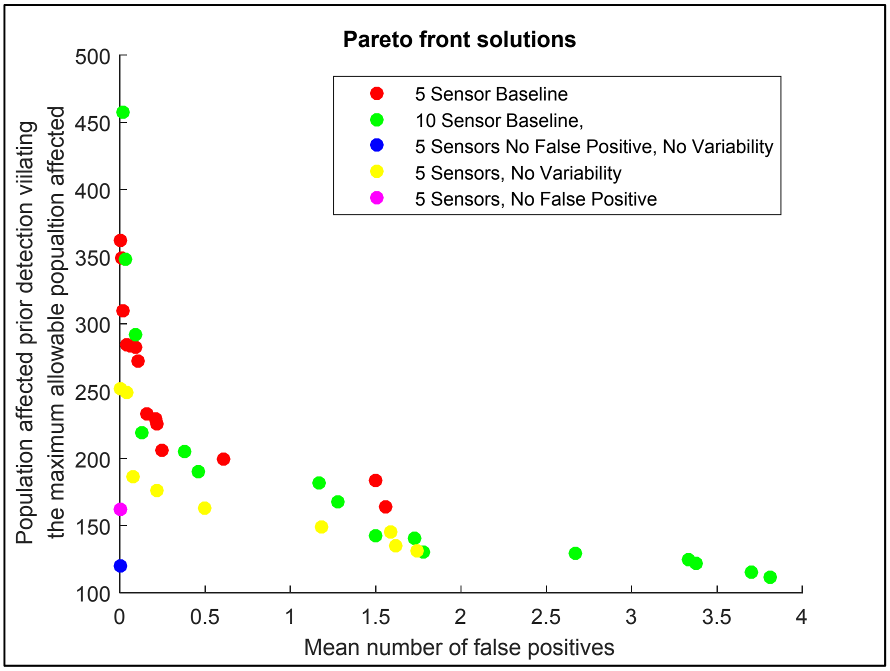

Figure 1: Evaluation of the generated sensor networks showing the ratio of contamination events where the population affected is less than the prescribed maximum allowable population affected. For evaluation, the maximum allowable population is 202 people

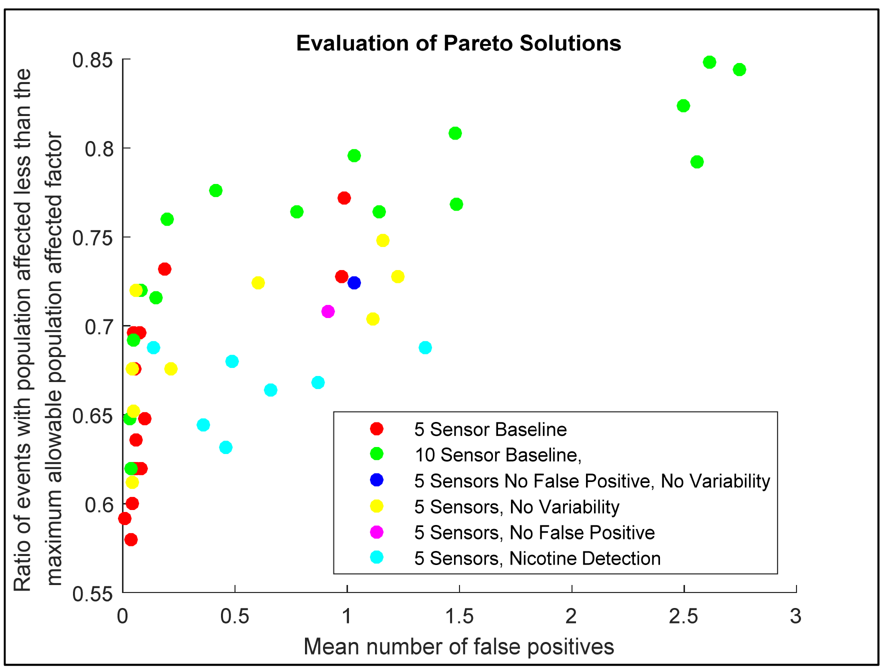

Figure 2: Evaluation of the sensor network designs.

All Pareto front solutions were evaluated against an independent test suite of randomly generated contamination events, and for each event, a randomly generated chlorine input pattern with 30% uniform variably about the baseline chlorine input concentrations was imposed in the simulation event. Results of evaluation are shown in Figure 2. It is clear that the sensor networks designed to detect the contaminant, nicotine, perform poorly when more realistic water quality data are used.

A small sensitivity analysis of the baseline 5 sensor solution was performed to understand the influence of the chosen parameters on the optimization results. The gamma safety factor in the objective function was set to 1, 3, and 5, and evaluated; and the maximum allowable population affected in the objective function was set to 13, 67, and 134 people and evaluated (shown in Figures 3(a) and 3(b)). For solutions with the largest populations affected violating the prescribed maximum allowable population affected (and lowest number of false positives) the gamma factor showed to be less influential than in solutions with lower populations affected. This is indicative that the solutions with lower populations affected generally have larger variability in the respective population affected due to water quality uncertainty. In contrast, changing the maximum allowable population affected all solution's performance equally.

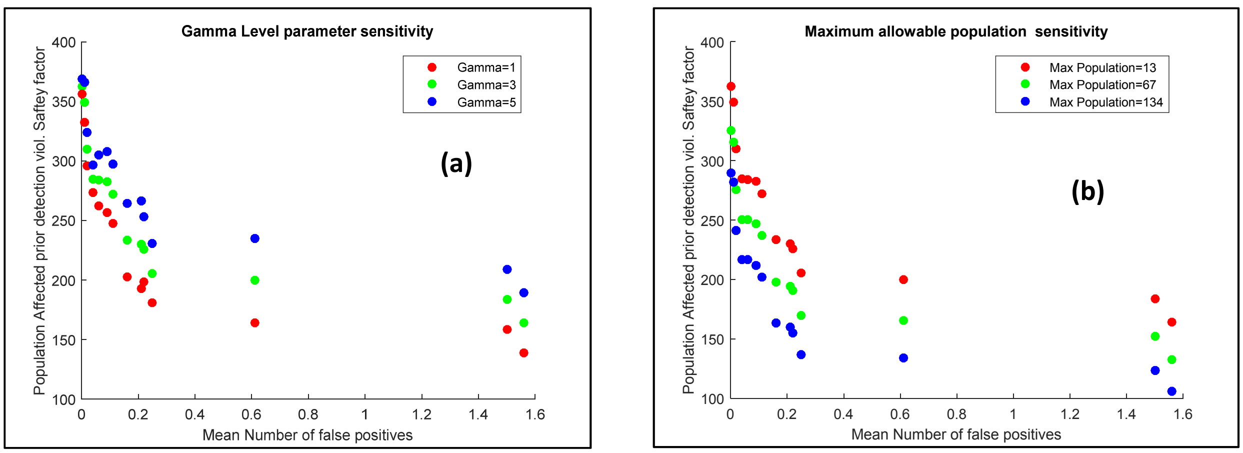

Figure 3: Pareto front sensitivity to the gamma parameter (a) and maximum allowable population affected (b) used in the objective function.

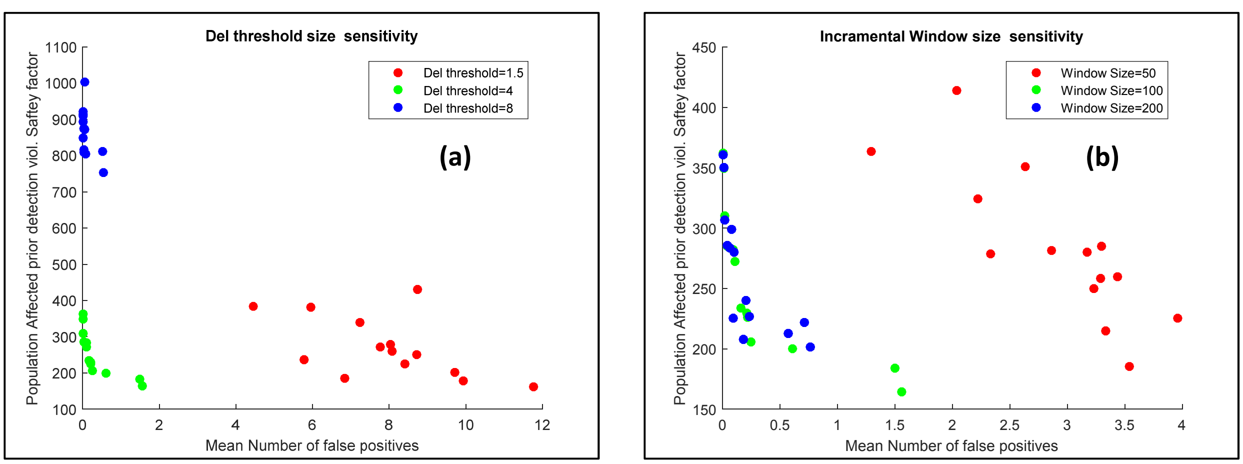

Figure 4: Pareto front sensitivity to the *del* threshold value (a) and incremental window size (b) used in the EDA.

Within the EDA the *del* threshold value was set to 1.5, 4, and 8 and evaluated; and the window size was set to 50, 100, and 200 timesteps (roughly 12, 24, and 48 hours) and evaluated, (Figures 4(a) and 4(b)). Decreasing the window size greatly reduced the performance of the sensor network and EDA, as the population affected violating the maximum allowable population affected increased, and the number of false positive detection increased. Increasing the window size showed inconclusive influence on the sensor network and EDS's performance; although the number of false positive detections are reduced, there are also greater populations affected by the contamination. Increasing or reducing the *del* threshold reduced the sensor network/EDA performance either via an increased number of false positive detections, or larger populations affected.

Conclusions

Previous work has shown the difficulty associated with detecting contamination events when using surrogate water quality parameters, and uncertainty in water quality would further increase the difficulty of event detection. This study proposes an optimization scheme designed to place water quality monitoring stations (sensors) at locations within a WDS that provide a strong water quality signal that is least sensitive to variability in background chlorine concentrations. Incorporating the population affected heuristic in the objective function provides numerous benefits. It implicitly minimizes the time required to detect a contamination event and maximizes the likelihood that a contamination event is detected; especially an event that occurs at a location and time that may affect large populations in the network. The population affected heuristic provides a unique surrogate metric to minimize the false negative detection rate, while minimizing the detection time of a contamination event. Incorporating a second objective to reduce the mean number of false positive detections then specifically choses monitoring station (sensor) locations where the water quality signal provides clear indication of a contamination event, promptly after the event takes place. The incorporation of water quality variability further drives the optimization algorithm to place sensors at network locations where the water quality signal and contamination response is most insensitive to inherent water quality variability in a water distribution system.

This study comprehensively formulated the water quality monitoring station (sensor) placement problem with the event detection problem. Relaxing the assumption of conservative contamination in the sensor placement problem introduces interdependencies between sensor network and event detection algorithm performance. Unifying these two problem may improve the holistic performance of a system to detect contamination in a public water distribution system and improve public water security.

Future work should investigate more advanced EDA algorithms and more effective optimization schemes, potentially even an optimization scheme where EDA parameters are also variables defined by optimization. The study herein was posed on a well-documented, small network; the study herein should be expanded to more realistic sized networks with a diverse selection of contaminants and consideration to more robust objectives and/or min-max objectives.

Acknowledgements

This study was supported by the United States - Binational Science Foundation (BSF), by the Technion Funds for Security research, by the joint Israeli Office of the Chief Scientist (OCS) Ministry of Science, Technology and Space (MOST), and by the Germany Federal Ministry of Education and Research (BMBF), under project no. 02WA1298.

References

Babayan, A., Kapelan, Z., Savic, D., and Walters, G. (2005). "Least-Cost Design of Water Distribution Networks Under Demand Uncertainty." *Journal of Water Resources Planning and Management.* 131(October), 375–382.

Berry, J., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., and Watson, J.-P. (2009). "Designing Contamination Warning Systems for Municipal Water Networks Using Imperfect Sensors." *Journal of Water Resources Planning and Management*, 135(4), 253–263.

Berry, J., Fleischer, L., Hart, W., and Phillips, C. (2005). "Sensor Placement in Municipal Water Networks." *Journal of Water Resources Planning and Management*, 131(June), 237–243.

Comboul, M., and Ghanem, R. (2013). "Value of information in the design of resilient water distribution sensor networks." *Journal of Water Resources Planning and Management*, 139(4), 449–455.

Goldberg, D. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Longman Publishing Co. Boston.

Holland, J. H., (1975) *Adaption in Natural and Artificial Systems.* MIT Press. Cambridge.

Jeffrey Yang, Y., Haught, R. C., and Goodrich, J. A. (2009). "Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: Techniques and experimental results." *Journal of Environmental Management*, Elsevier Ltd, 90(8), 2494–2506.

Kessler, A., Sinai, G., and Ostfeld, A. (1998). "Detecting Accidental Contaminations in Municipal Water Networks." *Journal of Water Resources Planning and Management*.

Lee, B. H., and Deininger, R. A. (1992). "Optimal Locations of Monitoring Stations in Water Distribution System." *Journal of Environmental Engineering*.

McKenna, S. A., Wilson, M., and Klise, K. A. (2008). "Detecting Changes in Water Quality Data." *Journal - American Water Works Association*, 100(1), 74–85.

Oliker, N., and Ostfeld, A. (2014). "Minimum volume ellipsoid classification model for contamination event detection in water distribution systems." *Environmental Modelling & Software*, 57, 1–12.

Oliker, N., and Ostfeld, A. (2015). "Network hydraulics inclusion in water quality event detection using multiple sensor stations data." *Water Research*.

Ostfeld, A., and Salomons, E. (2004). "Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security." *Journal of Water Resources Planning and Management*, 130(5), 377–385.

Perelman, L., and Ostfeld, A. (2013). "Application of Graph Theory to Sensor Placement in Water Distribution Systems." *World Environmental and Water Resources Congress*, 617–625.

Rossman, L. A., "EPANET 2 Users Manual." National Risk Management Research Laboratory, U.S. EPA, Cincinnati.

Shang, F., Uber, J. G., and Rossman, L. A., (2007) "EPANET multi species extension user's manual." National Risk Management Research Laboratory and National Homeland Security Research Center, U.S. EPA, Cincinnati.

Shastri, Y., and Diwekar, U. (2006). "Sensor Placement in Water Networks: A Stochastic Programming Approach." *Journal of Water Resources Planning and Management*, 132(3), 192–203.

Uber, J., Janke, R., Murray, R., and Meyer, P. (2004). "Greedy Heuristic Methods for Locating Water Quality Sensors in Distribution Systems." *Critical Transitions in Water and Environmental Resources Management*, 1–9.

Weickgenannt, M., Kapelan, Z., Blokker, M., and Savic, D. a. (2010). "Risk-Based Sensor Placement for Contaminant Detection in Water Distribution Systems." *Journal of Water Resources Planning and Management*, 136(6), 629–635.

Xu, J., Fischbeck, P. S., Small, M. J., VanBriesen, J. M., and Casman, E. (2008). "Identifying Sets of Key Nodes for Placing Sensors in Dynamic Water Distribution Networks." *Journal of Water Resources Planning and Management*, 134(4), 378–385.

Xu, J., Johnson, M. P., Fischbeck, P. S., Small, M. J., and VanBriesen, J. M. (2010). "Robust placement of sensors in dynamic water distribution systems." *European Journal of Operational Research*, Elsevier B.V., 202(3), 707–716.

Yang, X., Boccelli, D. L., and Asce, A. M. (2016). "Dynamic Water-Quality Simulation for Contaminant Intrusion Events in Distribution Systems." 142(10).

Importance Sampling of Water Distribution System Contamination Events Based on Nodal Neighborhood Populations

Nathan Sankary[1] and Avi Ostfeld, F.ASCE[2]

[1]Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel. E-mail: nesankar@mtu.edu
[2]Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel. E-mail: ostfeld@tx.technion.ac.il

Abstract

A goal of water distribution system security is to ensure that clean water is delivered to consumers. Potential for contaminations and cross connections make it difficult to ensure that the water being delivered to consumers is truly of high quality. The placement of water quality monitoring stations within a WDS system has proven a successful method to prevent the delivery of contaminated water. The locations of monitoring stations in a WDS is critical to the performance of a water quality monitoring station network (early warning system or EWS); ideally sensors will be placed at a limited number of locations which can quickly detect all contamination events. Designing a EWS to protect against every possible contamination event is computationally infeasible; however it is crucial that high impact contamination events will be detected. In this study a contamination event is defined as an intrusion taking place at a specific junction and time in a WDS. A probability distribution is generated according to the portion of a network's population served by water that flows "downstream" from a specific junction at a specific time, within a defined interval of time; this portion of population would be most at risk to exposure of the corresponding junction and time. The newly generated probability distribution is then used for sampling a set of contamination events used to design an EWS. The downstream nodes are calculated using breadth first search and the nodal populations are calculated according to the temporal demands. The relative consequence of a junction being contaminated is calculated using steady state and dynamic hydraulic models; elevating the need to perform numerous complex water quality simulations. Monitoring station networks designed using the proposed importance sampling technique and traditional random sampling are compared.

Introduction

Delivery of clean, high quality water to consumers is the primary goal of a water distribution system (WDS). Ensuring that the water delivered at consumer taps is truly clean can be a difficult task for numerous reasons, including: water quality deterioration with time, the size and complexity of water distribution system's pipe network, and the general spatial and temporal sparsity of points of known water quality. The large distances and long travel times between points of know water quality and user taps introduces vulnerability in to WDS operation. Placing monitoring stations at critical locations throughout a WDS has shown to be an effective method to reduce the risk of delivery of low-quality water.

Using strategically placed monitoring stations throughout a WDS has been extensively explored in prior research. Initially, critical sampling locations were identified (Lee and Deininger, 1992) to cover the largest fraction of network demand. Building off of the work of Lee and Deininger (1992), Kessler et al. (1998) proposed an early warning system for detecting accidental network contaminations. This work defined a specific "level of service" as the volume of contaminated water that can be delivered prior detection of a contamination based on all shortest paths of contamination transport identified within a network. Ostfeld and Salomons (2004) employed a Genetic Algorithm (GA) for identifying the best locations within a WDS for placement of a monitoring station. In this case, explicit hydraulic and water quality simulations were performed, and the GA identified the set of monitoring station locations for up to 7 sensors, which met a defined level of service and maximized the detection likelihood, and detection redundancy. Berry et al. (2005) formulated the water quality monitoring station (hereby referred to as a fixed WQ sensor) placement problem as a mixed-integer problem, and employed the CPLEX® solver for optimization and placed sensors throughout networks of up to 470 junctions in size. The Battle of the Water Sensor Networks (BWSN) (Ostfeld et al. 2008) compared the performance of numerous sensor design algorithms (greedy algorithm, greedy randomized adaptive search (GRASP) heuristic, genetic algorithms, and multiple evolutionary algorithms) to minimize objectives including: the mean population affected, the mean detection time, the mean volume of contaminated water delivered, and maximized the ration of detected events.

In fixed WQ sensor placement studies performance is often evaluated against a suite of contamination event scenarios, the basis for this evaluation lies in the uncertainty in knowing what contamination event may be imparted on to a true WDS. Thus, an EWS must be designed to best protect against any possible contamination event. Ideally an EWS would be designed and evaluated against every possible contamination event, however, this can be feasibly impossible for larger, real sized or full resolution WDS models. In consequence an EWS is typically designed and evaluated against a subset of all contamination event scenarios, such that this subset is large enough to represent the "average" effect of a contamination in the network. Using the contamination event subset EWSs are designed to perform "optimally" against this subset of contamination events. In many cases (Uber et al. 2004, Ostfeld et al. 2008, Krause et al. 2008, Xu et al. 2010) EWSs are designed to minimize the mean or worst case of the objective (detection time, volume of contamination water delivered prior contamination detection, population affected prior contamination detection, etc.) evaluated against all possible scenarios (for smaller networks), or a subset of all possible scenarios.

Monte Carlo (MC) sampling is commonly used to generate a random sample of contamination scenarios. A shortcoming of random MC sampling lies in the difficulty to randomly sample rare events; in the case of sampling WDS contamination scenarios, equal probabilities are typically assigned to all contamination events and in large networks the probability of sampling any single event quickly becomes $< 10^{-6}$. For optimizing an mean or expected value of an objective, it may be adequate to use random MC sampling, however for designing a more robust EWSs it is desirable to ensure that the most detrimental contamination events will be detected promptly. Thus, it is crucial that these most detrimental contamination events are included in the contamination event suite used for evaluation. There have been relatively few studies where the most detrimental contamination events are explicitly incorporated in the contamination event suite. Perelman and Ostfeld (2010) employed Cross Entropy (Rubinsstein and Kroese, 2004) to iteratively reconstruct a sampling distribution to favor sampling rare, detrimental events to be included in EWS design. Weickgenannt et al. (2010) employed a simple flow based heuristic to re-weight specific node—time pairs within the sampling space. For each possible node and time of contamination, the respective node – time pair of potential contaminant input was assigned a weight equal to the volume of water that emanated from the node during a time window beginning at the assigned node—time pair time. In the case of Weickgenannt et al. (2010) the time window was equal to the contamination input duration, 16 hours. This methodology is advantageous because it only requires a single hydraulic simulation, and for a given time window does not depend on any specific contaminant characteristics. Building off of these two importance sampling (IS) methodologies, this study proposes and evaluates an IS scheme based on the size of the population served within a defined hydraulic travel time from each node, at each hour of a hydraulic simulation.

Methodology

In the event of a water distribution system contamination, a potentially harmful substance will be transported from the point of intrusion, "downstream" throughout the network. Thus, as water flows from sources to consumers in a WDS, a contamination would likely have the greatest effect on the consumers immediately downstream from the intrusion point. The contamination events that that would have the greatest expected effect on a WDS's population would be those which are located just upstream from large populations. Using only the downstream population as a nodal weighting would place the largest weights on the nodes located near the sources of the networks. However, there may be large travel times between the sources and a majority of the consumers, and a large travel time provides a large window of detection where an event may be detected via a monitoring station or even consumer complaints before large populations are affected. Thus the most detrimental events are expected to take place at locations where large downstream populations are within a short hydraulic travel time of the intrusion point and large amounts of contaminant are consumed before the contamination is realized.

The IS scheme proposed herein is formulated to re-weight node—time pairs (possible contamination intrusion locations and times) according to the size of the local population served downstream from the node. A time window is defined to constrain the downstream population by a hydraulic travel time. Thus, a weighting is assigned to each node-time pair according to the size of the population served at nodes within the defined hydraulic travel time (or neighborhood) from the node of intrusion. This neighborhood represents the all possible consumers that could be delivered contaminant in the event of an intrusion emanating from the respective node and time within a defined time window after intrusion.

For computational analyses, a WDS can be represented as a graph, $G = (V, E)$ where network junctions or nodes represent vertices (V) , and network pumps and pipes represent edges (E) connecting vertices. Given this representation, various graph theory techniques can be employed to study the characteristics of the network including. In this study, a simple breadth first search (BFS) and shortest path algorithm are employed to determine the hydraulic travel times between network junctions located downstream from a network's junction.

A difficulty arises in modeling a WDS as a graph due to dynamic directions and magnitudes flow within pipes. It is difficult to assign a single direction and magnitude to a network's pipe (graph's edge) to represent its behavior based on a dynamic simulation. In this study, network flows are dynamic throughout the defined simulation length, driven by variable consumer demands and associated pump schedules. To represent the dynamic network as a single directed graph, pipe flow directions and velocities are averaged over the defined time window, beginning at the simulated time of intrusion (ie. the time associated with a node time pair).

Proposed Algorithm

1. Using the averaged directed graph as a representation of a WDS, BFS searches through the graph to determine all connected downstream junctions accessible from the current initial index node *(i)*. Equation (1) below shows the calculations of this set of junctions, ds, where N is the set of all nodes in the specific time window averaged graph, and $(i,\ t)$ represents the node—time pair of node i and time t.

$$ds(i,t) = \{n \in N: n\ is\ accessable\ from\ node\ i\ after\ time\ t\} \tag{1}$$

2. For each node identified using BFS (the set $ds(i,t)$), a shortest path algorithm (Dijkstra's) determines the shortest travel times from the initial node *(i)*, to each downstream node in $ds(i,t)$, where time window averaged graph edges (pipes) are weighted with the average time to travel the entire pipe length during the defined simulation time window.

3. The total population neighborhood of the nodes identified using BFS and within the defined travel time of the initial node is assigned to the respective node—time pair. Equation 2 and 3 show the calculations of the nodal neighborhood populations.

$$dsw(i,t) = \{n \in ds(i,t): traveltime(i,n) \leq windowtime\} \tag{2}$$

$$dsPop(i,t) = \sum_{n \in dsw(i,t)} Pop(n) * {d_w(n)}/{d_{total}(n)} \tag{3}$$

where: dsw is the set of downstream nodes within the defined hydraulic time window, $traveltime$ is the calculated shortest travel time from node i to node n, $windowtime$ is the defined hydraulic travel time window, $dsPop$ is the downstream population within the $windowtime$ travel time of the input node, $Pop(n)$ is the population served by node n, d_w is the cumulative demand during the defined $windowtime$, and d_{total} is the total daily demand delivered by node n.

4. The re-weighted node—time pairs are used to estimate a sampling distribution proportional to the potential consequence of a contamination and each node—time pair. This distribution is then sampled and used for generating a contamination event suite used for design of an EWS.

Figure 1 below shows the resultant node sets identified using the proposed algorithms.

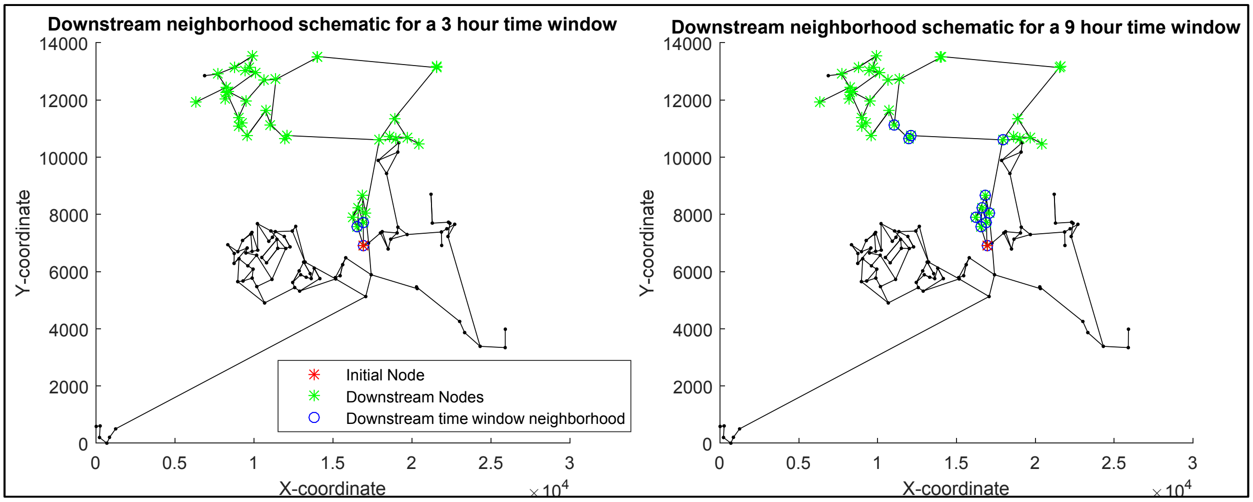

Figure 1: Schematic representation of the identified node sets using the proposed algorithm.

Objective Function

EWS design was driven to minimize the population affected prior contamination detection objective (Ostfeld et al. 2008). The population affected objective was chosen due its relationship to the populations served by a WDS, and its implicit incorporation of undetected contamination events (assuming that after 36 hours the event is realized via human detection.) The population affected prior event detection is calculated in the two equations below: equation 3, the cumulative mass of contaminant consumed, and equation 4, the respective total affected population.

$$M_{i,t,e} = M_{i,t-1,e} + \gamma \Delta t * C_{i,t,e} * \rho_{i,t} \tag{3}$$

$$\mathrm{Pa_e(t)} = \sum_{i=1}^{\mathrm{nodes}} \left[\phi \left\{ \beta \log_{10} \left(\frac{[M_{i,t,e}]/\mathrm{W}}{\mathrm{D_{50}}} \right) \right\} * \mathrm{P_i} \right] \tag{4}$$

where: $M_{i,t,e}$ is the mass consumed (mg), $C_{i,t,e}$ is the concentration of contaminant at junction *i* at time *t*, of event *e (*mg/L*)* such that M(t=0)=0, γ is the daily individual water consumption rate (L/day), Δt is the simulation time step, ρ_i is the temporal demand parameter of junction i at time t (-) calculated as the temporal demand divided by the average demand of node i, $\mathrm{Pa_e(t)}$ is the affected population at time t, of event e, such that $\mathrm{Pa_e(0)} = 0$ φis the standard normal cumulative distribution function, β is a probit slope parameter, W is the mean individual's weight (70 kg), $\mathrm{D_{50}}$ is the dose level that affected an individual with a 50% probability (mg/kg), and $\mathrm{P_i}$ is the population of node i. For further explanation the reader is directed to Ostfeld et al (2008).

Case Study

Experiments were conducted on the BWSN1 network (Ostfeld et al. 2008) composed of 126 nodes, 168 pipes, 1 reservoir, 2 tanks, 2 pumps, and 8 valves. Figure 1 shows schematic representations of the BWSN1 network. EPANET 2 (Rossman, 2000) was used for all hydraulic analysis and sensor network evaluations. Numerous analyses were performed to understand the trends in the importance sampling scheme proposed herein, described below. Contamination event suites were generated using 35, 75, 400,