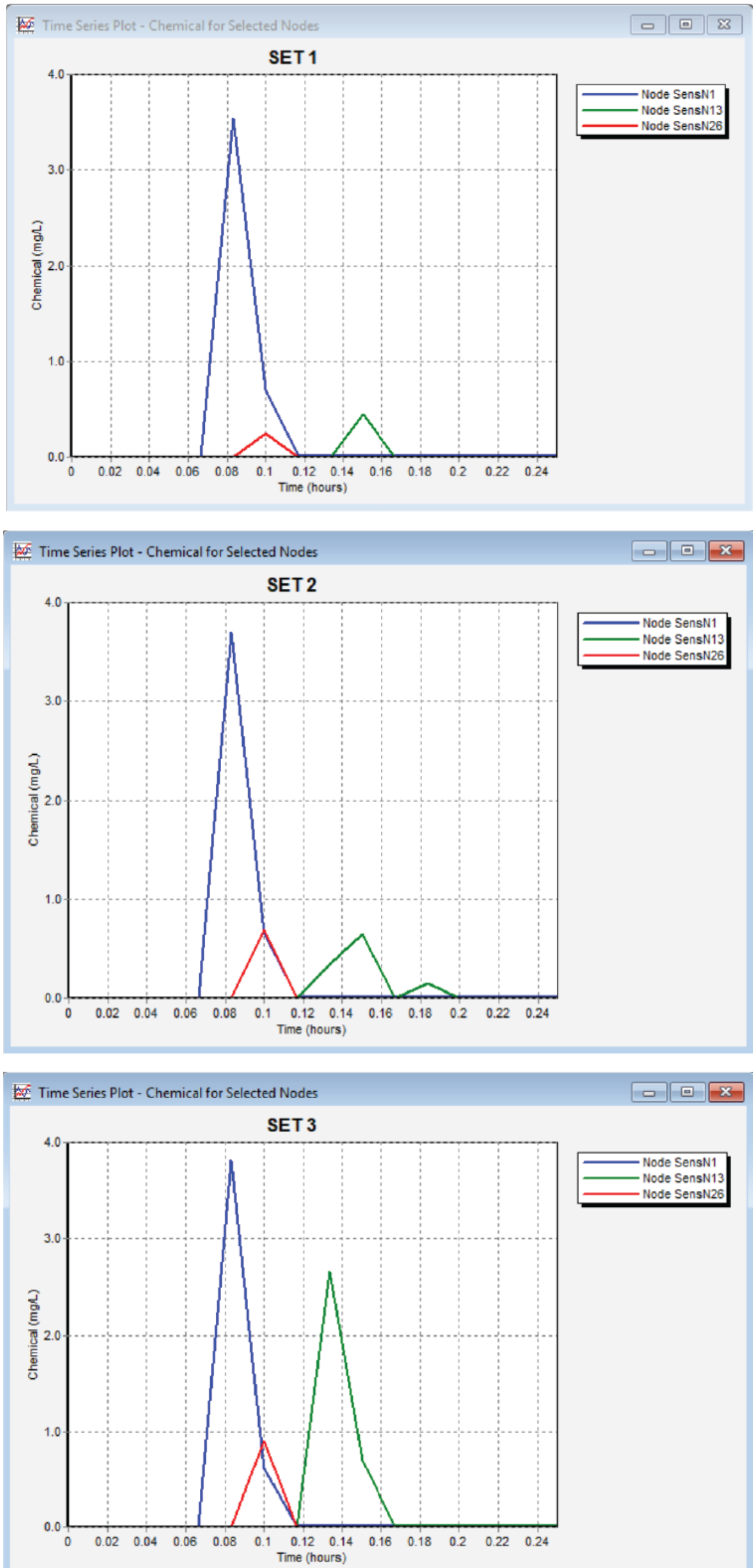


Figure 15: Sensors' response to intrusion at N4 for tests in Set 1, 2 & 3.




Times Options

Property	Hrs:Min
Total Duration	0:15
Hydraulic Time Step	0:01
Quality Time Step	0:01
Pattern Time Step	0:01
Pattern Start Time	0:00
Reporting Time Step	0:01
Report Start Time	0:00
Clock Start Time	12 am
Statistic	None

Figure 16: Sensors' response to intrusion at N4 using EPANET for tests in Set 1, Set 2 & Set 3. Time options used for the simulations is also shown.

4.2.2 Intrusion at node N23

Similar to intrusion at node N4, exact test setup was implemented except intrusion node is N23 instead of N4. The same 9 tests from 3 different sets were performed and the graphs are generated. From the graphs in the Figure 16, it is obvious that when N29 is not open, the intrusion chemical (acetic acid) flows through the sensor located at N13 and N26 (Set 4) while sensor at N1 detects no changes in pH value. It seems that when N29 is open as in both cases in Set 5 and Set 6, the intrusion chemical is discharged through N29 and does not spread into the system network hence none of the sensors can detect any changes in water quality. The response time for the sensors at N13 and N26 for the intrusion at N23 is 35 seconds and 2:17 seconds respectively. This response pattern proves that the intrusion chemical is flowing in laminar flow and not perfectly mixed. The way the WDN was designed (Figure 1), it seems obvious that some concentration of the intrusion should reach sensor at N26 in all cases since the distance to travel from N23 to nearest junction N22 which transfers water to N29 and N26 is short. The response from sensor at node N13 is particularly strange as it is taking more than to 2 minutes to response where intrusion from N4 (Figure 14) and intrusion from N31 shows more unique response pattern.

Testing the WDN model in EPANET for the configuration of Set 4, 5 & 6, the sensors' activity at N1, N13 and N26 are displayed in Figure 17. Graph of chemical concentration reaching the sensors in Set 4 is similar to what achieved from the pilot scale model. N26 is receiving chemical concentration of approximately 3.5 mg/L while N13 receives around 2 mg/L. However, in case of Set 5 EPANET simulations show some sensor activities at N26 and N13 with chemical concentration of 1 mg/L and 0.25 mg/L. Considering 0.25 mg/L of chemical concentration at N13 is negligible, 1 mg/L concentration reaching at N26 and this is not replicating during the real-time tests is a matter of concern. In addition, when N28 and N29 are open, the intrusion is carried out directly through N29 and therefore no sensors can detect any changes in pH.

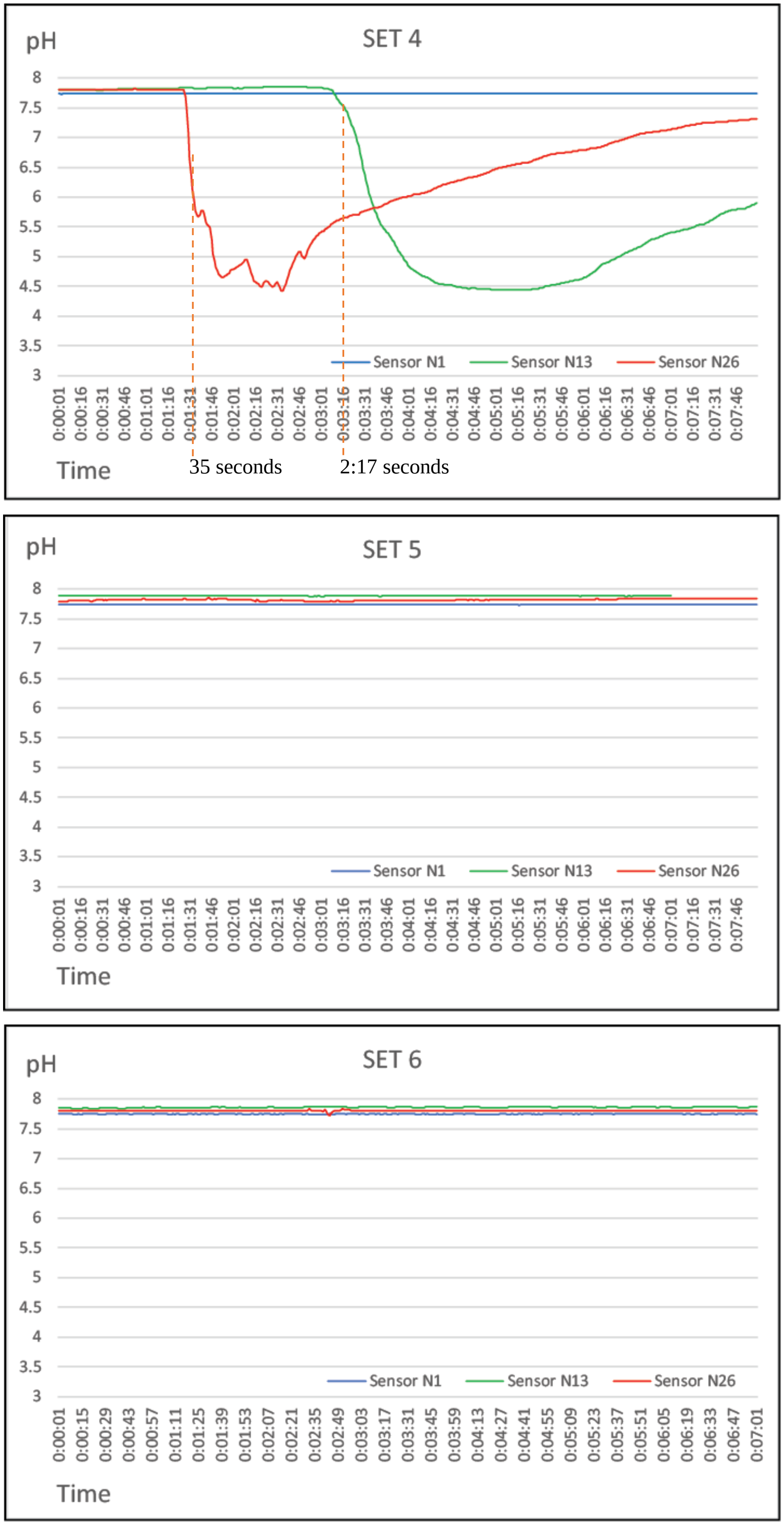


Figure 17: Sensors' response to intrusion at N23 for tests in Set 4, 5 & 6.

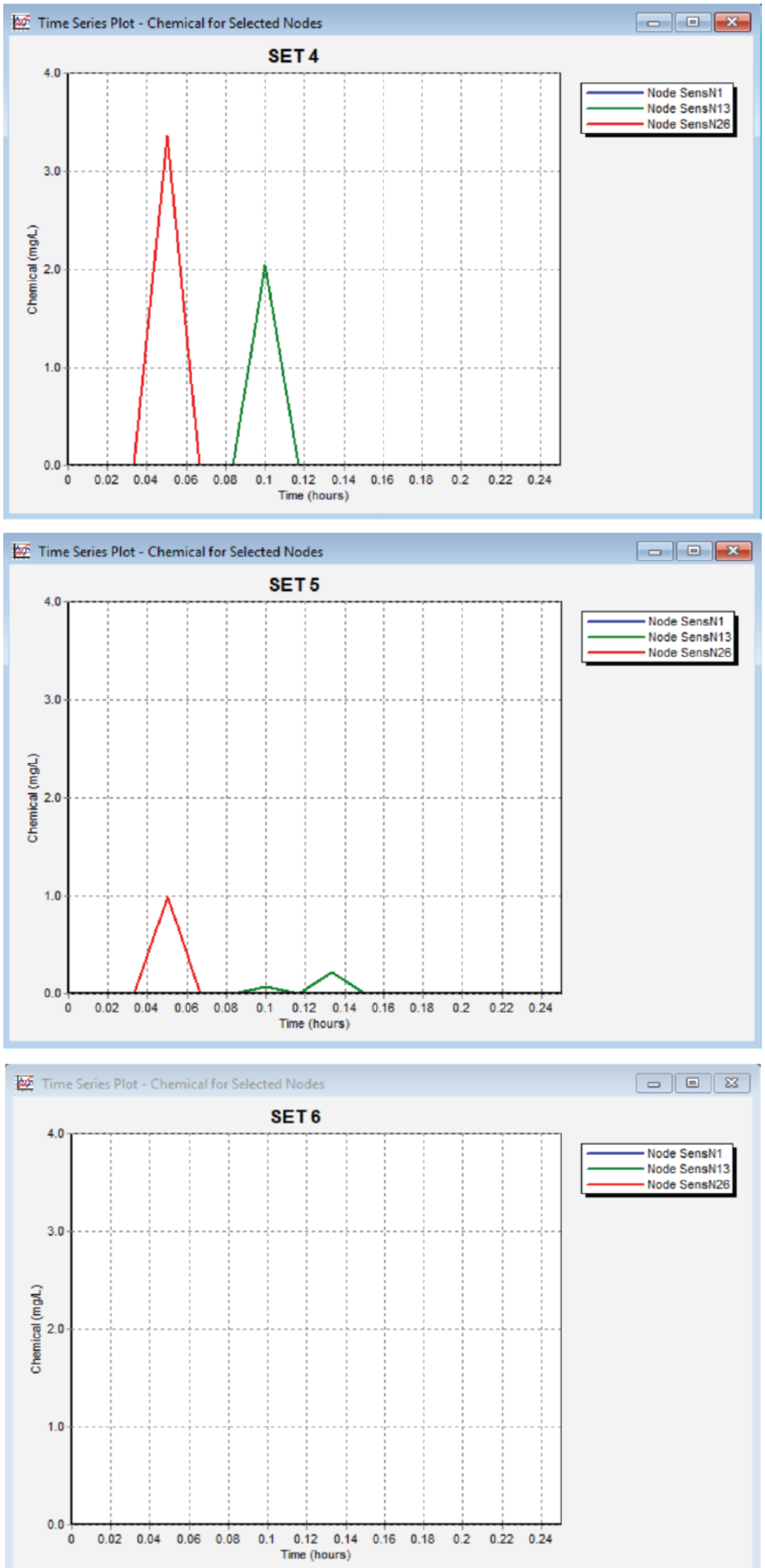


Figure 18: Sensors' response to intrusion at N23 using EPANET for tests in Set 4, Set 5 & Set 6.

4.2.3 Intrusion at node N31

The sensors activities for node N31 are very interesting as it is the most upstream node and placed in the center pipe line of the network. The spread of intrusion chemical is therefore extensive and most of the cases the sensors can detect the change in pH of the water. Analyzing Set 7 (Figure 18) where only node N27 and N28 are open for outflows, intrusion chemical follows the flow and pass through sensors at N1 and N13 and no response is observed from sensor placed at N26. Analyzing the graph and reading the time to respond from the HOBOware, it is evident that sensor at N13 is quicker to respond to pH change than sensor at N1 even though N13 is the furthest from N31. Response time for N13 after intrusion is approximately 1 minute while for N1 is 1 minute 18 seconds. Similar response is also obtained from the tests in Set 8 and Set 9 where sensors at N13 receives contaminated water faster than N1. This pattern of response again justifies the nature of laminar flow of intrusion, where the chemical is carried along the stream and no perfect mixing happening.

However, for Set 8 when N28 is closed and N29 is opened, all the three sensors responded to the intrusion and same goes for Set 9 where only node N28 and N29 are open. In Set 8, the response time for sensors at N1 and N13 is the same which is approximately 1 minute and 22 seconds; while the closer sensor location N26 detects intrusion in 37 seconds after the injection from N31. The results from Set 9 ensures the sensors' activity and how the intrusion is flowing in the network that is in laminar flow along the stream. Since node N28 and N29 are close sensor location at N13 and N26 respectively, the responses are quicker than Set 8. Here the intrusion from N31 reaches sensor location at N26 in 33 seconds, at N13 in 1 minute 9 seconds and at N1 in 1 minute 30 seconds.

The sensors' response from EPANET simulations shown in Figure 19, displays that all the three sensors locations are affected by the intrusion at N31. Surprisingly, in the EPANET simulations sensor at node N1 receives contaminated water before node N13 which is not the case in real-time tests. However, one common phenomenon is observed especially for sensor at N26. When concentration in EPANET is lower than 1 mg/L, no changes in pH are visible in real time data. In Set 7 there is no response from N26 for pilot scale model (Figure 18) but in case of EPANET simulation N26 does show a concentration of 1mg/L (Figure 19). Nevertheless, responses from N26 in Set 8 and Set 9 are more reasonable and share similar pattern of sensor activities with real-time tests done using the scaled WDN model.

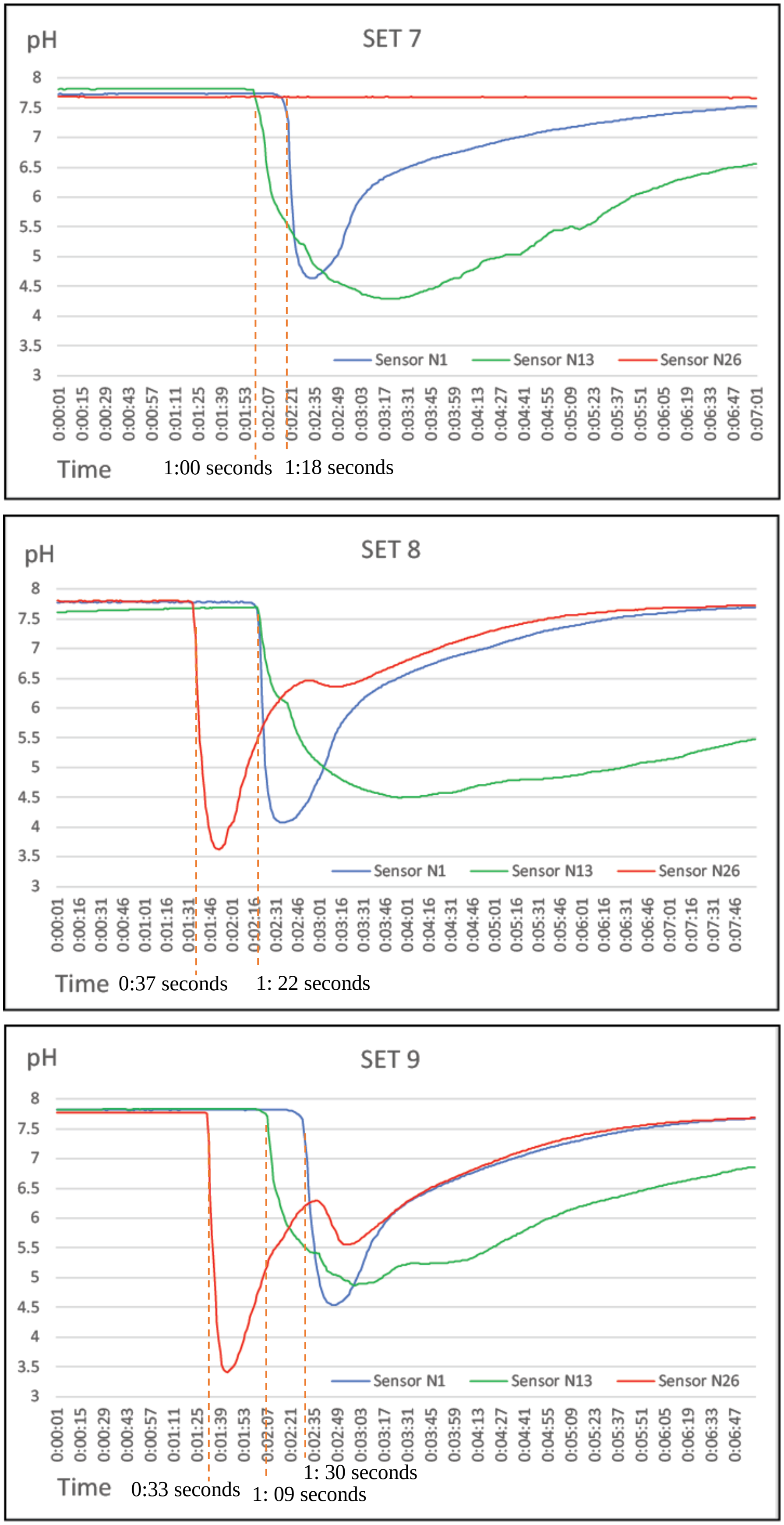


Figure 19: Sensors' response to intrusion at N31 for tests in Set 7, 8 & 9.

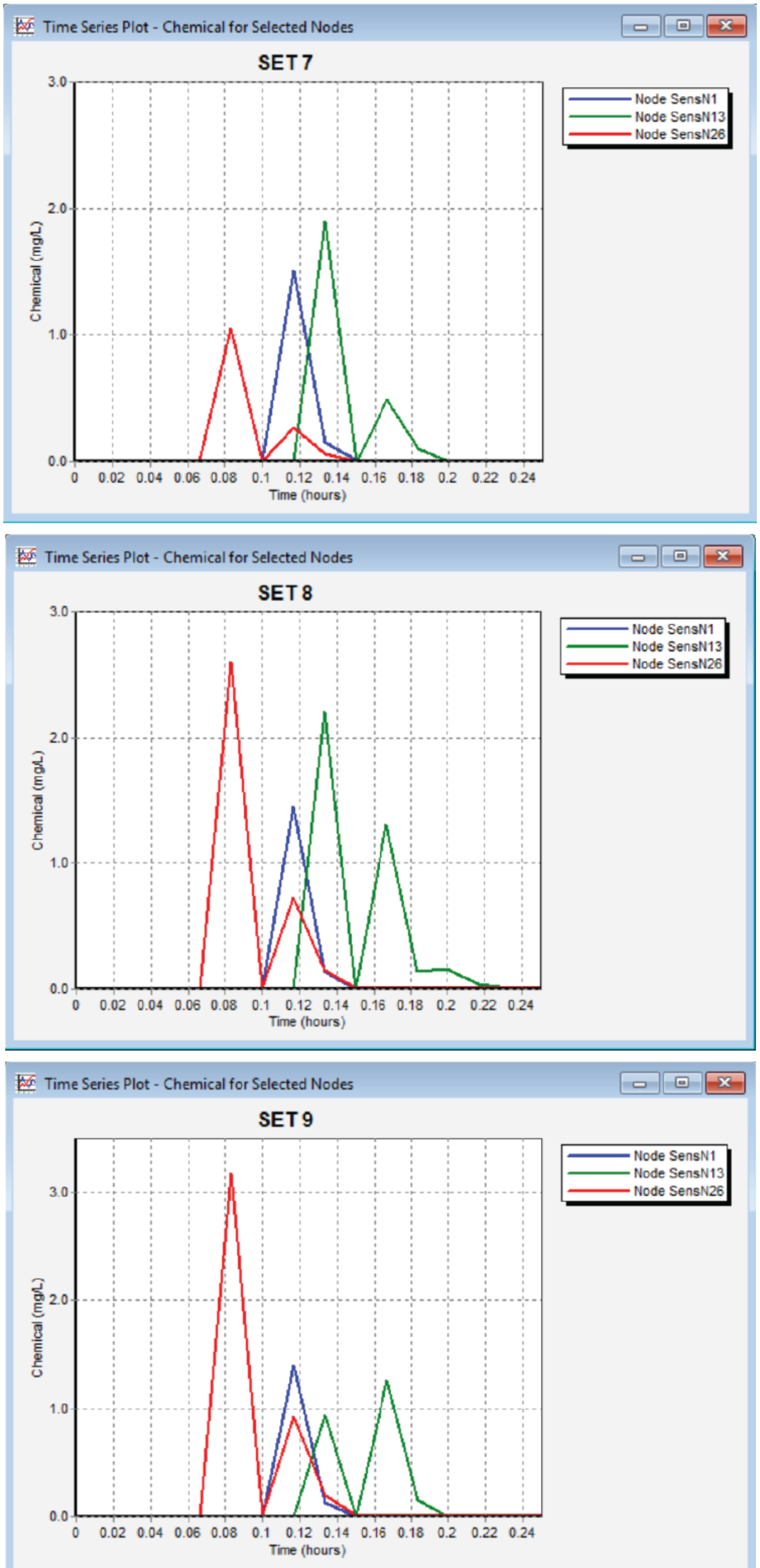


Figure 20: Sensors' response to intrusion at N31 using EPANET for tests in Set 7, Set 8 & Set 9.

4.3 Findings and limitations

In any research work key findings and limitations of the study are important part of the discussions. Working with EPANET and less detailed pilot scaled WDN network gave rise to several issues which were discussed in the previous chapters. However, this section of the chapter hopes to review the major findings from the experimental data and simulations in EPANET as well as the limitations of the study.

4.3.1 Pilot scaled laboratory experiment

It is possible to experiment with sensors placement and intrusion event detection in pilot scale WDN model. The experimental study shows some interesting facts of the hydraulic paraments, flow pattern in a confined pipe network. The study also demonstrates simple setup with less resource utilization and low-tech devices to perform critical experiments such as contamination detection and flow pattern. If the sensor response data are closely observed, the flow pattern for the chemicals in the pipe network from different intrusion nodes can be visualized as presented in the Figure 20, 21 and 22.

In the event of intrusion from node N4, three different set of sensor activities are obtained. Combining the data set and pattern of how the sensors responded, the following Figure 20 can be presented.

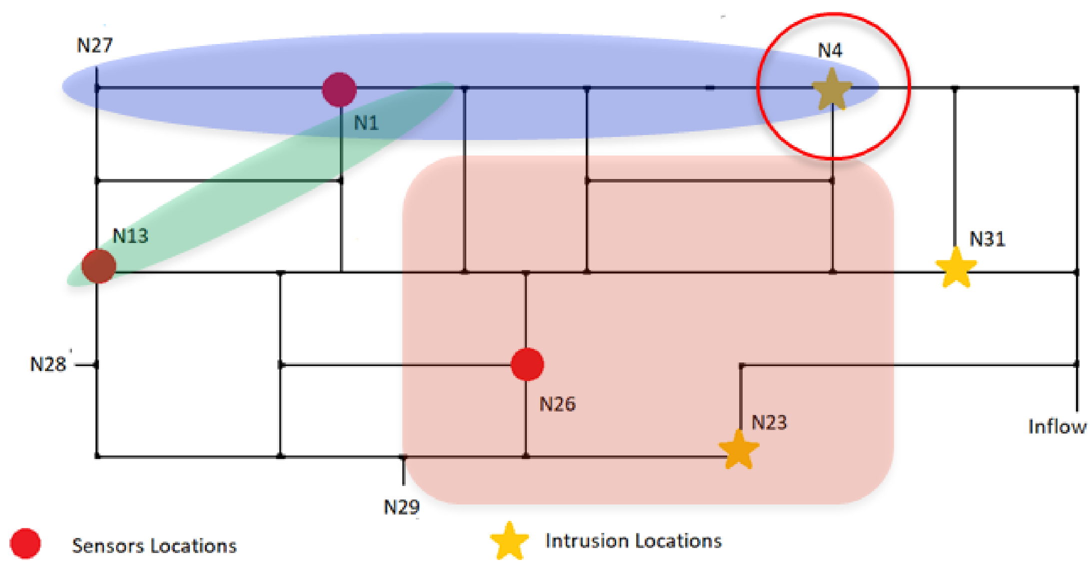


Figure 21: Chemical flow pattern when intruded from node N4. Colors differentiate sensor activities and shapes represents flow of chemical (i.e. elliptical shape: chemical flow; rectangular: no chemical flow).

It is easy to understand that the chemical when intruded from node N4, it tends to flow in laminar without significant dispersion. As a result, most of lower parts of the WDN is unaffected since the chemical did not spread in the system throughout the experiments in Set 1, 2 and 3. Similar results are also obtained for the intrusion scenario at node N23 and node N31. Figure 21 which shows the chemical flow pattern in case of intrusion at N23 (Set 4, 5 and 6), also indicates that intruded chemical flows along the water in laminar without spreading in whole network.

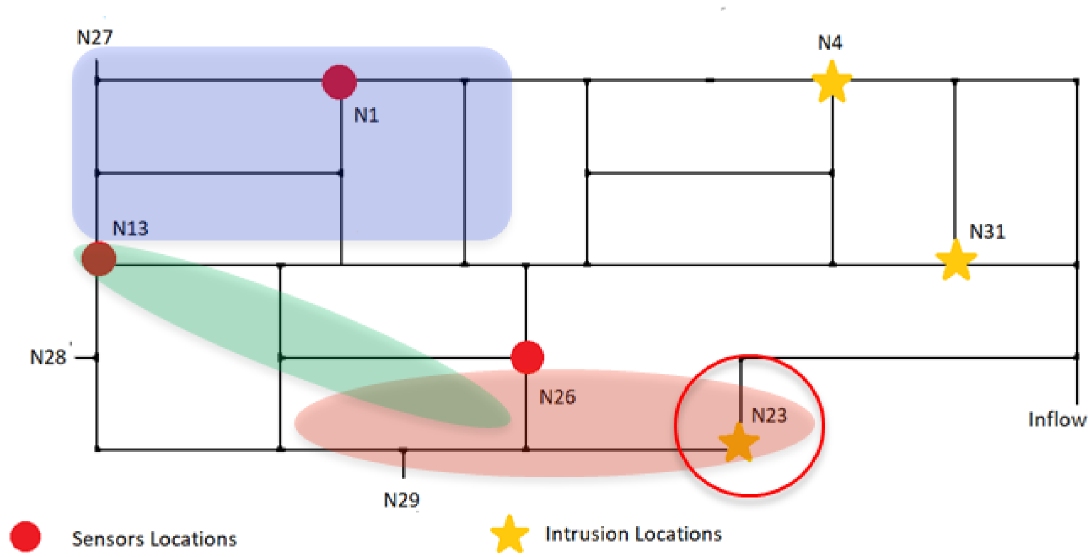


Figure 22: Chemical flow pattern when intruded from node N23. Colors differentiate sensor activities and shapes represents flow of chemical (i.e. elliptical shape: chemical flow; rectangular: no chemical flow).

In addition, the findings from the above intrusion scenario verified by the intrusion scenario for node N31 (Set 7, 8, and 9). Node N31 being located at the center and furthest downstream of the network and intruded chemical passes through most of the other nodes; spreading of the chemical occurs more significantly throughout the network as it can be seen in Figure 22. However, the laminar behavior of the chemical flow in the water is noticeable. Separate experiment was carried out to investigate chemical flow from node N31 to N13 by placing sensors at N19, N12 and N13, which clearly shows that the chemical concentration is higher in the pipe connecting N31 and N13. Data also indicates that pH value at node N19 and N26 are similar while pH values of N1 and N12 are similar which can suggest dispersion in the junctions. More details about the investigative experiment carried out will be mentioned later in this section when analyzing the behavior of sensor placed at node N13.

Looking carefully at the sensors' response curves for all the sets of experiments it is noticeable that sensor activity at node N13 is different from the other two sensor activities. Sensor at N13 is placed far in the downstream of the network and responses slowly to intrusion events. It is also evident that the sensor takes longer time to reach neutral pH value. Of course, some of the phenomenon are not easy to explain or find their exact reason, so some assumptions were tested and investigated.

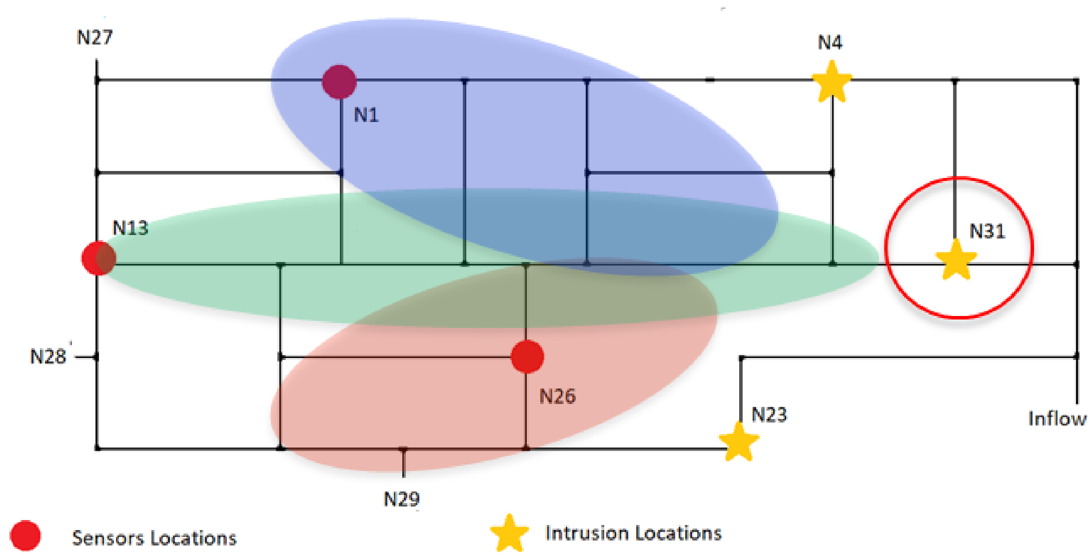


Figure 23: Chemical flow pattern when intruded from node N31. Colors differentiate sensor activities and shapes represents flow of chemical (i.e. elliptical shape: chemical flow; rectangular: no chemical flow).

Sensor electrode L7881 HD was placed at node N13 and initially assumed if the sensor's conductivity is weaker than L8281 HD sensors which were placed at N1 and N26. However, a separate experiment was carried out where the sensors were swapped. Node N13 was placed with L8281 HD and node N1 was placed with L7881 HD. The response was the same for intrusion injection at node N31 (Figure 23) and cancels out the assumptions of weaker signals that previously was made for L7881 sensor electrode.

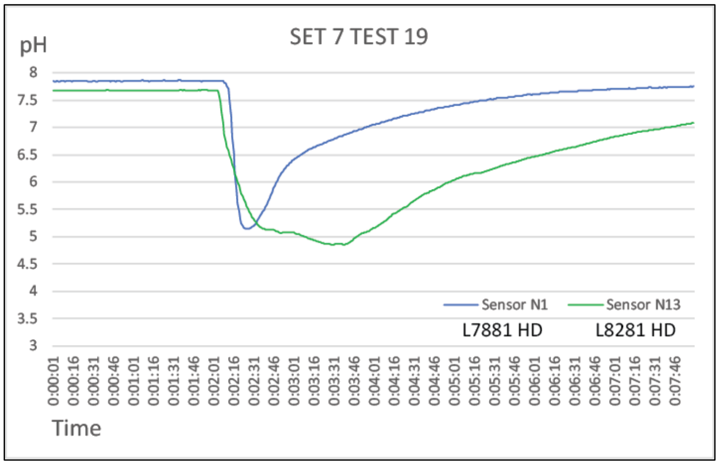


Figure 24: Sensor response after swapping sensor electrode at node N1 and N13

Apparently, the intrusion concentration reaching at node N13 is something interesting to learn more about. Since the piping of the network is non-transparent material, it is hard to investigate the reason behind the phenomenon of slow responsiveness of that node junction.

Another hypothesis was made regarding this response pattern which concerns with plume concentration in flow within closed pipe system. Theory of plume concentration or dispersion in fluids may explain the behavior of the intrusion chemical in the pipe network and the sensor activity at node N13. Dispersion in laminar flow is defined as the spreading or diffusion of mass of particles along the streamline with longitudinal mixing over the cross-section of the pipe. This means that the intrusion concentration will dilute and spread over larger cross-sectional area of the pipe (Figure 24). It is therefore assumed that intrusion is dispersed and reaches sensor location node N13 with low concentration initially (slow drop of pH) then attains peak concentration (highest pH value) before it starts to fall again (slowly regaining neutral pH). This is what the all curves from sensor at N13 suggests.

However, the theory of dispersion long the length of the pipe is not justified when looking at the curves for sensor at node N1 for intrusion event at node N4. The response curve for N1 shows sharp drop and quick regain in neutral pH even though it has similar laminar flow characteristics. Moreover, another separate experiment was carried out to investigate the intrusion flow pattern in pipe length from node N31 to node N13. In this case, sensors were placed at node N12 and node N19 which are 2 m and 3.5 m apart from N13 respectively.

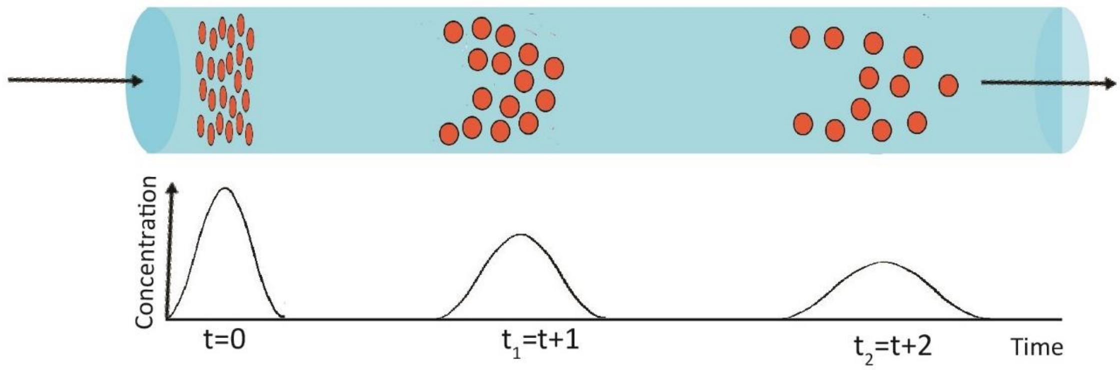


Figure 25: Concept of longitudinal dispersion under laminar flow with respect to time.

The responses from these locations are also surprising as they do not show the dispersion effect as it is seen for sensor at N13 (Figure 25). The investigation was done under the setup of Set 7 where N27 and N28 were open as outflows.

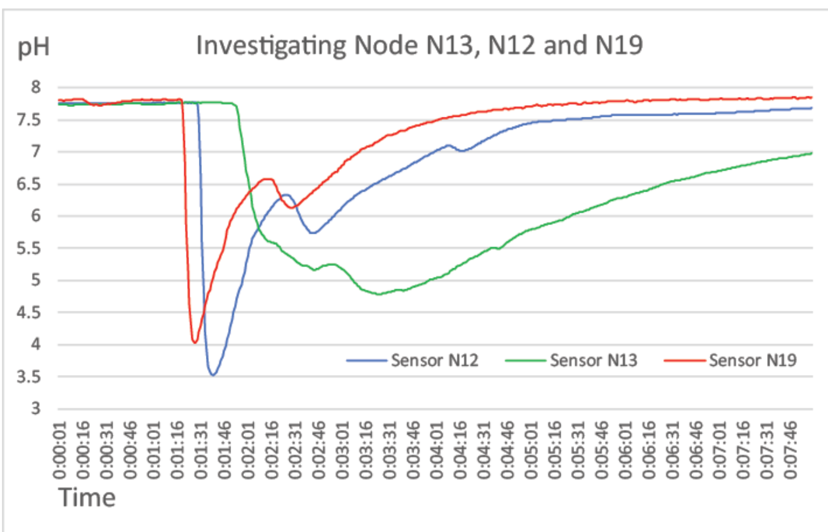


Figure 26: Investigating dispersion effect from intrusion at N31 in the pipe extending to N13.

Dispersion is dependent on the high/low velocities of flow and dispersion coefficient. Advective transport eliminates dispersion effect during high flow velocity and low dispersion coefficient (Axworthy & Karney, 1996). The theories are contradictory and hence no suitable explanation is available for the response pattern for node N13. However, it can be assumed that the N13 is the furthest node junction of the network and water cannot flow through the junction to another pipe along the streamline rather experience turbulent mixing at the junction and therefore demonstrates such slow response and long duration to regain neutral pH.

4.3.2 Usability of EPANET software

Introduction to the EPANET model software was presented in Chapter 3 where the interface and the basic settings were discussed for the WDN model in this study. Analyzing the data and outputs from both EPANET and pilot scaled model, it is logical to comprehend that sensor activities are possible to project from the all the other nodes available in the WDN model using EPANET simulations. The software also requires less computational resources and runs smoothly in a standard computer. The computer used for this study had the configuration of Intel Core i5 6th Gen CPU with 2.30 GHz processor speed and 8 GB RAM.

Success of EPANET in this study was to simulate contamination event for the nodes and the possibility to generate concentration quantity and trace data for different intrusion scenarios. Since it is not feasible to test all the nodes and place such quantity of sensors in pilot scaled model, EPANET simulations have benefited in this matter. Therefore, in developing any methodologies for optimization of sensors placement in water distribution network, data from EPANET can be used without much hesitation to feed the algorithms to run numerical calculations. Such data can also be used in developing and testing artificial intelligent (AI) programs in the future.

One of the objectives of this study was contamination source tracking, which could only be possible using EPANET as it requires less resources and less time consuming. However, with advancement of the study, EPANET fails to meet the requirements and incapable of analyze one key aspects source tracking; time-based flow and contamination distribution. As a result of which EPANET is not suitable for use of intrusion event and detailed mass transport spreading for space-time analysis.

The time settings (Figure 3 and Figure 15) in the EPANET simulations is user defined and results generated for every scenario is affected by the different time settings. The pattern of the graphs or the concentration receiving at a particular node is remains the same but the time to response change. The time settings in the EPANET browser toolbar, is dependent on user inputs, i.e. user can set any times (hydraulic time step, quality time step etc.) and the report is produced accordingly. It seems the time settings reflects the number of steps from one node to other. For example, in the WDN network (Figure 2) contamination entering in Node 31, requires 5 steps to reach Node N26. Now if the time settings is set to 1 minute (Figure 26a), the graph or table will show 5 minutes (Figure 27). If the time settings is set to 15 seconds (Figure 26b), the same graph and table will be produced but the time will be 1 minutes 15 seconds (5 x15 seconds). Desired simulations process is where the software will calculate itself using the hydraulic parameters (mixing type, flow rates etc.) and design values (pipe length, diameter etc.) to show where a certain volume of water will be in the network within a certain time frame.

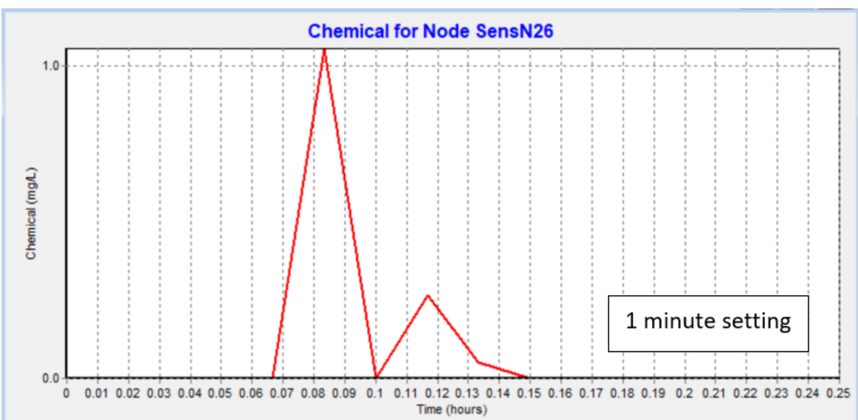
Times Options

Property	Hrs:Min
Total Duration	0:15
Hydraulic Time Step	0:01
Quality Time Step	0:01
Pattern Time Step	1:00
Pattern Start Time	0:00
Reporting Time Step	0:01
Report Start Time	0:00
Clock Start Time	12 am
Statistic	None

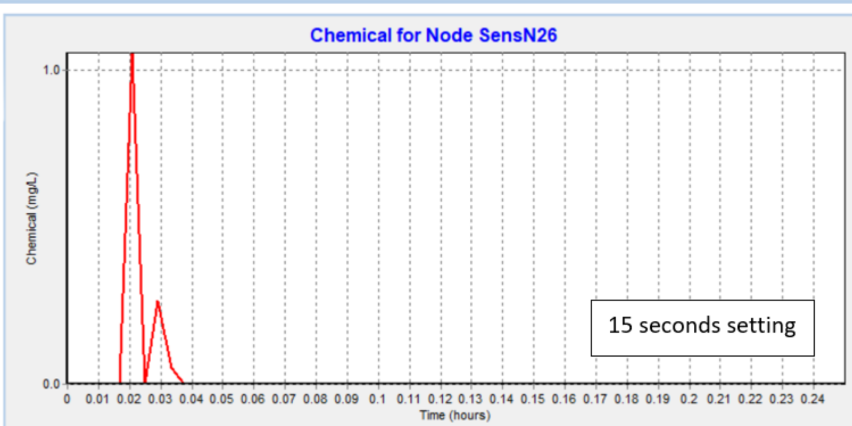
Times Options

Property	Hrs:Min
Total Duration	0:15
Hydraulic Time Step	0.0042
Quality Time Step	0.0042
Pattern Time Step	1:00
Pattern Start Time	0:00
Reporting Time Step	0.0042
Report Start Time	0:00
Clock Start Time	12 am
Statistic	None

Figure 27: Time setting in EPANET for a) 1 minute and b) 15 seconds.



Time Hours	Chemical mg/L
0:00	0.00
0:01	0.00
0:02	0.00
0:03	0.00
0:04	0.00
0:05	1.05
0:06	0.00
0:07	0.27
0:08	0.05
0:09	0.00
0:10	0.00
0:11	0.00
0:12	0.00



Time Hours	Chemical mg/L
0:00	0.00
0:00	0.00
0:00	0.00
0:00	0.00
0:01	0.00
0:01	1.05
0:01	0.00
0:01	0.27
0:02	0.05
0:02	0.00
0:02	0.00
0:02	0.00
0:03	0.00

Figure 28: Chemical concentration at Node N26 for time settings of 1minute and 15 seconds.

4.3.3 Possibilities of intrusion back-tracking

Intrusion back-tracking could be an interesting aspect of the study but then again, the simplicity of the data from pH sensors and the limitation of time setting in EPANET; further analysis was not possible. Efficient source back-tracking requires information which are time-variant allowing to project contamination flow concentration both in forward direction and backward direction. Such analysis also requires high computational work and data of mixing properties of chemicals used for intrusion scenario. However, developing a probabilistic method for each node can provide some important outcomes to narrow down the search area of intrusion.

The probabilistic approach is inspired by the probabilistic logic when there are higher number of uncertainties and difficulties to conclude to a unique solution. The probabilistic approach can be suitable for this study because the data analysis show that the chemicals tend to flow in laminar. Just to provide simple example how the method can be used to identify possible nodes contaminated using EPANET. For intrusion at node N4, and outflow N28 and N29 being open (Set 3), it is possible to generate a contour map of how intrusion is flowing in the network and show which nodes are contaminated (Figure 29). Considering first sensor to respond to intrusion as reference for all the other sensors, the time for other sensors to respond can be monitored to localize possible intrusion nodes. In order to resemble the exact setup of the pilot scaled WDN, the following (Figure 28) EPANET model was adopted where extra pipes are used from the outflow nodes to join at one outlet. This is done to eliminate the inconsistent individual flow rate at each outflow nodes and join them to one outlet node having same outflow as inflow (22 l/min). This setup also helped EPANET to adjust flow in the pipes automatically rather than trying to meet the base demand at each outflow nodes.

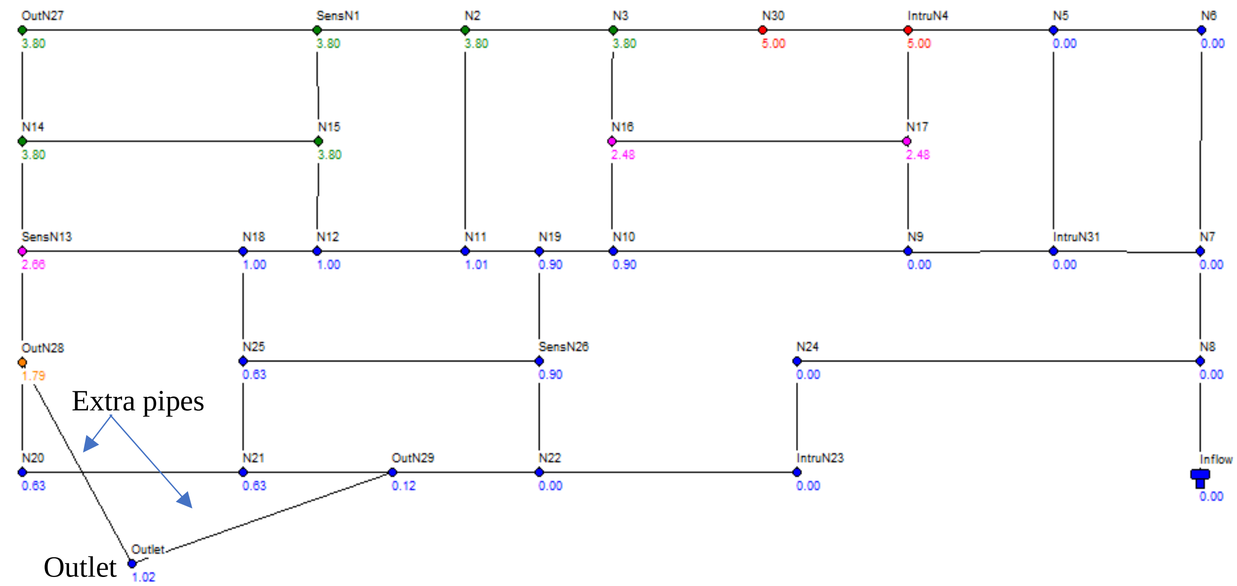


Figure 29: EPANET simulation for intrusion at node N4 and contaminated nodes with concentration values when outflow at N28 and N29 are active.

From the above figure, it is possible to see which nodes are being contaminated due to intrusion at node N4 and their respective concentration values. Taking sensor at node N1 as reference and using the sensor responses from set 3, 6 and 9 (Figure 14, 16, and 18), intrusion at node N23 can be rejected since sensor at N1 shows no activities for this certain intrusion node. Now checking for intrusion where both sensor N1 and sensors N13 are responding, Set 3 and Set 9 matches the criteria. Investigating further, check for sensors response from N1 and N26. If there is no response from sensor at N26, it is reliable to conclude that the intrusion occurred at node N4 (basing on this experiment data) or somewhere close to node N4. If N26 shows response, then the intrusion occurred at node N31. Figure 30 simplifies the sensor response data for source back-tracking.

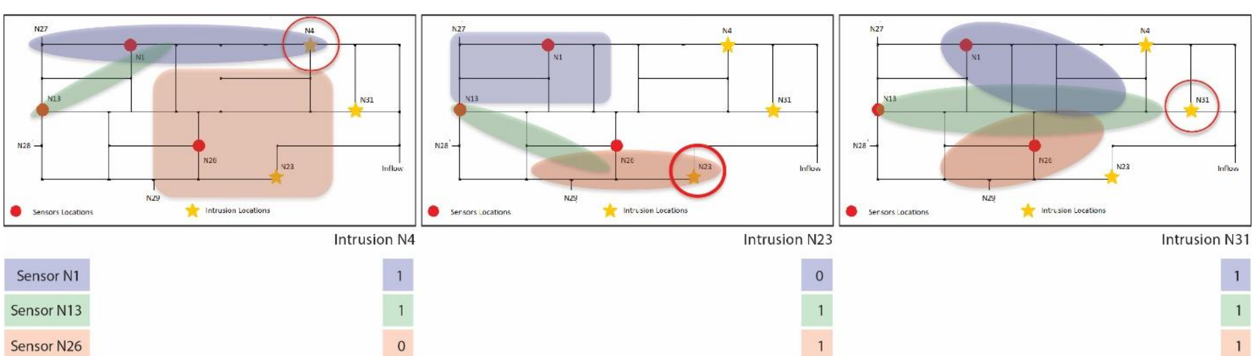


Figure 30: Source back-tracking using sensor response data from different intrusion scenarios.

However, pin-pointing the exact node will require time-variant data, e.g. flow rate from one node to another, dispersion of chemicals etc. and all the pipe lengths connected to the sensor nodes in addition to computational work to cross-reference hundreds of data. Nevertheless, to identify a specific area to search for intrusion, the contour map as shown in Figure 29 can be very beneficiary and help reduce the time to localize intrusion point. The contour lines explain how the chemical is flowing with different concentration. As it seems, the lower part of the network and further downstream nodes can be removed from the search operation for intrusion.

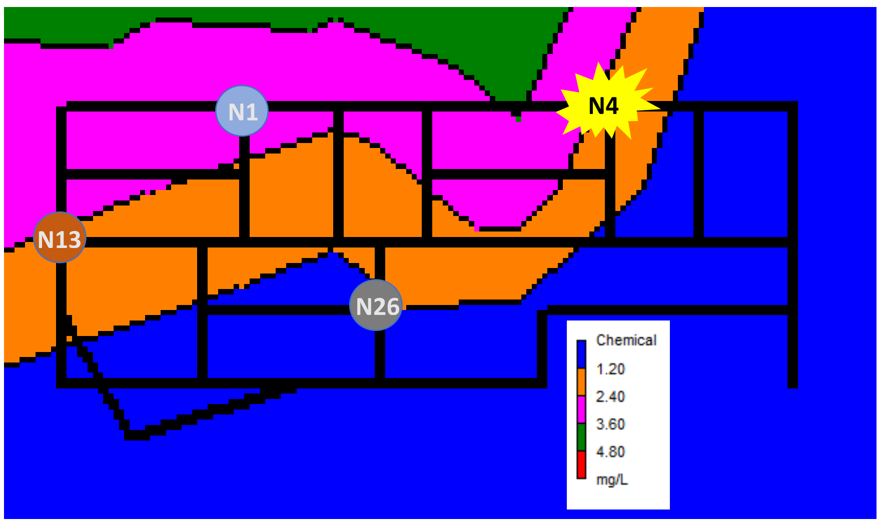


Figure 31: Contour mapping of the chemical concentration flow in network when intruded from N4.

As simple as it seems in this example on how the probabilistic approach can be used to localize intrusion zone or nodes, in real life scenario it can be complicated when there will be thousands of nodes and margin to assimilate a particular node from other can be very narrow. Moreover, large preset database on intrusion events and their flow pattern for a network needs to be stored for usage for referencing a future event. This is where the latest technologies and algorithms will play an important role and advancement of artificial intelligence and machine learning can contribute significantly in this area of research.

5

CONCLUSIONS

The study in this report investigates and share lights into the recent concerning issues of water contamination in distribution network. There has not yet been any perfect and universal solution for the sensors placement location in the water distribution network (WDN) and back-tracking of the source of intrusion. The major issues remain as the cost of sensor placements in large number and computational resources required for storing and analyzing the continuous water quality measurement. Technologies to use the water quality data and predict source of intrusion or back-tracking the contamination spread are limited to uncertainties and ever-changing variables in the hydraulic properties and flow dynamics.

Some of the major outcomes from this study can be pointed out as following:

- Possibilities of correlating the data and chemical transport behavior from EPANET and pilot scaled WDN model.
- Simple matrix and ranking theories are adequate for localization sensor placement in the distribution network.
- Graph theory and complex network theory can add much more to the optimization of sensor location placement in comparison to complex algorithms.
- Considering the real-time data analysis more realistic, EPANET is limited to some degree in modelling and tracking of concentration flow within the pipe network.
- The flow pattern of chemicals in the pipe system is perceived to be laminar and apparently no complete mixing at the nodal junctions is observed.
- Advance sensors measuring different water quality can contribute better understanding of concentration flow which was not possible in this study due to usage of simple pH sensor.
- Back tracking of intrusion is dependent on concentration flowing through the sensors and time-variant data such as flow rate, dispersion rate, diffusivity etc.
- Contour plot (Figure 29) is helpful to describe important contaminated zones and prioritixe nodes in the region.

- Time settings in EPANET is a major drawback of the software for dynamic modelling (e.g. chemical transport in varying time)

The results of the study and the analysis on sensor placement will surely contribute to the future research and provide the authority of water utilities an overview of current issues of contamination detection. The water utility services can also investigate into the simple methods and develop specific sensors components and software for monitoring and tracking contamination intrusion.

6

FURTHER WORKS AND RECOMMENDATION

6.1 Further works

Water distribution network is one of the most important aspects of any community of any country and key part in urban development plans. Ensuring safe clean water is also part of the Sustainable Development Goals (#6) established by UN in 2015. This field of study with water security in distribution network is very important for the sustainable future and the rising awareness concerning deliberate attacks with biological and chemical weapons in recent times. Further work is always appreciated to fill the gaps in research scope, especially when implementing the research in real life projects where there are real threats and numerous uncertainties.

To extend the scope of this study, different methodologies of sensor placement can be investigated and compared to find suitable and possible unique solutions to the problems. Different type of sensors can also be experimented to see the sensitivity and effectiveness that may vary in larger distribution systems with higher flows. However, the spread of contamination in the closed system is something which is really important in source tracking for intrusion event. Using concentration data and flow dynamics in the network it is possible to estimate the location of intrusion but of course with help of computational resources and mathematical models. In more advanced engineering, use of artificial intelligence (AI) and machine learning technologies can contribute in monitoring the distribution network and make quick changes when necessary such as changing the course of flow by closing and opening certain valves during event of contamination.

6.2 Recommendation

The analysis using the experimental results from the real-time experiments are unique and more reliable than what is produced in the EPANET model, hence any software that can overcome the complexity of the issues with the EPANET model for such research work is highly recommended. Apart from public domain software, popular commercial software such as Mike Urban by DHI, WaterGEMS and WaterCAD by Bentley should be tested and experimented. Furthermore, computer program can be made specific to the requirements for source tracking is necessary for future integration to these software or as extensions to support the similar studies with water distribution network.

Most critical task for such scenario of contamination is for water utility companies, who can investigate further into the findings to improve the service and security in distribution. Initially, a small area can be taken as study site and implement research techniques as this report suggests. The results and findings can then be up scaled to implement in larger water distribution networks. It would be interesting and beneficial to see the application the complex network theory and ranking of nodal junction in real scale network and test for the effectiveness and efficiency of sensors' contamination detection.

7

REFERENCES

Axworthy, B. D. H., & Karney, B. W. (1996). Modeling Low Velocity/High Dispersion Flow in Water Distrubution Systems, *122*(3), 218–221.

Aydin, N. Y. (2018). Identifying critical components in water networks using time- dependent data, (July), 1–6.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. Cambridge, UNITED STATES: MIT Press. Retrieved from http://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=3339142

Craun, G. F., Brunkard, J. M., Yoder, J. S., Roberts, V. A., Carpenter, J., Wade, T., … Roy, S. L. (2010). Causes of outbreaks associated with drinking water in the United States from 1971 to 2006. *Clinical Microbiology Reviews*, *23*(3), 507–528. https://doi.org/10.1128/CMR.00077-09

Deuerlein, J., Meyer-harries, L., & Guth, N. (2017). Efficient Online Source Identification Algorithm for Integration within Contamination Event Management System, 1–8.

Dorini, G., Jonkergouw, P., Kapelan, Z., & Savic, D. (2010). SLOTS : Effective Algorithm for Sensor Placement in Water, *136*(December), 620–628.

F. Colombo, A., Lee, P., & Karney, B. (2009). *A selective literature review of transient-based leak detection methods*. *Journal of Hydro-environment Research* (Vol. 2). https://doi.org/10.1016/j.jher.2009.02.003

Hrudey, S. E., Payment, P., Huck, P. M., Gillham, R. W., & Hrudey, E. J. (2003). A fatal waterborne disease epidemic in Walkerton , Ontario : comparison with other waterborne outbreaks in the developed world, *2*, 7–14.

Hu, C., Ren, G., Liu, C., Li, M., & Jie, W. (2017). A Spark-based genetic algorithm for sensor placement in large scale drinking water distribution systems. *Cluster Computing*, *20*(2), 1089–1099. https://doi.org/10.1007/s10586-017-0838-z

Hu, C., Tian, D., & Yao, H. (2015). A cloud-based optimization for the placement of water quality sensors in water distribution system. *Proceedings - 2014 International Conference on Mechatronics and Control, ICMC 2014*, (Icmc), 698–702. https://doi.org/10.1109/ICMC.2014.7231644

Kramer, O. (2017). Genetic Algorithm Essentials, *679*, 11–20. https://doi.org/10.1007/978-3-319-52156-5

Krause, A., Leskovec, J., Guestrin, C., Vanbriesen, J., Asce, M., & Faloutsos, C. (2009). Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks, *134*(6), 516–526.

Laine, J., Huovinen, E., Virtanen, M. J., Snellman, M., Lumio, J., Ruutu, P., … Kuusi, M. (2011). An extensive gastroenteritis outbreak after drinking-water contamination by sewage effluent, Finland. *Epidemiology and Infection*, *139*(7), 1105–1113. https://doi.org/10.1017/S0950268810002141

Nardo, A. Di, Giudicianni, C., Greco, R., Herrera, M., Santonastaso, F., Scala, A., & Civile, I. (n.d.). Sensor placement in water distribution networks based on spectral algorithms, (July 2018), 1–8.

Nicklow, J., Asce, F., Reed, P., Asce, M., Savic, D., Dessalegne, T., … Asce, M. (2010). State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management, *136*(August), 412–432.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J., Hart, W. E., Phillips, C. a, … Pierro, F. (2008). The Battle of the Water Sensor Networks „ BWSN …: A Design. *Journal of Water Resources Planning and Management*, *134*(6), 556–568.

Perelman, L., & Ostfeld, A. (2013). Application of Graph Theory to Sensor Placement in Water Distribution Systems. In *World Environmental and Water Resources Congress 2013*. https://doi.org/10.1061/9780784412947.060

Propato, M., Sarrazy, F., & Tryby, M. (2010). Linear Algebra and Minimum Relative Entropy to Investigate Contamination Events in Drinking Water Systems. *Journal of Water Resources Planning and Management*, *136*(4), 483–492. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000059

Rathi, S., Gupta, R., & Ormsbee, L. (2015). A review of sensor placement objective metrics for contamination detection in water distribution networks. *Water Science and Technology: Water Supply*, *15*(5), 898–917. https://doi.org/10.2166/ws.2015.077

Rossman, L. a., & Boulos, P. F. (1996). Numerical Methods for Modeling Water Quality in Distribution Systems: A Comparison. *Journal of Water Resources Planning and Management*, *122*(2), 137–146. https://doi.org/10.1061/(ASCE)0733-9496(1996)122:2(137)

Sadeghi, J., Sadeghi, S., Taghi, S., & Niaki, A. (2014). Optimizing a hybrid vendor-managed inventory and transportation problem with fuzzy demand : An improved particle swarm optimization algorithm. *Information Sciences, 272*, 126–144. https://doi.org/10.1016/j.ins.2014.02.075

Sankary, N., & Ostfeld, A. (2017). Incorporating operational uncertainty in early warning system design optimization for water distribution system security. *Procedia Engineering*, *186*, 160–167. https://doi.org/10.1016/j.proeng.2017.03.222

Sankary, N., & Ostfeld, A. (2017). Inline Mobile Sensors for Contaminant Early Warning Enhancement in Water Distribution Systems. *Journal of Water Resources Planning and Management*, *143*(2), 4016073. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000732

Sankary, N., & Ostfeld, A. (2017). Scaled Multiobjective Optimization of an Intensive Early Warning System for Water Distribution System Security. *Journal of Hydraulic Engineering*, *143*(9), 4017025. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001317

Sela Perelman, L., Abbas, W., Koutsoukos, X., & Amin, S. (2016). Sensor placement for fault location identification in water networks: A minimum test cover approach. *Automatica*, *72*, 166–176. https://doi.org/10.1016/j.automatica.2016.06.005

Seth, A., Klise, K. a, Siirola, J. D., Haxton, T., Laird, C. D., & Asce, a M. (2016). Testing Contamination Source Identification Methods for Water Distribution Networks, *142*(4), 1–11. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000619.

Simone, A., Ridolfi, L., Laucelli, D. B., Berardi, L., Giustolisi, O., Bari, P., … Duca, C. (2018). Centrality metrics for Water Distribution Networks, (July), 1–8.

Ung, H., Ph, D., Piller, O., Ph, D., Gilbert, D., Ph, D., & Mortazavi, I. (2017). Accurate and Optimal Sensor Placement for Source Identification of Water Distribution Networks, *143*(8), 1–12. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000777.

USEPA, US Environmental Protection Agency, (2000). Epanet 2 users manual, (September).

Xu, J., small, M., VanBriesen, J., Fischbeck, P., & Johnson, M. (2010). Robust placement of sensors in dynamic water distribution systems. *European Journal of Operational Research*, *202*, 707–716. Retrieved from http://www.efdinitiative.org/publications/robust-placement-sensors-dynamic-water-distribution-systems-0

Zeng, D., Gu, L., Lian, L., Guo, S., Yao, H., Hu, J. :O. cost-efficient sensor placement for contaminant detection in water distribution systems. (2016). On Cost-Efficient Sensor Placement for Contaminant Detection in Water Distribution Systems, *12*(6), 2177–2185.