Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security

Avi Ostfeld, M.ASCE,[1] and Elad Salomons[2]

Abstract: Deliberate contamination is generally viewed as the most serious potential terrorist threat to water systems. Chemical or biological agents could spread throughout a distribution system and result in sickness or death among the people drinking the water. Since September 11, 2001 the U.S. Environmental Protection Agency's water protection task force and regional offices have initiated massive actions to improve the security of the drinking water infrastructure. A methodology is presented for finding the optimal layout of an early warning detection system (EWDS). The detection system is comprised of a set of monitoring stations aimed at capturing deliberate external terrorist hazard intrusions through water distribution system nodes—sources, tanks, and consumers. The optimization considers extended period unsteady hydraulics and water quality conditions for a given defensive level of service to the public, defined as a maximum volume of polluted water exposure at a concentration higher than a minimum hazard level. Such a scheme provides an EWDS for a deliberate terrorist external hazard intrusion, as well as for accidental contamination entries under unsteady conditions—a problem that currently has not been solved. The methodology is cast in a genetic algorithm framework for integration with EPANET and is demonstrated through two example applications.

DOI: 10.1061/(ASCE)0733-9496(2004)130:5(377)

CE Database subject headings: Water distribution; Monitoring; Optimization; Evolutionary computation; Water quality; Security; Terrorism.

Introduction

Since September 11, 2001 the U.S. Environmental Protection Agency's (U.S. EPA) water protection task force and regional offices have initiated massive actions to improve the security of drinking water and wastewater infrastructure. This has been carried out by providing grant assistance to large publicly owned drinking water facilities, supporting the development of tools, offering training and technical assistance for small and medium drinking water and wastewater utilities, and promoting information sharing and research to improve treatment and detection methods.

Water distribution systems are spatially diverse. As such, they are inherently vulnerable to physical, chemical, or biological threats. In general, physical disruptions can result in significant economic cost inconvenience, but the direct threat to human health is limited. Contrary to that is contamination intrusion, chemical and/or biological, which is one of the most serious potential threats to water distribution systems consumers.

For harmful contaminants to reach the consumer taps, a deliberate intrusion must occur within the sources or within the pipe network. A contaminant can be injected at any connection to a water distribution system using a pump or a mobile pressurized tank, capable of overcoming the system pressure. Although backflow preventers provide an obstacle, they do not exist at all connections and some might not be functional. Thus, identification of an intrusion must be made with information acquired within the distribution system itself. The ability to monitor water quality within the system is a significant concern. No formal procedures or guidelines exist at present on where and how to locate water quality monitoring stations for drinking water distribution systems security, subject to unsteady hydraulics and water quality conditions.

Presented herein is the development and demonstration of a methodology for finding the optimal layout of an early warning detection system (EWDS), comprised of a set of monitoring stations aimed at capturing deliberate external terrorist hazard intrusions through water distribution system nodes, such as sources, tanks, and consumers. Consideration is given to extended period unsteady hydraulics and water quality conditions for a given defensive level of service to the public, defined as a maximum volume of polluted water exposure at a concentration higher than a minimum hazard level. The main limitation of the proposed methodology is the real-time assumption of the monitoring equipment, a state that is rapidly changing with the recognition of the critical need for on-line water quality hazard monitoring equipment.

Literature Review

Efficient water quality monitoring is one of the most important tools to guarantee a reliable potable water supply to consumers of drinking water distribution systems. A decline in pressure at one or more system nodes can reduce the quantities supplied to consumers, while the accidental entry of contaminants or self-

[1]Senior Lecturer, Dept. of Civil and Environmental Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: ostfeld@tx.technion.ac.il

[2]Director, OptiWater, 25 David Pinsky St., Haifa 34351, Israel. E-mail: selad@optiwater.com

Note. Discussion open until February 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on August 6, 2003; approved on November 10, 2003. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 130, No. 5, September 1, 2004. ©ASCE, ISSN 0733-9496/2004/5-377–385/$18.00.

deterioration of the water quality within the network itself can severely harm public health (Geldreich 1991).

Ideally every system node should be monitored for the maximum protection of public health, which obviously leads to a monitoring system of maximum cost. On the other hand, if one assumes that the water quality does not change as it is distributed through the network, only the sources would need to be monitored. Since phenomena in the network itself [e.g., a deliberate terrorist intrusion, hazards, corrosion, or trihalomethanes (THM) formations] can cause water quality changes, there is a need to monitor within the entire network, and not just the sources. The goal of an EWDS is to reliably identify a contamination event (accidental or deliberate) in source water or distribution systems, in sufficient time to allow an effective response that reduces or avoids adverse impacts that may result from such an event.

Real-time devices for monitoring biological and chemical indicators of contaminations (e.g., e-coli) are the subject of intensive ongoing research activity. In 1990 the EPA promulgated rules requiring that water quality standards must be satisfied at the consumer taps rather than at the source treatment plants. This initiated the need for water quality modeling and raised other problems to assist utilities in meeting that goal.

The control of pressure is commonly achieved by booster pumps and pressure reducing valves, while water quality can be controlled by booster chlorine injections (Pool and Lansey 1997; Boccelli et al. 1998; Boccelli and Uber 2001) and by monitoring. Since the injection process does not affect water flow, it can be shown that if chlorine decay follows a first order reaction, then the response at a node is linear with respect to an injection at a booster location (Boccelli et al. 1998). In a linear system, the response can be scaled by the injection rate. This linearity relationship has been used as the basis of a linear programming problem to determine the booster injection rates that minimize the total mass of chlorine injected, while maintaining a desired chlorine level at all nodes for all times (Pool and Lansey 1997; Boccelli et al. 1998). The decision variables in this problem are the amounts of chlorine to be injected in situ at the boosters.

Optimal booster locations are an extension of the booster operation problem. Here, the decisions are the location of the boosters and their injection rates for a typical day. The objective of the location/operation problem is to minimize the mass of chlorine injected for a desired number of satellite boosters or to minimize the cost of new booster stations, while limiting the amount of chlorine injected. This problem can be regarded as an inverse type of the monitoring location problem, in that the sources of the injected material are to be identified to satisfy desired levels at all locations. The monitoring problem is to identify the source of the contaminant from measurements throughout the system. Pool (2002) formulated and solved the location problem as a mixed integer linear programming problem (MILP) using linear response functions. Tryby and Uber (2001) solved an MILP model using water age as the basis for determining sample "representativeness."

Monitoring water quality varies throughout a distribution system, and verifying that acceptable water quality is delivered to consumers is not straightforward. EPA regulations require that samples be taken at locations representative of the water quality in the system. "Representative" has not been explicitly defined, and several approaches have been developed to quantify this term.

Lee and Deininger (1992) developed a procedure based on steady-state flow under one or more demand patterns. The premise is that sampling at a location supplied by upstream nodes provides information about water at the upstream nodes. With this idea, a set covering problem for the monitoring locations is formulated to maximize the representativeness of the water samples. Nodal contributions are determined using a steady-state source contribution analysis (Boulos and Altman 1993). After flow paths have been defined, the source contribution from all sources can be determined. Each node in succession is then considered as a source node, and the distribution of flow from this node is computed. A matrix is then formed that identifies upstream nodes supplying a downstream node, and the proportion of water that passed through the upstream node (defined as the water fraction) to the downstream junction. Subsequently, the sampling design question is how to maximize the coverage of the water in the distribution system with a minimum number of measurements. Lee and Deininger solved the optimization problem using an integer programming method. Next, Kumar et al. (1997) applied a greedy heuristic-based algorithm to the same problem. The algorithm provided similar results for the small example solved by Lee and Deininger, but no proof of global optimality was shown. Al-Zahrani and Moied (2001) applied a genetic algorithm (GA) to solve the same model.

From an engineering perspective, the demand coverage method (Lee and Deininger 1992; Kumar et al. 1997; Al-Zahrani and Moied 2001) has several limitations. First, it only considers steady-state water quality conditions that are usually not achieved. Second, the method does not consider the residence time that water spends in the system and temporal variation of the water quality. Finally, information is only considered in the upstream direction and coverage does not extend in the downstream direction. Harmant et al. (1999) modified the objective function to introduce time dependence and water quality into the demand coverage model. The new form weighs the sampling toward larger flows and older water. To emphasize nodes with lower water quality, Woo et al. (2001) further modified the objective by applying weights at each term by normalizing the concentrations by the source values. Thus, nodes with lower water quality received higher weights in the objective function.

All of the aforementioned models use the notion that if the downstream water quality is acceptable, then water supplied before reaching that node must be acceptable. The tendency in the optimal solution is then to install meters at downstream locations and where a mixture of flows exists. This basic idea holds for the EWDS scheme, except that installing monitors at distant points will miss or delay the detection of intrusions until many consumers have been affected.

Kessler and Ostfeld (1998), Kessler et al. (1998), and Ostfeld and Kessler (2001) presented and applied a design methodology for detecting random accidental contamination intrusions in municipal water distribution networks. The methodology is capable of identifying an optimal set of monitoring stations, for a given level of service, allowing the capture of an accidental contamination intrusion to the system. The level of service is defined as the maximum volume of consumed contaminated water prior to detection. The methodology involved (1) the establishment of an auxiliary network representing all possible flow directions in the system for a typical demand cycle; (2) the use of an all shortest paths algorithm for the identification of domains of pollution; and (3) a minimum covering set algorithm for choosing the optimal locations of the monitoring stations. The main shortcoming in the work of Kessler and Ostfeld (1998), Kessler et al. (1998), and Ostfeld and Kessler (2001) lies in not taking into account water dilution and water quality changes as they are distributed in the network, and in not explicitly addressing the un

J. Water Resour. Plann. Manage., 2004, 130(5): 377-385

steady state of the hydraulics and contamination flow.

Recently, Bahadur et al. (2001, 2003) developed the PipelineNet model that simulates the fate and transport of potentially introduced contaminants in water distribution systems. The model simulates the flow and concentration of a biological or chemical agent injected at a system node and assesses the resultant population at risk.

Definitions and Assumptions

An *early warning detection system* is defined as a distributed set of monitoring stations that constantly (i.e., in real time) monitor hazardous water quality constituents; a *pollution event* is a deliberate terrorist injection of a contaminant at one or several nodes of the distribution system at given times, injection durations, and concentrations; the *level of service* (LOS) is the *maximum volume* (MV) of polluted water exposed to the public at a concentration higher than a *minimum hazard level* (MHL) (e.g., an MV of 1,000 m^3, at a concentration higher than an MHL of 0.3 mg/L); a *domain of pollution event* (DOPE) is the set of all contaminated nodes that correspond to a given pollution event and LOS (i.e., all nodes whose concentrations are higher than the MHL, and whose cumulative volume equals the MV); a *domain of detection of a node* (DODN) is the set of all pollution events detectable by a monitoring station located at a specific node; and *redundancy* is the relative number of pollution events detected by more than one monitoring station.

It is assumed that any node (i.e., source, tank, or consumer) at any given time can become a source of pollution due to an external injection, the spreading of a pollutant downstream of a contaminated node is dominated by advection, and that the detection system is capable of providing real-time monitoring data [i.e., to an Supervisory Control and Data Acquisition (SCADA) system].

Methodology

The methodology is composed of two main stages—construction of a randomized pollution matrix (RPM), and maximum column coverage of the RPM using a GA.

Stage 1—Construction of Randomized Pollution Matrix

The first stage is the construction of an RPM, which is based on the pollution matrix (PM) concept proposed by Kessler et al. (1998). The need for expanding the PM to an RPM arises from the pollution events considered herein—explicit multiple random injections. Thus, realizations are needed to quantify possible pollution events.

The RPM is an $(n \times m)$ binary matrix, where the rows (n) refer to the water distribution system nodes (i.e., sources, tanks, and consumers), the columns (m) refer to a randomized set of pollution events, and 1 and 0 refer to contaminated/noncontaminated nodes, respectively.

The jth column lists all of the contaminated nodes due to a pollution event up to the LOS (i.e., each column is a DOPE), while the ith row lists all of the pollution events that can be detected by a monitoring station located at Node i (i.e., each row is a DODN). The RPM therefore provides a stochastic representation of the consequences of a set of randomized injection events, imposed at the system nodes.

Stage 2—Maximum Column Coverage of Randomized Pollution Matrix Using Genetic Algorithm

GAs (Holland 1975; Goldberg 1989) are heuristic combinatorial search techniques that imitate the mechanics of natural selection and natural genetics of Darwin's principle of evolution. The basic idea is to simulate the natural evolution mechanisms of chromosomes (represented by string structures) involving selection, crossover, and mutation. This is accomplished by creating a random search technique that combines survival of the fittest among string structures with randomized information exchange.

A typical GA form involves the following three main stages: (1) *initial population generation*, where the GA generates a bundle of strings (termed population or generation), with each string being a coded representation of the decision variables (not necessarily feasible); (2) *computation of the strings' fitness*, where the GA evaluates each string's fitness (i.e., the value of the objective function corresponding to each string), giving a "fitness penalty" to infeasible strings; and (3) *generation of a new population*, where the GA generates the next population by performing selection, crossover, and mutation. Selection involves the process of choosing chromosomes from the current population for reproduction according to their fitness values. Crossover involves the partial exchange of information between pairs of strings. Mutation refers to a random change in the location of one of the

Table 1. GA Parameters Used

GA parameter	Description
String	Integer—size NOMS; 1 . . . nn values for each bit
Selector	Roulette
Crossover type	Blending—a linear combination of the selected two parents
Elitism	The best chromosome in each generation is included unchanged in the next generation.
Crossover probability	0.95
Mutation probability	0.02
Number of generations	50
Population size	50
Running time	Maximum of 2 min for a test case with $nn=19$ [Anytown, U.S.A. (Walski et al. 1987)]; RPM size of $(nn, 5nn^2)$; machine—PC 850 MHz, 512 mB random-access memory

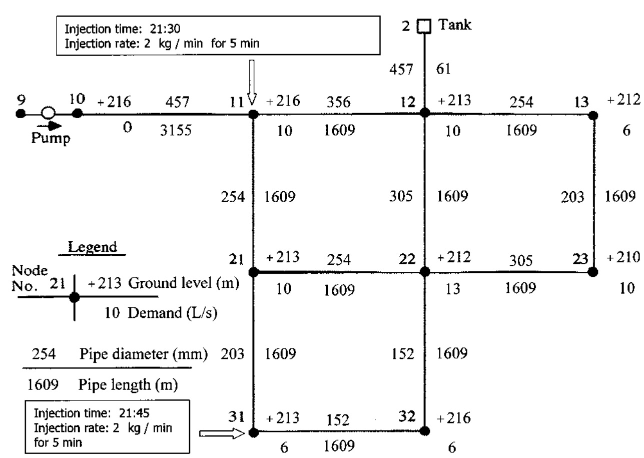

Fig. 1. Example 1—network layout and contamination details of Pollution Event 3 (Fig. 2)

Table 2. Example 1—Demand Pattern

Time of day	Multiplier of average demand
00–02	1.0
02–04	1.2
04–06	1.4
06–08	1.6
08–10	1.4
10–12	1.2
12–14	1.0
14–16	0.8
16–18	0.6
18–20	0.4
20–22	0.6
22–24	0.8

strings. Strings may have binary, integer, or real values. During the last decade, GAs became one of the more robust optimization techniques used in water distribution systems management.

Given a number of monitoring stations that comprise a candidate detection system, a genetic algorithm is applied on the RPM rows for searching the set of rows with maximum column matching. The fitness of each such set corresponds to the coverage of the columns, with a maximum value of 1 if all pollution events are detectable by a candidate set of monitoring stations. The outcome of this stage is an optimal detection system for a given number of monitoring stations and a level of service.

The Appendix provides a summary of the proposed methodology. Table 1 presents the GA parameters.

Applications

The methodology is demonstrated on two example applications through base runs and sensitivity analysis runs. Example 1 is a small "three loop" illustrative distribution system taken from the EPANET user's manual (Example 1) (U.S. 2001) and Example 2 is Anytown U.S.A. (Walski et al. 1987).

Example 1

Example 1 consists of 12 pipes, eight consumers, a source, a pumping station, and an elevated storage tank (Fig. 1). The system is subject to a 24-h demand pattern, as shown in Table 2. Fig. 2 shows an example of an RPM (upper part), and its corresponding polluted volumes and order (lower part), for 10 randomized pollution events, applied to Example 1. Pollution Event 3, for instance, consists of an injection at Node 11 at 21:30 and at Node 31 at 21:45 of a pure pollutant at a rate of 2 kg/min for 5 min (Fig. 1). As a result of that, Nodes 11, 12, 31, and 2 are contaminated up to the level of service defined as 95 L (25 gal.). The minimum hazard level is 0.3 mg/L, and the probability of all nodes to be randomly injected is considered even.

Fig. 3 presents the outcome of Stage 2 of the methodology if applied to the RPM as in Fig. 2, considering two candidate monitoring stations, and given that two exist—at Node 9 (the source) and at Node 2 (the tank). The optimal locations selected are at Nodes 21 and 32. The corresponding detection likelihood (DL) is 1 (i.e., all contamination events are covered), and the redundancy is 4/9 (i.e., four pollution events are detectable by more than one monitoring station). Note that Pollution Event 7 is considered a "success" (i.e., a column of zeros), as "only" 38 L (out of an LOS of 95 L) of polluted water were consumed (Fig. 2). The trade-off between the LOS and the number of monitoring stations that comprise a candidate detection system (NOMS) will be explored further later. Figs. 4–7 provide sensitivity analysis for Example 1.

Fig. 4 shows the trade-off between the DL and NOMS for one, two, and three injections. The RPM for one injection was 3,168 [i.e., number of nodes×(60 min/5 min)×24 h] DOPEs, each corresponding to a single pollutant injection event at each of the nodes, for each 5 min interval, up to 24 h. For two and three

Node	Random contamination events																			
	1		2		3		4		5		6		7		8		9		10	
	N	T	N	T	N	T	N	T	N	T	N	T	N	T	N	T	N	T	N	T
	9	17:30	32	5:45	11	21:30	12	22:05	32	18:20	11	21:45	13	20:10	21	8:50	32	2:10	9	4:25
	10	17:45	21	6:00	31	21:45	9	22:20	21	18:35	10	22:00	32	20:25	31	9:05	23	2:25	21	4:40
10	1		0		0		1		0		1		0		0		0		1	
11	1		0		1		1		0		1		0		0		0		0	
12	0		0		1		1		0		1		0		0		0		0	
13	0		0		0		0		0		0		0		0		0		0	
21	0		1		0		0		1		0		0		1		0		1	
22	0		0		0		0		0		0		0		0		0		0	
23	0		0		0		0		0		0		0		0		0		0	
31	0		0		1		0		0		0		0		1		0		0	
32	0		1		0		0		1		0		0		0		1		0	
9	1		0		0		1		0		0		0		0		0		1	
2	0		0		1		1		0		1		0		0		0		0	
	Polluted volumes and polluted order																			
10	0 (2)		0		0		0 (4)		0		0 (2)		0		0		0		0 (2)	
11	94.6 (3)		0		18.9 (1)		37.9 (5)		0		18.9 (1)		0		0		0		0	
12	0		0		56.8 (3)		28.4 (1)		0		56.8 (3)		0		0		0		0	
13	0		0		0		0		0		0		18.9 (1)		0		0		0	
21	0		47.3 (2)		0		0		47.3 (2)		0		0		24.9 (1)		0		66.2 (3)	
22	0		0		0		0		0		0		0		0		0		0	
23	0		0		0		0		0		0		0		0		0		0	
31	0		0		12.6 (2)		0		0		0		0		69.4 (2)		0		0	
32	0		37.9 (1)		0		0		31.5 (1)		0		18.9 (2)		0		44.2 (1)		0	
9	0 (1)		0		0		0 (3)		0		0		0		0		0		0 (1)	
2	0		0		0 (4)		0 (2)		0		0 (4)		0		0		0		0	
CTPV	94.6		85.2		88.3		66.3		78.8		75.7		37.8		94.3		44.2		66.2	

Legend: N = node; T = Time; CTPV = consumed total polluted volume (liters); 94.6 (3) = polluted volume of 94.6 liters (25 gallons) (third node has been polluted)

Fig. 2. Example 1—example of randomized pollution matrix and its corresponding polluted volumes and order for 10 randomized pollution events

J. Water Resour. Plann. Manage., 2004, 130(5): 377-385

MS	Node	Random contamination events																			
		1		2		3		4		5		6		7		8		9		10	
		N	T	N	T	N	T	N	T	N	T	N	T	N	T	N	T	N	T	N	T
		9	17:30	32	5:45	11	21:30	12	22:05	32	18:20	11	21:45	13	20:10	21	8:50	32	2:10	9	4:25
		10	17:45	21	6:00	31	21:45	9	22:20	21	18:35	10	22:00	32	20:25	31	9:05	23	2:25	21	4:40
	10	1		0		0		1		0		1		0		0		0		1	
	11	1		0		1		1		0		1		0		0		0		0	
	12	0		0		1		1		0		1		0		0		0		0	
	13	0		0		0		0		0		0		0		0		0		0	
SMS	21	0		1		0		0		1		0		0		1		0		1	
	22	0		0		0		0		0		0		0		0		0		0	
	23	0		0		0		0		0		0		0		0		0		0	
	31	0		0		1		0		0		0		0		1		0		0	
SMS	32	0		1		0		0		1		0		0		0		1		0	
EMS	9	1		0		0		1		0		0		0		0		0		1	
EMS	2	0		0		1		1		0		1		0		0		0		0	

Redundancy – detection by more than one monitoring station

Legend: N = node; T = time; MS = monitoring stations; SMS = suggested monitoring station; EMS = existing monitoring station

Fig. 3. Example 1—example of optimal locations of two suggested monitoring stations and their corresponding redundancy for randomized pollution matrix of 10 pollution events

injections, the RPM column size was 605 (i.e., 5×11^2). The LOS was 190 L (50 gal.), $\Delta T=15$ min, MHL=0.3 mg/L, the pollution injection rate (PIR)=2 kg/min, PIR duration (DT)=5 min, and uniform probabilities were assigned to all nodes to be randomly injected. Fig. 5 presents the corresponding redundancy of the results in Fig. 4.

Fig. 4 shows an increase in the DL as the number of injections decreases, given a fixed NOMS, and a decrease in the differences between the detection likelihoods of the different injections, as the NOMS increases (e.g., four monitoring stations versus seven monitoring stations). Fig. 5 shows the opposite trend. That is, given an NOMS (e.g., four), as the number of injections increases, so does the redundancy. This can be explained as follows. Since there are more injections situated into the system, the DL decreases for a given NOMS, but the redundancy increases. As more pollutants are diffused throughout the network, the likelihood of more than one monitoring station revealing a contaminant increases.

Fig. 6 shows the trade-offs between the LOS of 95, 190, 285, and 380 L, respectively, to the NOMS, for a single pollutant injection event [i.e., the RPM size is (11, 3,168)], uniform node pollution probabilities, MHL=0.3 mg/L, PIR=2 kg/min, and DT=5 min. It can be seen from Fig. 6 that as the NOMS increases, so do the DL and the redundancy, regardless of the LOS. For a given NOMS, the DL decreases as the LOS increases (e.g., given five monitoring stations DL=0.9997 for LOS=380 L, versus DL=0.8665 for LOS=95 L).

Fig. 7 shows a base run (BR) and 11 sensitivity analysis runs performed with respect to the BR parameters. Sensitivity Analysis 3 (SA3), for example, shows the algorithm response to setting a probability of zero to have a pollution event at Node 21, which results in a DL increase from 0.9842 at the BR to 0.9955. For SA4, the model response to having only one monitoring station at Node 9 results in a DL decrease from 0.9842 (BR) to 0.9725.

Example 2

Example 2 presents the model application to Anytown, U.S.A. (Walski et al. 1987) (Fig. 8). Anytown, U.S.A. is comprised of 34 pipes, 16 consumers, two 1,610 m^3 (56,832 FE^3) elevated storage tanks, one pumping station, and one well. The tanks are assumed to be cylindrical between their minimum and maximum levels. The water level in the well is maintained at an elevation of 3 m (10 ft). The pipes, nodes, tanks, pumping station characteristics, and the 24-h demand flow pattern are exactly as in Anytown, and thus are not repeated herein.

Figs. 9 and 10, and Table 3 provide results for a level of

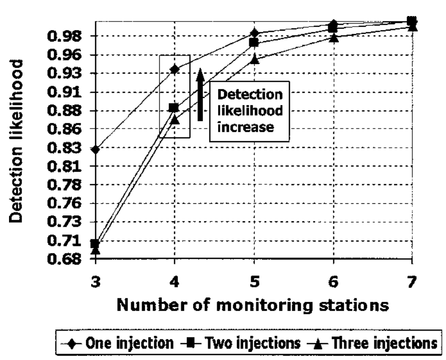

Fig. 4. Example 1—trade-offs between detection likelihood and number of monitoring stations, for one, two, and three injection events

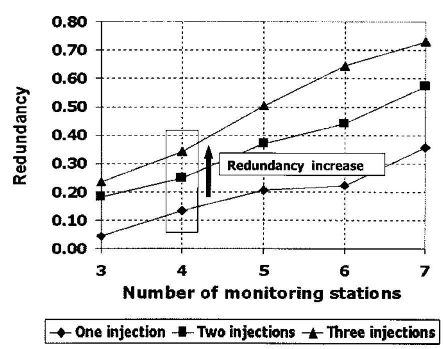

Fig. 5. Example 1—trade-offs between redundancy and number of monitoring stations, for one, two, and three injection events

J. Water Resour. Plann. Manage., 2004, 130(5): 377-385

LOS (liters)	NOMS	Existing monitors		Proposed detection system									Detection likelihood	Redundancy
		Source (Node 9)	Tank (Node 2)	10	11	12	13	21	22	23	31	32		
95	3	+	+		+								0.6528	0.0821
	4	+	+		+				+				0.7715	0.0824
	5	+	+		+				+		+		0.8665	0.0836
	6	+	+		+		+		+		+		0.9201	0.0900
	7	+	+		+	+	+		+		+		0.9618	0.1852
	8	+	+		+	+	+	+	+			+	0.9959	0.2459
	9	+	+		+	+	+	+	+	+	+		1.0	0.2727
190	3	+	+						+				0.8273	0.0445
	4	+	+		+				+				0.9353	0.1348
	5	+	+		+			+	+				0.9842	0.2064
	6	+	+		+		+	+	+				0.9962	0.2232
	7	+	+		+	+	+	+	+				1.0	0.3573
285	3	+	+						+				0.9069	0.0789
	4	+	+		+				+				0.9836	0.1825
	5	+	+		+			+	+				0.9962	0.2955
	6	+	+		+	+		+	+				0.9984	0.3933
	7	+	+		+	+		+	+	+			1.0	0.4258
380	3	+	+						+				0.9331	0.0975
	4	+	+					+	+				0.9981	0.2431
	5	+	+		+				+		+		0.9997	0.2939
	6	+	+		+		+	+	+				1.0	0.3485

Legend: + = optimal proposed monitoring station location; LOS = level of service; NOMS = number of monitoring stations

Fig. 6. Example 1—trade-off between level of service and number of monitoring stations for single injection event

service of 1,893 L (500 gal.). Fig. 9 graphs the trade-off between the number of monitoring stations and the DL for one, two, and three pollutant injection events. Table 3 gives detailed results for two pollutant injections, including the monitoring station locations, the DL, and the average contaminated volume consumed at first detection. Fig. 10 illustrates the trade-off between the number of monitoring stations and the corresponding redundancy for pollution events involving two injections.

Conclusions

Following September 11, 2001, a potential attack on water distribution systems by the injection of contaminants is considered one of the most serious threats to society.

Real-time pollutants monitoring when combined with network modeling can play an important role in tracking and containing the spread of contaminants throughout a water distribution sys-

Case	DATA										RESULTS		
	EML	Pollution event parameters						Genetic algorithm parameters			PMSL	DL	R
		NPP	NIE	DBIE	LOS	MHL	PFD	NOG	PS	MP			
BR	2 ; 9	even	1	NA	190	0.3	2	50	50	0.02	11 ; 21 ; 22	0.9842	0.2064
SA1	2 ; 9	even	2	15	190	0.3	2	50	50	0.02	11 ; 22 ; 31	0.9702	0.3702
SA2	2 ; 9	even	3	15	190	0.3	2	50	50	0.02	11 ; 21 ; 22	0.9488	0.5041
SA3	2 ; 9	P21 = 0	1	NA	190	0.3	2	50	50	0.02	11 ; 13 ; 22	0.9955	0.1667
SA4	9	even	1	NA	190	0.3	2	50	50	0.02	12 ; 21 ; 22	0.9725	0.2109
SA5	2 ; 9	even	3	120	190	0.3	2	50	50	0.02	12 ; 21 ; 32	0.9455	0.3603
SA6	2 ; 9	even	1	NA	190	0.3	1	50	50	0.02	11 ; 21 ; 22	0.9842	0.1910
SA7	2 ; 9	even	1	NA	190	0.1	2	50	50	0.02	11 ; 21 ; 22	0.9852	0.2216
SA8	2 ; 9	even	1	NA	380	0.3	2	50	50	0.02	11 ; 21 ; 22	0.9997	0.3280
SA9	2 ; 9	even	1	NA	190	0.3	2	50	80	0.02	11 ; 21 ; 22	0.9842	0.2064
SA10	2 ; 9	even	1	NA	190	0.3	2	50	50	0.04	11 ; 21 ; 22	0.9842	0.2064
SA11	2 ; 9	even	1	NA	190	0.3	2	80	50	0.02	11 ; 21 ; 22	0.9842	0.2064

Legend

BR = base run; SA1 = sensitivity analysis 1; NA = not available; P21 = 0 – probability zero for node 21 to be contaminated.

DATA

EML = existing monitors locations (nodes); NPP = nodes pollution probability; NIE = number of injection events; DBIE = duration between injection events (minutes); LOS = level of service (liters); MHL = minimum hazard level (mg/L); PFD = pollutant flow discharge (kg/min) (up to a total duration of 5 minutes); NOG = number of generations; PS = population size; MP = mutation probability.

RESULTS

PMSL = proposed monitoring stations locations; DL = detection likelihood; R = redundancy.

Fig. 7. Example 1—sensitivity analysis

J. Water Resour. Plann. Manage., 2004, 130(5): 377-385

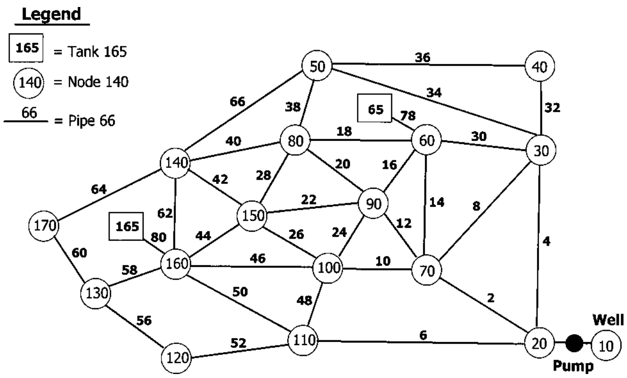

Fig. 8. Example 2—network schematic for Anytown, U.S.A.

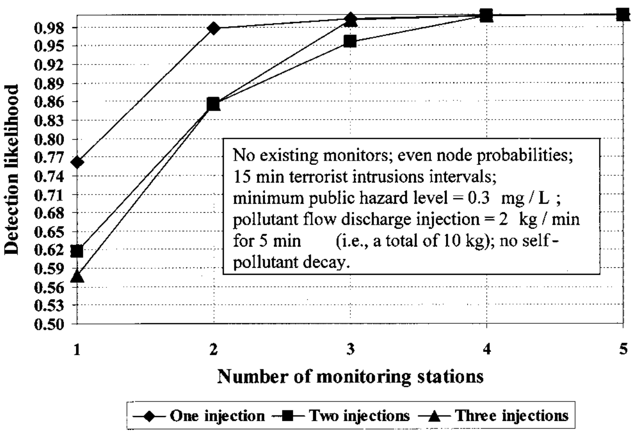

Fig. 9. Example 2—trade-off between number of monitoring stations and detection likelihood for level of service of 1,893 L (500 gal.)

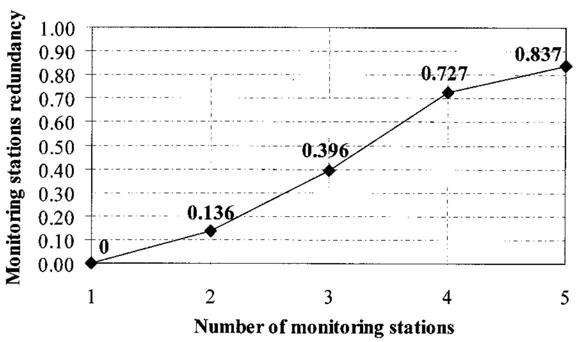

Fig. 10. Example 2—trade-off between number of monitoring stations and their redundancy for two injections and level of 1,893 liters (500 gal.) (Fig. 9)

tem, thus providing an effective early warning system for enhancing drinking water distribution systems security. Presented and demonstrated herein is a methodology for finding the optimal layout of an early warning detection system, comprised of a set of detection stations, for municipal water distribution systems security. The methodology, composed of two main stages (construction of a randomized pollution matrix and maximum column coverage of the RPM using a GA), was demonstrated using two example applications.

The main limitation of the proposed model is the real-time assumption of the monitoring equipment. On-line monitors exist for turbidity, residual chlorine, pH, etc., serving as indirect indicators of pollutant intrusion levels. However, this situation is rapidly changing, with the recognition of the critical need for on-line water quality hazard monitoring equipment, on one hand, and advances in research technology prototype products, on the other.

Future research is suggested to cope with the following topics: considering contamination injections along pipelines (i.e., not necessarily through system nodes); using different LOS measures; taking explicitly into account a delay of the monitoring equipment in enlightening hazard intrusions; using more sophisticated Monte Carlo simulation techniques to reduce the size, and thus the computational efforts of constructing the RPM; and testing the methodology on more complex water distribution systems.

Table 3. Example 2—Detailed Results for Two Injections (Fig. 9)

Number of monitoring stations	Monitoring station locations	Detection likelihood	Average contaminated volume at first detection (L)
One	160	0.618	693
Two	70; 160	0.855	500
Three	90; 110; 140	0.956	628
Four	70; 80; 90; 160	0.998	367
Five	80; 90; 110; 140; 160	0.999	424

Acknowledgments

This research was supported by the Fund for the Promotion of Research at the Technion, by the Technion Grand Water Research Institute, and by the Technion's Counterterrorism Competition.

Notation

The following symbols are used in this paper:

GAG (gc) = genetic algorithm generation at Generation gc;
nn = number of WDS nodes; and
NOPE = number of pollution events.

Appendix. Summary of Proposed Methodology

This appendix summarizes the proposed methodology.

Stage 1—Construction of the Randomized Pollution Matrix

INPUT

1. *Water distribution system data* are as follows: (1) WDS topology; (2) consumers' demand patterns; (3) pipes (diameters, Hazen-Williams friction coefficients, and lengths); (4) pumps (flow-head curves); (5) tanks (elevations, diameters, and initial, minimum, and maximum levels, respectively); (6) constant head sources (constant head data); (7) valves (types and settings); and (8) WDS operational rules.
2. *Pollution event data* are as follows: (1) locations of existing monitoring stations (e.g., at sources or at tanks); (2) settings of node injection probabilities [i.e., probability settings of specific nodes that have a higher likelihood than others to be potential injection locations (e.g., sources or tanks); with no specific data, all nodes are set to have even probabilities to be injected]; (3) pollution event settings—number of injections (NOI) per pollution event (up to three injections), times between consecutive injections (ΔT) (e.g., 20 min), level of

service (e.g., 1,000 m^3), minimum hazard level (e.g., 0.3 mg/L), pollutant injection rate and its duration (DT) [e.g., 2 kg/min for 5 min (i.e., a total amount of 10 kg of an injected pollutant at 100% pure contamination concentration)]; (4) number of pollution events (i.e., the number of columns of the RPM); and (5) maximum simulation duration (MSD)—a maximum predefined simulation duration (MSD$\gg$demand cycle).

3. *RPM construction* is as follows:

Repeat

While tracking contamination movement for quantity and quality

Randomly select an injection starting time

Repeat

- Randomly select a WDS node,
- Inject pollutant at 100% concentration at PIR for DT at the selected node, and
- Move ΔT in time

Until NOI

UNTIL min (time to LOS, MSD) for all NOI

Construct the DOPE 0-1 column in compliance with the given LOS and MHL input data *Until NOPE*.

Stage 2—Maximum Column Coverage of RPM Using Genetic Algorithm

Initialization

- Set the generation counter (gc=0),
- Generate a random initial population genetic algorithm generation [(GAG)(0)], and
- Evaluate the GAG(0) strings fitness.

Main Scheme

Repeat

- Set gc=gc+1,
- Generate GAG (gc) using GAG (gc-1), through selection, crossover, and mutation, and
- Evaluate the GAG (gc) strings fitness

Until gcMax, where gcMax=maximum number of generations.

General Comments

1. Each GA chromosome is an integer string of size NOMS, where NOMS=number of monitoring stations that comprise a candidate detection system. Each integer number of the string can receive any integer value up to nn (total WDS nodes number) [e.g., if NOMS=3, and $nn=15$, then a possible GA string has the form of 7, 6, 3 (i.e., candidate monitoring stations at Nodes 7, 6, and 3)]; 14, 12, and 3 stand for candidate monitoring stations at Nodes 14, 12, and 3, and so on. If a monitoring station already exists at a specific location of a candidate string, or if a node injection probability equals zero at a specific chosen node, then the string is considered nonfeasible.
2. The fitness of each string corresponds to the columns' coverage (i.e., the DL), with a maximum value of 1 if all pollution events are detectable by the candidate set of monitoring stations. For each DL value, a corresponding redundancy is defined as the relative number of instances in which more than one monitoring station has revealed the pollution event.
3. In most cases, an NOPE$=5\times nn^2$ was sufficient to receive a stable solution for two or three pollution event injections. In the case of a pollution event of one injection, each of the nodes was injected at a resolution of 5 min, up to a total of 24 h—the demand cycle considered.
4. If a pollution event has not "reached" the given LOS (i.e., the simulation has reached the MSD), then its corresponding column is filled with zeros. Such an event is considered a success (i.e., no need for monitoring stations for that event, as the pollution event is not hazardous).

References

Al-Zahrani, M. A., and Moied, K. (2001). "Locating optimum water quality monitoring stations in water distribution system." *Bridging the gap: Meeting the world's water and environmental resources challenges, State of the Practice—Proc., World Water and Environmental Resources Congress*, (CD-ROM), ASCE, Reston, Va.

Bahadur, R., Pickus, J., Amstutz, D., and Samuels, W. (2001). "A GIS-based water distribution model for Salt Lake City, UT." *Proc. 21st Annual ESRI Users Conf.*, (CD-ROM), Environmental Systems Research Institute, Redlands, Calif.

Bahadur, R., Samuels, W. B., Grayman, W., Amstutz, D., and Pickus, J. (2003). "PipelineNet: A model for monitoring introduced contaminants in a distribution system." *Proc., World Water & Environmental Resources Congress 2003 and Related Symposia*, (CD-ROM), ASCE, Reston, Va.

Boccelli, D. L., Tryby, M. E., Uber, J. G., Rossman, L. A., Zierolf, M. L., and Polycarpou, M. M. (1998). "Optimal scheduling of booster disinfection in water distribution systems." *J. Water Resour. Plan. Manage.*, 124(2), 99–111.

Boccelli, D. L., and Uber, J. G. (2001). "Evaluation of multi-component model of chlorine residual in water distribution systems." *Water software systems, Proc., CCWI 2001*, U. Bogumil, B. Coulbeck, and J. Rance, eds., Vol. 1, Research Studies Press Ltd., Leicester, U.K., 195–210.

Boulos, P., and Altman, T. (1993). "Explicit calculation of water quality parameters in pipe distribution systems." *Civ. Eng. Sys.*, 10, 187–206.

Geldreich, E. E. (1991). "Investigating the outbreak in Cabool, Missouri for a water supply connection." *Proc., AWWARF/EPA Conf. on Water Quality Modeling in Distribution Systems*, Environmental Protection Agency, Washington, D.C., 55–56.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, New York.

Harmant, P., Nace, A., Kiene, L., and Fotoohi, F. (1999). "Optimal supervision of drinking water distribution network." *Proc., 26th Annual Water Resources Planning and Management Conf. '99—Preparing for the 21st Century*, Chapter 8F180, (CD-ROM), ASCE, Reston, Va.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*, The University of Michigan Press, Ann Arbor, Mich.

Kessler, A., and Ostfeld, A. (1998). "Detecting accidental contaminations in municipal water networks: Application." *Water Resources and the Urban Environment, Proc., 25th Annual Conf. on Water Resources Planning and Management*, ASCE, New York, 272–278.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks." *J. Water Resour. Plan. Manage.*, 124(4), 192–198.

Kumar, A., Kansal, M. L., and Arora, G. (1997). "Identification of monitoring stations in water distribution system." *J. Environ. Eng.*, 123(8), 746–752.

Lee, B. H., and Deininger, R. A. (1992). "Optimal locations of monitoring stations in water distribution system." *J. Environ. Eng.*, 118(1), 4–16.

Ostfeld, A., and Kessler, A. (2001). "Protecting urban water distribution systems against accidental hazards intrusions." *Proc., IWA 2nd Conf.*, CD-ROM, IWA.

Pool, S. (2002). "Optimal operation and location of chlorine boosters." Master's thesis, Univ. of Arizona, Tucson, Ariz.

Pool, S., and Lansey, K. (1997). "Locating and operating disinfection boosters in water networks." *Aesthetics in the constructed environment, Proc., 24th Annual Water Resources Planning and Management*

J. Water Resour. Plann. Manage., 2004, 130(5): 377-385

Conf., ASCE, New York, 738–743.

Tryby, M. E., and Uber, J. G. (2001). ''Representative water quality sampling in water distribution systems.'' *Bridging the gap: Meeting the world's water and environmental resources challenges, State of the Practice—Proc., World Water and Environmental Resources Congress*, CD-ROM, ASCE, Reston, Va.

U.S. Environmental Protection Agency (U.S. EPA). (2001). ''EPANET 2.0.'' ⟨http://www.epa.gov/ORD/NRMRL/wswrd/epanet.html⟩, (July 1, 2004).

Walski, T. M. et al. (1987). ''Battle of networks models: Epilogue.'' *J. Water Resour. Plan. Manage.*, 113(2), 191–203.

Woo, H.-M., Yoon, J.-H., and Choi, D.-Y. (2001). ''Optimal monitoring sites based on water quality and quantity in water distribution systems.'' *Bridging the gap: Meeting the world's water and environmental resources challenges, State of the Practice—Proc., World Water and Environmental Resources Congress*, CD-ROM, ASCE, Reston, Va.

J. Water Resour. Plann. Manage., 2004, 130(5): 377-385