An investigation of the possible scenarios for the optimal locating of quality sensors in the water distribution networks with uncertain contamination

Hamideh Jafari, Taher Rajaee and Sara Nazif

ABSTRACT

One of the ways to reduce the risk of contaminated water consumption is to optimally locate the quality sensors. These sensors warn users in the case of contamination detection. Analyzing the actual conditions of the contamination which enters the network is faced with many uncertainties. These uncertainties include the dose of contamination, time and location of its entry which have received less attention. Also, the uncertainty in the nodes' water demand causes changes in the distribution and contamination diffusion within the network. The main impetus of the present study is to determine the optimal quality sensor locations in the water distribution network in order to reduce the damage caused by contaminated water consumption prior to the contamination event detection. For this purpose, a parameter is defined as the maximum possible damage for calculating which the vulnerability and importance of the nodes have been considered in addition to the uncertainties in the location and time of the contamination entry. The importance of each node differs from that of other ones. Ranking the importance of the nodes is influenced by both land use and covered population ratio. In this study, six scenarios are defined for the contamination event in the water distribution network. These scenarios consider the effects of varying pollutant dose and the contamination input from nodes which are prone to its entry. Also, the NSGA-II has been utilized in order to minimize the damage with minimum number of sensors. The proposed model is evaluated on a real network in Iran. The results indicate that adding only one or two contamination warning sensors to the proposed locations can lead to the decreasing damage caused by the contaminated water consumption from 54 to 82%. According to the proposed method, the best answer for scenarios 1–6 was obtained for 7, 6, 6, 2, 2 and 2 sensors, respectively. The results showed that the slope of the pollution rate diagram does not change much from 6 sensors upwards in the first three scenarios, and from 4 sensors upwards in the second three scenarios. In scenarios 1–3, with 7, 6 and 6 sensors, respectively, in different nodes, the best placement is for 203–224 equivalent attack population, and in scenarios 4–6, with sensors in nodes 4 and 43, the best placement is for 225–279 equivalent attack population.

Key words | contamination intrusion, optimization, uncertainty, water distribution networks, water quality sensor placement

Hamideh Jafari (corresponding author)
Taher Rajaee
Department of Civil Engineering,
University of Qom,
Qom,
Iran
E-mail: *h.jafari@stu.qom.ac.ir*

Sara Nazif
Department of Civil Engineering,
University of Tehran,
Tehran,
Iran

HIGHLIGHTS

- To optimize the quality sensors placement for the most critical possible scenario based on the robust optimization models.
- Using the new concept of minimizing the probable damage caused by the affected population.

doi: 10.2166/wh.2020.099

- In addition to the uncertainty in the contamination entry site, this approach takes into account the time of entry.
- The importance of nodes within the objective function is defined.

INTRODUCTION

Water distribution networks are vulnerable to the intentional or accidental contamination due to the wide range of accessible points. The water contamination possibility from any point of the water distribution network and its diffusion is a serious concern (Yousefi *et al.* 2018, 2019). Contamination that enters the network can seriously threaten the health of individuals in the community. One of the most effective means of ensuring the provision of drinking water with the desirable quality for the consumer is the water quality monitoring within the water distribution network. Accordingly, the optimal sensor placement in order to reduce the effect of contamination entry to the network is of utmost importance. The research background on the placement of quality sensors in the water distribution network goes back to the 1990s.

Lee & Deininger (1992) employed the mixed integer linear programming (MILP) based on the steady-state flow under one or more demand patterns in order to optimally locate the quality sensors with the aim of maximizing the water demand coverage by them. Kumar *et al.* (1997) developed the method used in the study of Lee & Deininger (1992). In order to calculate the water quality at each node, they simulated the residual chlorine concentration and estimated the total water demand coverage values based on it. In this approach, the time between contamination injection and its diagnosis is ignored.

Despite the above-noted problems, many scholars have based their investigations on Lee & Deininger (1992) and continued the monitoring studies of the water distribution network on the basis of this study. The study of Berry *et al.* (2005) was one of the first investigations conducted on the water distribution network monitoring in which some nodes were more likely to be attacked than others. In this study, the MILP method was developed for the location of the sensors with the aim of minimizing at-risk populations. Only four 6-h consumption patterns were used in this study. Shastri & Diwekar (2006) did not agree with the deterministic method used by Berry *et al.* (2005) for quantifying the uncertainties. For this purpose, they presented a nonlinear probabilistic optimization program in order to solve the sensor placement problem. In most of these studies, an estimation of the effect of an attack on a specific node of the network has been considered in the sensor placement optimization problem. This estimation was obtained via analyzing the flow direction and velocity, pipe lengths and the water demand amount of the nodes. The disadvantage of this formulation is that the dilution of the contaminant, concentration and method of attack are not modeled.

Recent research dealing with the optimal quality sensors placement have directly used the pollutant transmission simulation in order to minimize the risk of contamination entry. Berry *et al.* (2006) improved their previous approach in 2005 by considering the contamination concentration at different times. They generalized the optimization problem from a single-objective one (minimizing people affected by contamination) to a more comprehensive formula taking an impact factor for each node into consideration. In their paper, the mixed integer programming method was also considered as a p-median problem. This method is a mixed approach introduced by Resende & Werneck (2004) and is called RW. A large number of contamination events needs to be simulated while implementing optimizations which use the pollutant transmission simulations. For this reason, the implementation of simulations for evaluating any new alignment of the sensors will make the application of optimization techniques such as evolutionary algorithms very costly. To overcome this problem, conducting off-line simulation prior to performing the optimization has attracted the attention of many researchers (Propato 2006; Krause *et al.* 2008). Hence, the simulation time is separated from that needed to run the optimization program. Comboul &

Ghanem (2013) developed the uncertainty analysis for the optimal sensor locating with the aim of maximizing the likelihood of detecting contamination intrusion in the network. Rathi & Gupta (2015) considered the two objectives of maximizing the water demand coverage and likelihood of the contamination diagnosis for the optimal sensors placement. The authors normalized these two goals and weighted in a linear single-objective function. Hu *et al.* (2017) developed a Spark-based genetic algorithm (GA) for the sensor placement in a large water distribution network. In their investigation, the sensor locating criterion in the network was to minimize the detection time by the sensors. Using the complex network theory, Nazempour *et al.* (2018) considered the two objectives of minimizing the contamination coverage time and maximizing the water demand coverage of the nodes for the optimal location of the sensors. Nasirzade *et al.* (2018) developed a probabilistic model considering the possibility of simultaneous contamination entry from multiple nodes and implemented the NSGA-II in order to minimize the risks and costs. Winter *et al.* (2018) integrated various objectives, including the likelihood of contamination diagnosis and detection by the sensor, average time during which the sensor detects the contamination and the estimated effect of the attack in a relationship, and developed a single-objective model based on the greedy algorithm for optimizing the sensors placement in the water distribution network. Recently, Khorshidi *et al.* (2018) utilized the game theory and considered the two objectives of minimizing the sensor cost and the contamination detection time for the optimal sensors placement.

The continuing studies on the optimal locating of water quality sensors indicate the developing place of this issue in the international arena. As a conclusion, given the limited cost, problem complexity and uncertainties existing in the type, time, mass, contamination entry location and water demand, the location of a limited number of quality monitoring sensors should be optimally investigated with the aim of minimizing the damage due to the network pollution.

The present paper aims to optimize the quality sensors placement for the most critical possible scenario based on the robust optimization models.

Also, in this study, a novel formulation is presented to solve the sensor locating problem using the new concept of minimizing the probable damage caused by the affected population (AP). For the first time in this paper, the importance of nodes within the objective function is defined. In addition to the uncertainty in the contamination entry site, this approach takes into account the time of entry. To optimize the sensors placement, using a simulation–optimization approach which directly uses the pollutant transmission simulation, a bi-objective GA-based optimization model (NSGA-II) has been developed in order to minimize the 'maximum possible damage' caused by infected populations along with the number of sensors. In this model, the decision variables are the number of nodes in which the quality sensors are placed. To calculate the damage, the amount of contaminated water consumed (before being detected by the first sensor) is taken into account. The 'maximum possible damage' refers to the maximum amount of damage that can be caused through the contamination entry from any node at any time step. To calculate the 'maximum possible damage' value, the damage matrices are used being calculated prior to the optimization. In other words, the optimal location of the quality sensors in the present model is selected so as to minimize the maximum damage due to the contamination entry (with uncertainties in the location and time of entry).

Also, for the first time in the optimization process, in addition to the population covered, the definition of the importance of nodes is used. In the formula presented for the damage estimating, the nodes are considered to be weighted in order to account for the differences between their importance levels. Six scenarios were defined in this study. The uncertainties in the pollutant mass injected into the nodes, the location of the contamination entry and corresponding time were among the parameters considered in these six scenarios.

CASE STUDY

The proposed algorithm has been implemented on a sample network for a city in Iran as shown in Figure 1. This network has no contamination detecting sensor and consists of 81 nodes, 121 pipes and a reservoir. Figure 1 indicates the nodes which have the potential of injecting contamination. The city's population is about 324,000.

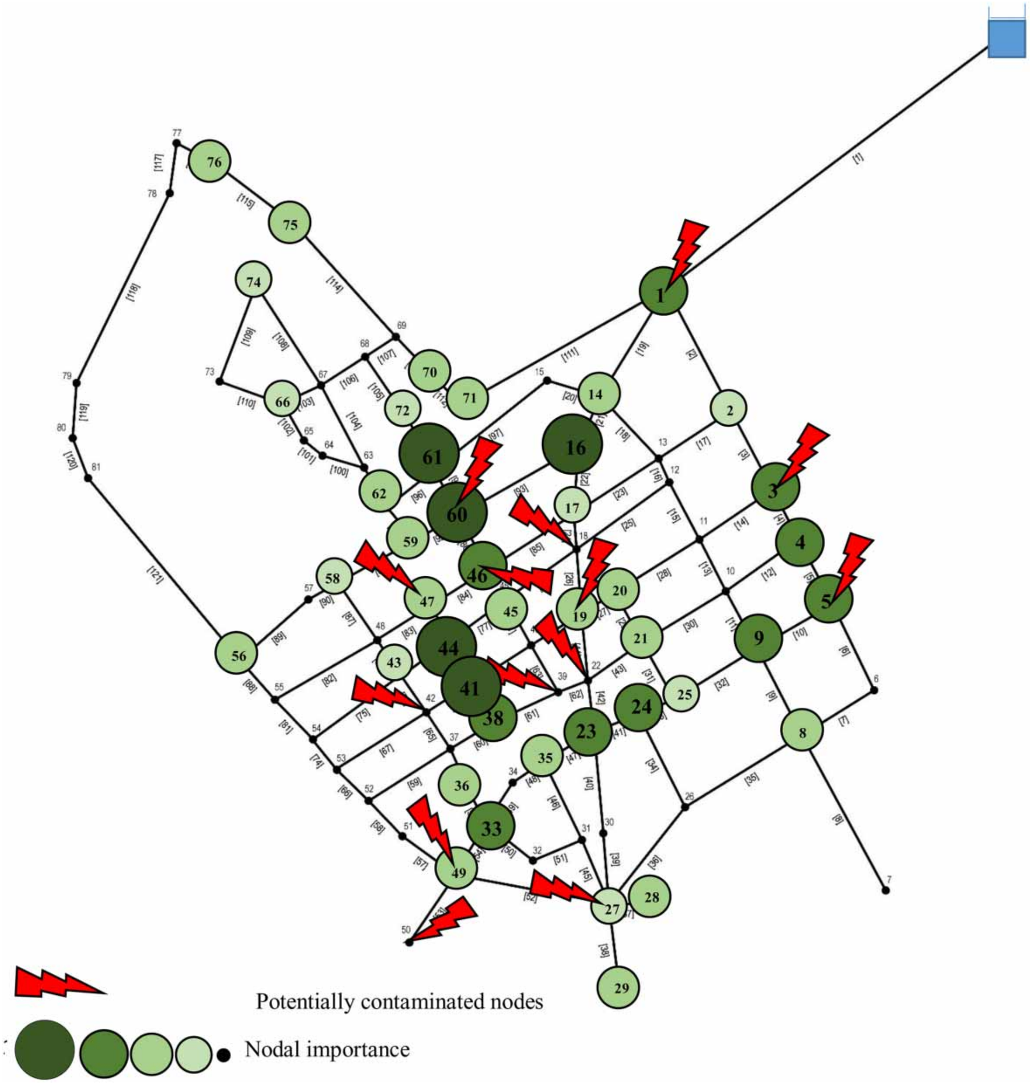

Figure 1 | Potential locations of contamination injection.

The hourly multipliers of the base demand are depicted in Figure 2.

To consider the importance of the nodes in terms of their applications, each node's coverage and population density were considered as five weight categories. Each node's weight is influenced by the importance that the corresponding application will create in increasing the damage caused by the AP if infected. Applications were categorized as residential areas, hospitals, hotels, restaurants, religious sites, universities and passenger terminals. Higher weights are dedicated to the nodes covering multiple uses simultaneously. According to Table 1, the classification of the nodes' spatial importance is performed. A significance coefficient of between 0.05 and 0.005 was assigned to the damage calculation formulation. The higher the significance coefficient, the greater the likelihood of selecting a scenario in which the critical population is covered by the sensor.

Scenarios

In this research, six scenarios were considered. In all scenarios, the contamination (arsenic) is injected gradually for 1 h. The contamination entry is observed every hour of the day. In the first scenario, it is assumed that the

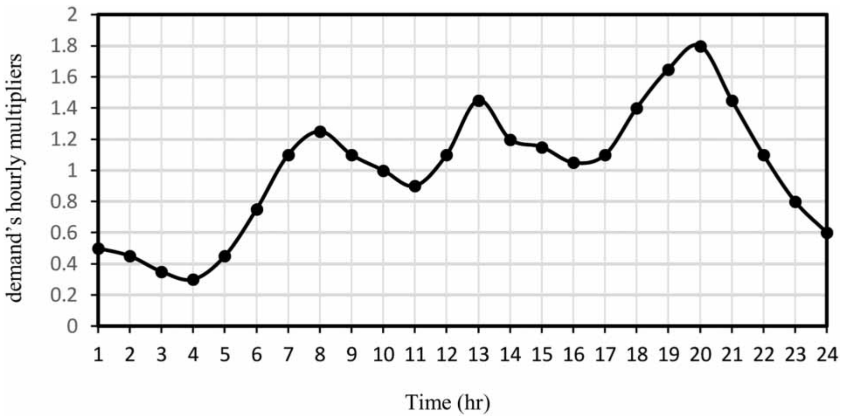

Figure 2 | Base demand's hourly multipliers during a day.

Table 1 | Classify the importance of nodes in terms of location

Criterion	Very high (0.05)	High (0.025)	Medium (0.01)	Low (0.0075)	Very low (0.005)
Nodes	16, 41, 44, 56, 60, 61	1, 3, 4, 5, 9, 23, 24, 33, 38, 46	2, 7, 8, 14,15, 19, 20, 21, 28, 29, 35, 36, 45, 47, 49, 59, 62, 70, 71, 75, 76	17, 25, 27, 43, 58, 66, 72, 74	6, 10, 11, 12, 13, 18, 22, 26, 30, 31, 32, 34, 37, 39, 40, 42, 48, 50, 51, 52, 53, 54, 55, 57, 63, 64, 65, 67, 68, 69, 73, 77, 78, 79, 80, 81

contamination can be entered from all nodes and the pollutant rate in this scenario is 100 mg/s. Similar to the first scenario, the contamination entry from all the nodes is completely accidental in the second and third scenarios. The contamination entry occurs through all the nodes every 24 h, but the pollutant rates in these scenarios are 150 and 200 mg/s, respectively. In these three scenarios, the possibility of contamination entry from all the nodes is identical. In the fourth to sixth scenarios, nodes which are prone to contamination were identified. Therefore, the probability of the contamination entry is 0 and 1. In this section, it is assumed that the ground level components are easily accessible for the attack and are considered as the possible target of contamination. Thus, only a group of nodes are known as vulnerable ones, and the probability of contamination entry through them is 1. In these scenarios, it is assumed that only a subset of nodes is a potential target of attack. The vulnerable nodes were selected on the basis of the ease of access, security features, pressure relief valves, firefighting valves of the storage tanks and pumping stations. One of the factors of the easy access to the nodes is the urban texture consideration. City centers are the initially developed sections that are very ancient with poor protection and high accessibility. Therefore, it can be referred to as a vulnerable node.

The details associated with each scenario are presented in Table 2.

METHODOLOGY

Network hydraulic and quality analysis

In order to be informed of how contamination is distributed through the water network, one must first simulate the flow and its movement in the network. Hence, the hydraulic model is provided according to the hydraulic and physical properties of the network components. Then, the quality simulation is used in order to simulate the pollutant movement within the water distribution network. In the quality analysis, the diffusion of a material in the network is examined, and its concentration is calculated at each time step and at each node. The quality analysis process in a network consists of the three processes of transmission, mixing and volumetric reactions. Khorshidi *et al.* (2018) considered the deterioration rate of arsenic in the water distribution

Table 2 | Water quality parameters of the developed EPANET simulation model

Variable/parameter	Scenarios 1–3 S1	S2	S3	Scenarios 4–6 S4	S5	S6
Time of injection in day	1–24	1–24	1–24	1–24	1–24	1–24
Mass of injection (mg/s)	100	150	200	100	150	200
Duration of injection (h)	1	1	1	1	1	1
Location of injection	All of nodes			13 locations, 1, 3, 50, 46, 47, 5, 60, 27, 49, 22, 19, 18, 39, 42		
Total duration of hydraulic and water quality simulations (h)	48	48	48	48	48	48

system (WDS) as a first-order reaction. In this respect, the transmission involves moving along the flow direction and also in all directions due to the concentration difference expressed by the law of conservation of mass and Fick's law of diffusion as follows:

$$\frac{\partial C_i}{\partial t} = -u_i \frac{\partial C_i}{\partial x} + r(C_i) \tag{1}$$

where u_i is the velocity and C_i stands for the concentration in the ith pipe which is a function of space (x) and time (t). Also, r defines the reaction rate which a function of concentration.

In the mixing process, assuming the complete and instantaneous mixing, the output fluid concentration at the node is equal to the total weight associated with the input flow concentrations to that node as follows:

$$C_j = \frac{\sum_{i \in I_j} Q_i C_i + q_j C_{0j}}{\sum_{i \in I_j} Q_i + q_j} \tag{2}$$

in which, C_j is the concentration at the jth node and C_{0j} stands for any input concentration from outside the network to this node.

As the third process in the qualitative analysis, the volumetric reaction is a function of the concentrations, rate and order of reaction and it is expressed as follows:

$$r = AC^{nr} \tag{3}$$

Here, A is the reaction rate coefficient, and nr stands for the reaction order constant. The value of A will be positive for the formation reaction and negative for the deterioration one.

The qualitative analysis of the network involves the simultaneous solution of the above equations together with those associated with the mixing process in the reservoirs.

The EPANET 2.0 has been implemented for preparing the hydraulic model of the network. This program includes a dynamic-link library called EPANET Programmers Toolkit that allows users to customize the EPANET computing engine with respect to their demands. The output of this program with the extension (.inp) is introduced as the input to the MATLAB environment. This software uses the gradient method to solve the network hydraulic equations and the Lagrangian transfer algorithm for solving the qualitative ones (Rossman 2000).

The simulation of the water distribution network was statically performed before 2004. In this way, the system response was only simulated to a particular and given state of flow (Watson et al. 2005). In this case, a specific moment was only considered as the maximum or minimum hourly flow, and qualitative analysis was ignored. Whereas, in this study, the pollutant transmission simulation is directly used for the quality analysis via the dynamic analysis of the flow. The pollutant materials mix with the water mass in the WDS. Assuming that the pollutant does not react with water, its dynamics is expressed by several parameters such as the water flow direction, intensity of the water flow and dilution of the pollutant concentration along the distance between the contamination entry site into the system and the location of the water consumption. Simulating the pollutant transmission in optimization requires analyzing a large number of contamination events. Accordingly, implementing the simulation to evaluate each new layout, will make the optimization methods very costly.

To solve this issue, the simulation is performed according to the scenarios being defined before conducting the optimization. Hence, the time simulation is separated from that required to run the optimization program.

In this study, arsenic was selected as the pollutant. Arsenic is a cheap and readily available toxic heavy metal, and it is lethal at low doses. Shafiee & Zechman (2013) proposed a relationship based on which the critical arsenic dose can be calculated. Accordingly, the critical dose (CD) of arsenic (mg) is estimated depending on the weight of person at risk of injury (WP) using the following relation:

$$\mathrm{CD} = \mathrm{WP} \times 5^{-8} \tag{4}$$

As in the study of Shafiee & Zechman (2013), it is assumed that the health of an adult weighing an average of 70 kg is endangered with 3.5 mg of arsenic. Therefore, in this study, the population consuming 3.5 mg of arsenic or more is called the affected population. Nasirzade *et al.* (2018) and Khorshidi *et al.* (2018) considered the bulk flow reaction (0.05 day^{-1}) in the quality simulation. They did not consider the wall reaction coefficient for the quality simulation as arsenic does not react with the pipe's wall materials. Janke *et al.* (2017) defined a concept known as the sensor detection limit. According to this definition, the detection limit of the sensor is the concentration threshold above which the sensor is capable of detecting values (binary function).

Optimization model formulation

The present problem approach is the optimal location of the quality sensors for the most critical possible scenario based on the robust optimization models. In this research, the assumptions are given as below:

1. The pollutant (arsenic) is stable, in other words, it does not react with the available substances in the water, and then the separation reactions take place in response to the contaminated water without any delay.
2. Each contamination event occurs at one point in the network.
3. When the sensor detects the contamination, an alert occurs and the network is closed.
4. Each node can be a possible location for the quality sensor placement.
5. For ease of the sensor locating problem analysis, the sensors are assumed to be flawless and detect the pollutant concentration to the allowable limit, without any negative or positive errors.
6. The contamination infusion starts at each hour of the day.

The shorter the time step, the greater the accuracy of the simulation. To evaluate the goals defined in this investigation, the water quality simulation has been performed with a 5 min time step for 2 days. Here, the damage caused by the contamination entry in the absence of sensors is defined as the AP. In this regard, assuming the contamination entry from kth node at hth time step, the damage caused at jth node at tth time step is obtained as follows:

$$M_{h,k,t,j} = G_{h,k,t,j} \times \lambda_t \times I_j \times P_j$$
$$\text{for } j = 1, \ldots, nj,\ t = 1, \ldots, nt,\ k = 1, \ldots, nk,\ h = 1, \ldots, nh \tag{5}$$

where $G_{y,h,k,t,j}$ is a binary variable to detect jth node as contaminated (1) or not being contaminated (0) at tth time step, when the pollutant enters the network from the kth node, at hth time step. Also, λ_t defines the consumption coefficient at the tth time step based on the coefficients of the consumption pattern. I_j is the weight of each node based on its importance (the importance of the nodes is weighted according to the application, the coverage of each node and population density). Besides, P_j stands for the population covered by the jth node.

Damage due to the contamination entry

The effect of contamination entry is usually considered as an objective function by an index. The important thing is that the AP consumes contaminated water before being detected by the sensors, and then, damage becomes zero due to the interruption in the network and no water consumption. For this reason, the damage matrices are calculated before optimizing the placement of the sensors. The damage matrix represents the amount of damage in all possible contamination entry cases (from the kth node at hth time step). For example, if the network has 10 nodes and 48 time steps, the dimensions of each damage matrix will be 48×10, with a total of 480 components to be calculated:

$$M_{h,k,t,j} = G_{h,k,t,j} \times \lambda_t \times I_j \times P_j$$
$$\text{for } j = 1, \ldots, nj, t = 1, \ldots, nt, k = 1, \ldots, nk, h = 1, \ldots, nh$$
$$G_{h,k,t,j} = \begin{cases} 1 & C_{h,k,t,j} \geq C_{allow} \\ 0 & C_{h,k,t,j} < C_{allow} \end{cases} \tag{6}$$

$$1M_{h,k} = \begin{bmatrix} M_{h,k,1,1} & M_{h,k,1,2} & \cdots & M_{h,k,1,nj} \\ M_{h,k,2,1} & M_{h,k,2,2} & \cdots & M_{h,k,2,nj} \\ \vdots & \vdots & M_{h,k,t,j} & \vdots \\ M_{h,k,nt,1} & M_{h,k,nt,2} & \cdots & M_{h,k,nt,nj} \end{bmatrix}_{nt \times nj} \tag{7}$$

Here, $M_{h,k,t,j}$ is the damage at jth node and tth time step due to the contamination entry from the kth node at hth time step, and $G_{h,k,t,j}$ defines the contamination binary concentration at jth node and tth time step due to the same contamination entry data. Also, $C_{h,k,t,j}$ stands for the observational concentration at jth node and tth time step due to the contamination entry from the kth node at hth time step. Furthermore, λ_t, P_j, I_j, n_j and n_t define the consumption pattern coefficients, population vector, importance vector of each node, total number of demand nodes and simulation steps in the network, respectively.

According to the damage function coefficients, it is obvious that the obtained damage represents the virtual population affected by the contamination inflow.

In general, contaminated water will be consumed until being detected by the sensor(s), and then, the damage will be zero. For this purpose, a binary coefficient (a) is used in the present study. Therefore, if there is a quality sensor in the network, the damage will be corrected as in the following relation assuming the flow cutoff upon the contamination detecting:

$$M'_{h,k,t,j} = a_{h,k,t} \times M_{h,k,t,j}, \quad a_{h,k,t} = \begin{cases} 1 & t \leq DT_{h,k} \\ 0 & t > DT_{h,k} \end{cases} \tag{8}$$
$$\text{for } t = 1, \ldots, nt, k = 1, \ldots, nk, h = 1, \ldots, nh$$

in which, $M'_{h,k,t,j}$ is the modified damage based on the sensors performance. Also, $DT_{h,k}$ is the time step in which the first sensor in the network detects the contamination coming from the kth node at the hth time step and the flow is then cut off in the network in order to avoid consuming the contaminated water. Minimizing this time will limit the number of infected people. This objective has been defined as an independent target in many studies such as Kumar et al. (1999); Chastain (2006); Cozzolino et al. (2006); Rathi & Gupta (2013); He et al. (2018); Khorshidi et al. (2018); Winter et al. (2018). However, it has been included within the damage calculation formula in the current research. This time is estimated in the contamination concentration matrix in terms of a time step's coefficients simulated as 5-min ones.

$a_{h,k,t}$ is a binary parameter which causes damage to be calculated only before the flow is interrupted. It is clear that these values depend on the arrangement of the sensors and concentration values.

As can be seen from the above equations, the node and time step of the contamination entry to the network also have a significant impact on the amount of damage in addition to the arrangement of the sensors. However, the total damage of the network is equal to the sum of the damage at all nodes and at all-time steps as:

$$Damage_{h,k} = \sum_{t=1}^{nt} \sum_{j=1}^{nj} M'_{h,k,t,j} \tag{9}$$
$$\text{for } k = 1, \ldots, nk, h = 1, \ldots, nh$$

where $Damage_{h,k}$ defines the damage caused by the contaminated water consumption across the entire network due to the contamination entry from the kth node at hth time step. In this study, to determine the maximum possible damage, the maximum (most critical) damage amount among the damage caused by the contamination entry from different nodes and at various time steps is considered:

$$CriticalDamage = \max_h (\max_k (Damage_{h,k})) \tag{10}$$

Optimization of the quality sensors placement

The location of the sensors should now be determined in such a way as to minimize the *CriticalDamage*. In other words, the location of the sensors should be such that the worst possible damage (assuming the uncertainty in the location and time of the contamination entry) is minimized.

Providing quality sensors is usually costly, and their location and number must be carefully selected. Given the N number of quality sensors, their location should be determined such that the least possible damage is created for the lowest number of the sensors. To this end, the 'maximum possible damage' should be minimized, given the uncertainty in the node and the time of contamination entry. In this regard, the following optimization model is proposed. In this model, Z_1 is the first objective function equal to the *CriticalDamage* minimization. X_s is the design variable which indicates the number of the node on which the Sth sensor is mounted. $\boldsymbol{X}$ is the vector of the decision variables of the problem, the length of which is the number of available sensors (N). Z_2 is the second objective function which minimizes the number of installed sensors.

$$\begin{aligned} Z_1 &= \min(CriticalDamage) \\ Z_2 &= \min(N) \end{aligned} \quad (11)$$

$$1X = [x_1 \quad x_2 \quad \ldots \quad x_s \quad \ldots \quad x_{ns}]_{1\times ns} \quad (12)$$

$$x_s \in \{0, 1\} \quad \text{for} \quad s = 1, 2, \ldots, ns$$

$$N = \sum_{s=1}^{ns} x_s \quad (13)$$

In the above problem, the bi-objective GA is used. This algorithm changes the location of the sensors enough to minimize the maximum possible damage to the network. The general outline of the method presented in this investigation is depicted in Figure 3.

RESULTS AND DISCUSSION

Scenarios 1–3: contamination entry from unspecified nodes at an unknown time

In this section, it is assumed that the contamination can enter the water supply network from all nodes. The pollutant mass is different for these scenarios and varies between 100, 150 and 200 mg/s. The Pareto optimal is shown in Figure 4.

Generally, as the number of available sensors increases, one can reduce the damage due to the contaminated water consumption with the optimal placement. The damage reduction meets higher slope when the number of sensors is lower. However, it decreases with lower slope when the number of sensors exceeds a specific limit (10 sensors).

This means that for more than 6 sensors, increasing the pollutant dose does not have a noticeable impact on Pareto fronts. In Figure 5, this issue is addressed separately. Figure 5 plots the damage variations due to the arsenic entry with the increasing mass for various numbers of sensors (1–6 sensors). Upon placing one sensor within the network, the contamination mass increase from 100 to 150 and from 150 to 200 mg/s, increases the damage by 35 and 33%, respectively. However, this value significantly decreases with an increment in the number of sensors, so that the increasing contamination mass leads to an increase of 25 and 19% in the damage by locating 3 sensors, respectively. Furthermore, the least damage increase occurs upon placing 5 sensors in the network.

The primary aim of a WDS multi-objective model is to reach a solution that is optimal for the single-objective functions. Therefore, one of the solutions set on the Pareto must be selected to report it as the end result of the multi-objective design. To achieve this variety, various methods of conflict resolution exist in this regard. This study applies Young's method (Young 1993), in which one point is selected by maximizing a mathematical equation based on the Pareto's gradient of different points (Young 1993). By using this method, points (Figure 4) are chosen and differentiated by a square, diamond and circle for three scenarios in Figure 4. The optimal layout of selected points in this study is shown in Figure 6 for each scenario. As would be observed, the proposed sensor locations are well distributed throughout the city in order to minimize the maximum possible damage. Investigating the location of sensors in different scenarios shows that this placement is very diverse and not necessarily the same in all scenarios.

The suggested approach in this paper places sensors in nodes that can protect areas of greater importance from contamination. The key idea of this paper is to consider the

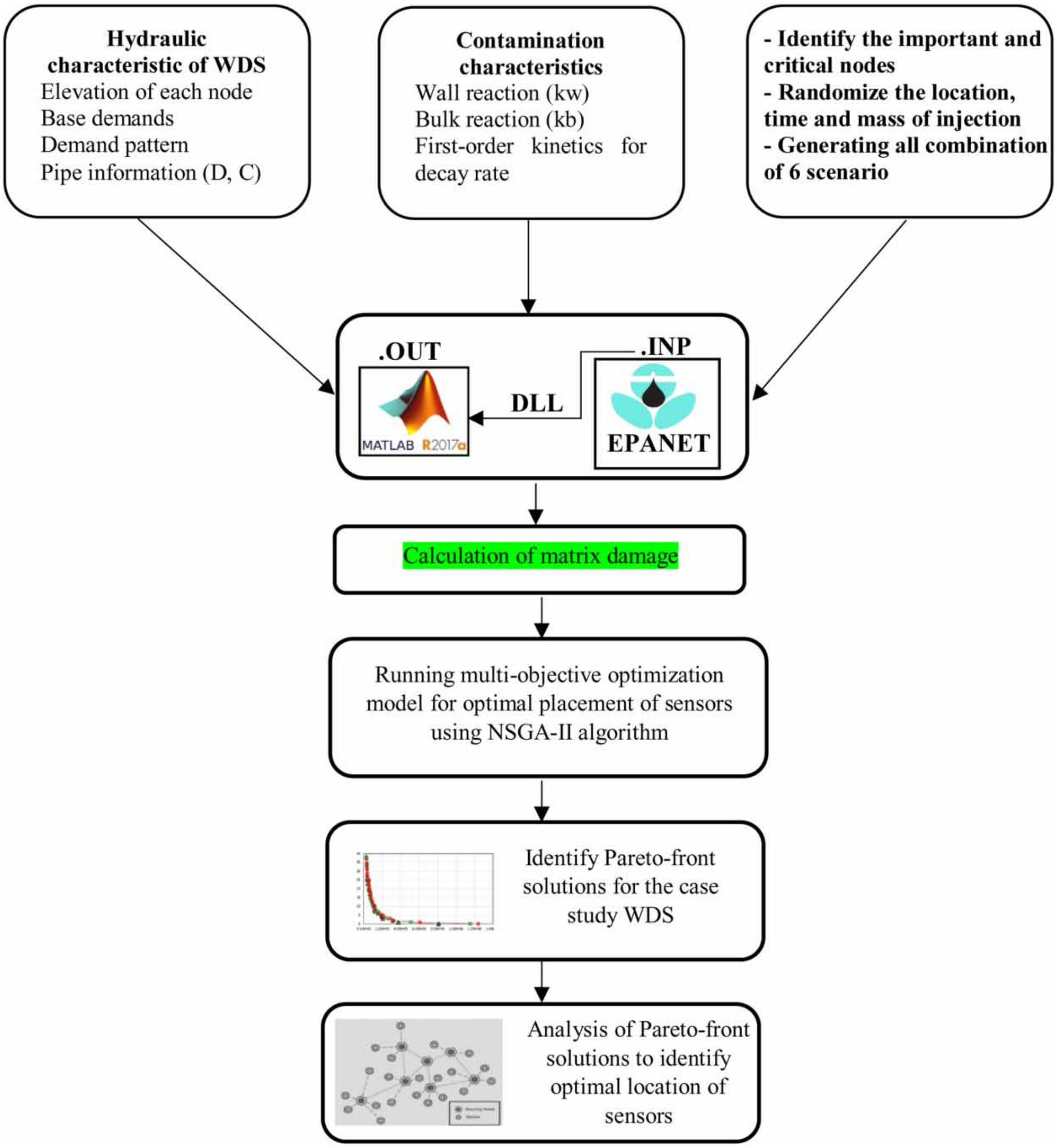

Figure 3 | Framework of simulation–optimization.

importance coefficient of nodes in detecting a critical scenario. Thus, the sensors are located in places in this situation which minimize the critical scenario damage. Bahadur *et al.* (2003) also pointed out the importance and necessity of the role of users' sensitivity on the issue of pollution tracking but did not consider a specific framework for this task. In this paper based on He *et al.* (2018), the effect of the importance of all nodes in calculating the damage was considered. He *et al.* (2018) considered the importance of nodes as an independent probabilistic objective function. Unlike the method of He *et al.* (2018), which considered nodes as two categories and defined the same value for very important nodes, in this paper, nodes are divided into five categories in terms of importance and the significance coefficient was defined within the main objective function.

According to Figure 6, the findings reveal that besides the population covered by each node, the location of the sensors is influenced by the significance coefficient of the

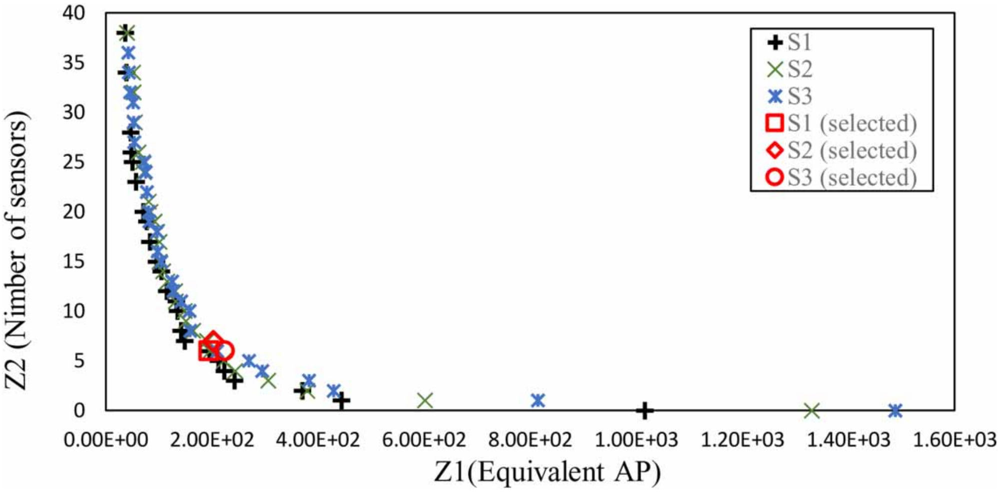

Figure 4 | The Pareto front for contamination from all nodes.

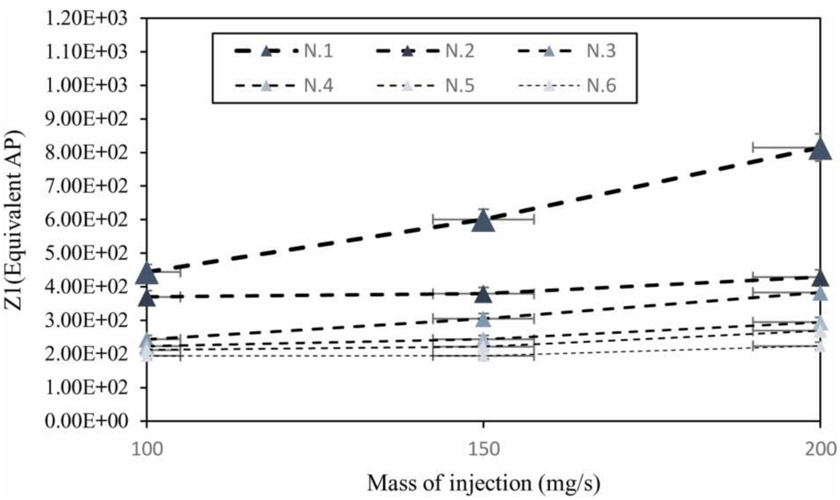

Figure 5 | Damage changes with increasing arsenic mass with number of sensors 1–6 (scenarios 1–3).

nodes. This finding is consistent with the He *et al.* (2018) findings.

Another important result discussed in the studied network is the examination of the most critical location and time step of the contamination entry into the network which cause the greatest possible damage. The entry of pollution is an independent scenario for measuring the damage at every time step. The reason for that is the hourly consumption pattern of each node's population. Therefore, it is expected that the most critical scenario for calculating damages is related to peak consumption hours.

In Figure 7, for each of scenarios 1–3, the most critical time and location of contamination entry are plotted corresponding to different number of sensors (1–10 sensors).

The results of most scenarios indicate that if the contamination enters the network at time steps within the interval 16–19, it will cause the highest damage. As expected, these results are consistent with the trend of the hourly consumption pattern.

The critical input node varies for different sensor placement conditions, but the range of the most critical nodes can be considered as illustrated in Figure 7. It seems that nodes

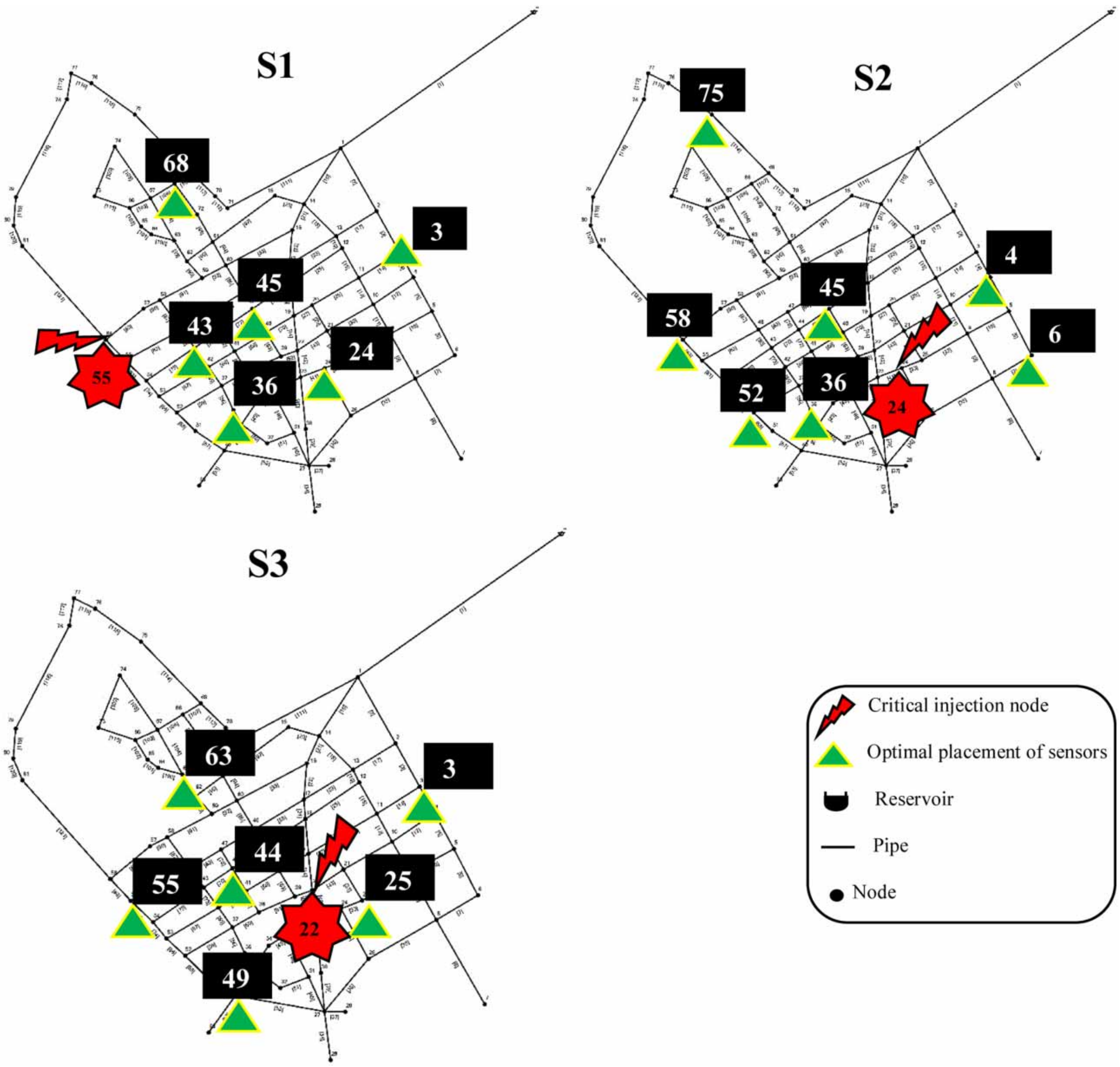

Figure 6 | The optimal layout of selected points determined by the proposed method for three scenarios.

1, 3, 18, 22, 46, 57 and 60 are the ones causing the highest damage on the contamination entry.

Scenarios 4–6: contamination entry from possible nodes at an unknown time

In this section, it is assumed that contamination can possibly enter from only specific nodes having the potential for pollutant entry in terms of the urban context and access to the water network facilities. In other words, nodes in this area can be exposed to intentional contamination with a very high probability. The pollutant mass in these scenarios is 100, 150 and 200 mg/s, respectively. The optimal Pareto is shown in Figure 8.

One can minimize the damage caused by the contaminated water consumption through increasing the number of sensors and their optimal locations. According to Figure 8, it can be seen that the slope of the diagram is higher when only one sensor is placed in the network compared with the cases with more sensors. The slope of the graph changes with a gentle slope as the number of sensors exceeds 3.

Similar to Figure 5, the damage variation due to the arsenic entry with increasing mass is plotted in Figure 9 for various numbers of the quality sensors. It can be

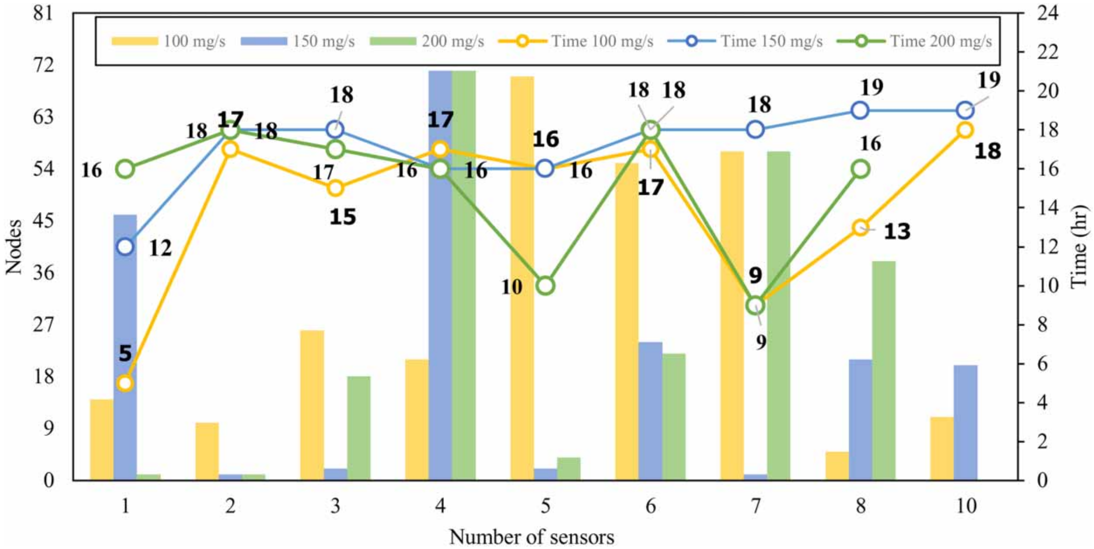

Figure 7 | The most critical time and place of injection corresponding to the number of sensors (scenarios 1–3).

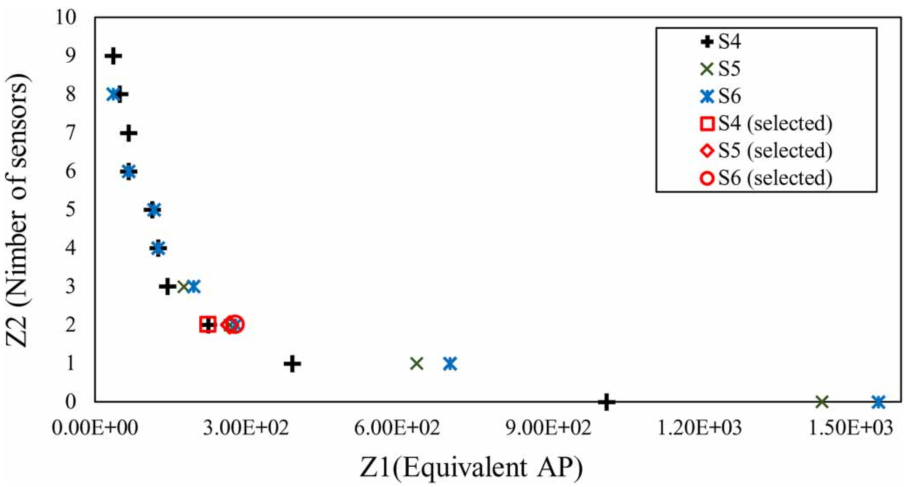

Figure 8 | The Pareto front for potentially contaminated nodes.

concluded that an increment in the contamination entry mass will not cause a noticeable change in the damage if the proper number of sensors with an optimal distribution are placed. In other words, by locating a sensor in the network, the contamination mass increase from 100 to 150 and from 150 to 200 mg/s, increases the damage by 62 and 10%, respectively. As the number of sensors increases, this value decreases markedly such that by locating 5 sensors, the increasing contamination mass only causes an increase of 0.03% in the damage and no change is observable in the network damage increase in the case of placing 6 sensors.

The optimal placement of the proposed algorithm for best solution based on Young's method (Young 1993) is shown in Figure 10 for each scenario.

As can be seen, the proposed locations for the sensors are well distributed throughout the city in order to reduce the maximum possible damage. The investigation of the location of the sensors in the different scenarios indicates that their location is approximately the same in all three.

In Figure 11, the most critical time and location of contamination entry associated with each of scenarios 4–6 are plotted for different numbers of sensors (1–10 sensors).

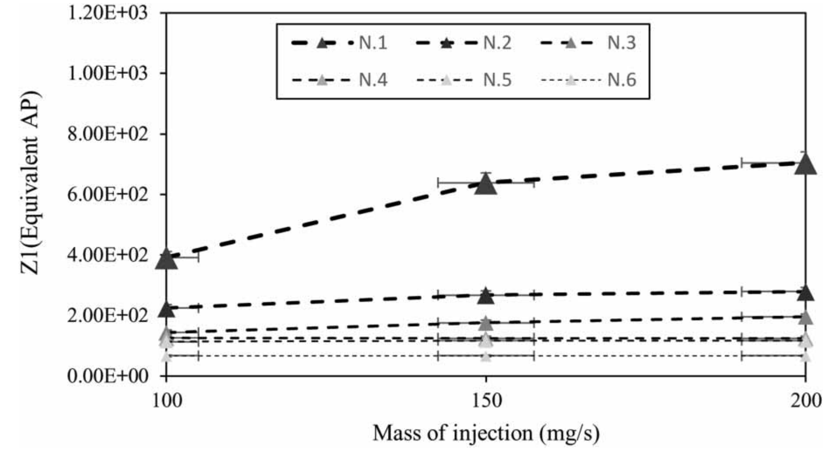

Figure 9 | Damage changes with increasing arsenic mass with number of sensors 1–6 (scenarios 4–6).

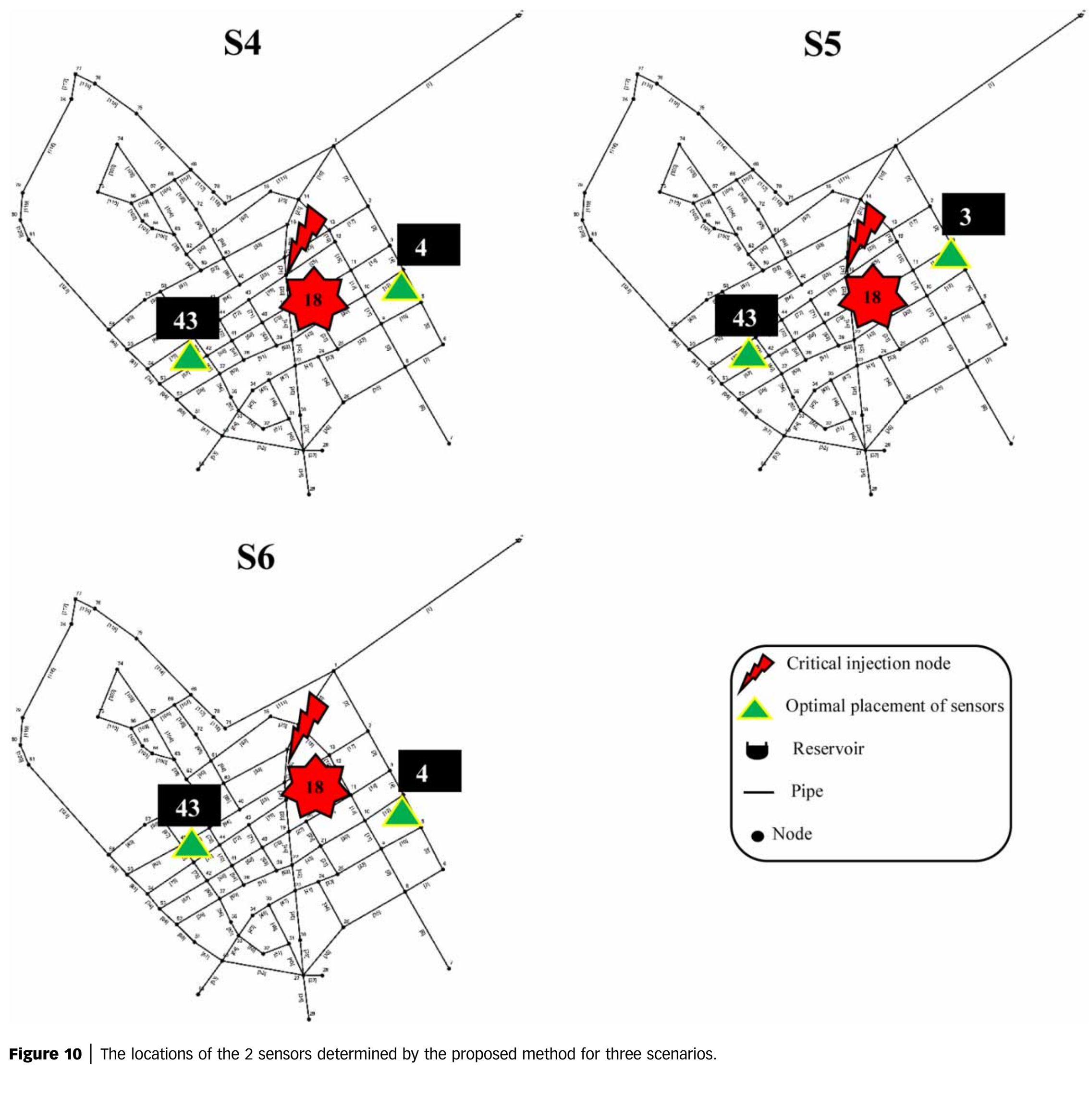

Figure 10 | The locations of the 2 sensors determined by the proposed method for three scenarios.

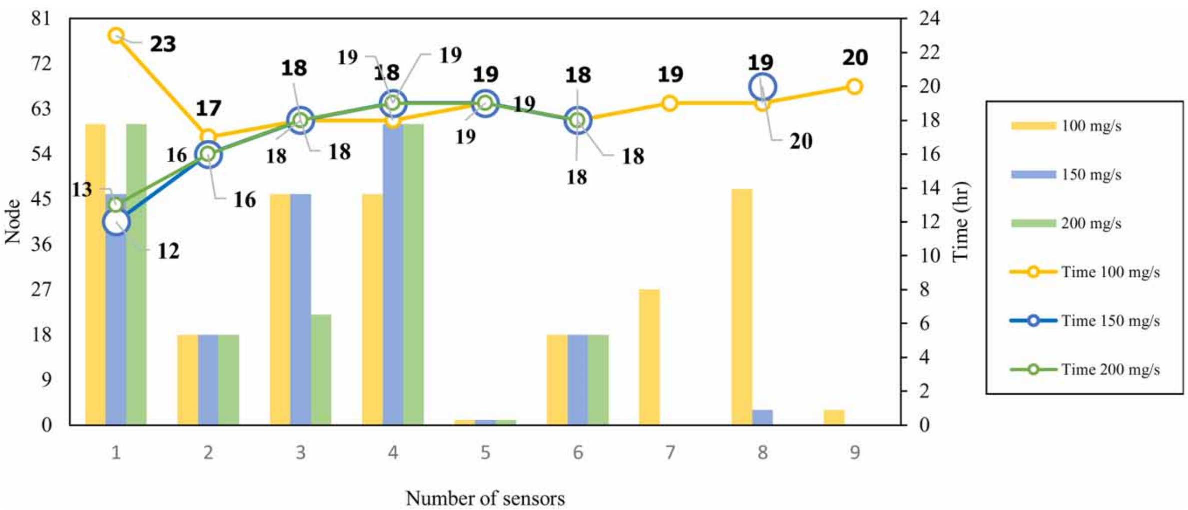

Figure 11 | The most critical time and place of injection corresponding to the number of sensors (scenarios 4–6).

The critical input node varies for different sensor placement conditions, but the range of the most critical nodes can be considered as illustrated in Figure 11. According to the figure, it seems that nodes 1, 3, 18, 22, 27, 46 and 60 are the ones exposing more people to damage upon the contamination entry.

Comparison of the scenarios

Details of the optimal sensors placement are listed in Tables 3 and 4 for various numbers of the sensors. In these tables, in addition to the critical contamination entry times and optimal sensors location, the damage changes are presented for the increasing number of sensors from 1 to 6 relative to the absence of any sensors. Limiting the possibility of contamination from some nodes seems to have a significant impact on Pareto outcomes and the location of sensors. The reason for this is the difference in the selection of the critical scenario in the optimization process of the two mentioned strategies. It should be noted that, in a situation where pollution enters only from certain nodes, the cost and complexity of optimization are much less than a situation where pollution enters from all nodes.

The obtained results indicate that the location of the sensors differs from one scenario to another. However, it can be seen that some nodes, such as nodes 3, 43, 45 and 49, have been replicated in different scenarios, assuming the contamination entry from all nodes or from those with higher potential and play an important role in the network contamination spreading. These nodes appear to be appropriate places to put the sensors.

As can be seen from Table 3, by considering all three scenarios, the damage can be reduced by 50–67% by increasing only one sensor at the optimal location. These values are shown for other scenarios and more sensors. According to the results presented in Table 4, it can be seen that the damage is reduced by 93–95% by adding 6 sensors compared with the case of sensor absence in the entry network.

CONCLUSION

Due to the complex nature of water distribution networks and the great importance of drinking water quality in the health of society, in recent years, several studies have been conducted to maintain and guarantee water quality, especially in the event of terrorist attacks. In this regard, many optimization techniques have been developed to identify the optimal targeting of sensors. The predominant assumption in these techniques was that the probability of contamination of all nodes was the same. While some nodes may be important or sensitive users, even with low

Table 3 | Details of optimal location sensors with a certain number of sensors in the scenarios 1–3

	S1			S2			S3		
Number of sensors	Critical time of pollution entrance	Damage changes due to increase in number of sensors (%)	Optimal placement of sensors	Critical time of pollution entrance	Damage changes due to increase in number of sensors (%)	Optimal placement of sensors	Critical time of pollution entrance	Damage changes due to increase in number of sensors (%)	Optimal placement of sensors
0	10	–	–	16	–	–	16	–	–
1	5	−53	{19}	12	−67	{45}	16	−50	{44}
2	17	−61	{19,61}	18	−72	{9,43}	18	−74	{25,59}
3	15	−74	{3,24,43}	18	−82	{23,45,62}	17	−76	{13,27,59}
4	17	−75	{19,31,48,61}	16	−82	{6,44,52,61}	16	−82	{4,31,41,48}
5	16	−76	{3,25,45,48,49}	16	−83	{6,44,52,61,75}	10	−83	{13,31,41,54,67}
6	17	−80	{3,24,36,43,45,68}	18	−86	{4,6,45,52,58,75}	18	−87	{3,25,44,49,55,63}

Table 4 | Details of optimal location sensors with a certain number of sensors in the scenarios 4–6

	S4			S5			S6		
Number of sensors	Critical time of pollution entrance	Damage changes due to increase in number of sensors (%)	Optimal placement of sensors	Critical time of pollution entrance	Damage changes due to increase in number of sensors (%)	Optimal placement of sensors	Critical time of pollution entrance	Damage changes due to increase in number of sensors (%)	Optimal placement of sensors
0	10	–	–	16	–	–	16	–	–
1	5	−61	{4}	12	−55	{3}	13	−54	{46}
2	18	−77	{4,43}	16	−81	{3,43}	16	−82	{4,43}
3	15	−85	{3,27,43}	18	−87	{3,27,43}	18	−87	{3,48,52}
4	17	−87	{3,27,40,48}	19	−91	{3,27,44,48}	19	−91	{3,32,40,48}
5	16	−88	{3,27,43,45,60}	19	−91	{3,27,40,48,60}	19	−92	{3,40,43,49,60}
6	17	−93	{1,3,40,43,49,60}	18	−95	{1,2,40,43,49,60}	18	−95	{1,3,40,43,49,60}

water consumption, the security of high-quality water should be guaranteed. In other words, water contamination of these users has social effects and more severe consequences than normal users.

In this paper, a method was proposed for optimizing the placement of quality sensors in the water distribution networks using two objectives of minimizing the number of sensors and maximum possible damage considering the minimum detection time. In this regard, six scenarios were defined in which uncertainties related to the location and time step of the contamination entry into the network were considered. The damage matrices were utilized in order to calculate the maximum possible damage. The significance coefficient of each node was taken into account while formulating the present problem.

The importance of nodes according to their location and use is a very important factor in safety of water quality which can be very effective in the identification of optimal solutions. The results revealed that by identifying nodes from which the probability of contamination is very high, it is possible to reduce computational costs and increase the probability of achieving an optimal sensing strategy. The observed results also indicate the necessity of this point. In other words, when there are less than six sensors, the damage and the dose of pollution that enters the network are directly related. The main model for sensors locating is a GA-based bi-objective optimizer that is used to minimize the maximum possible damage with the least number of sensors. This model is usable for any desired network. A real network was selected in Iran as the case study and its results were evaluated in six scenarios. The present achievements illustrated that adding only one sensor can reduce the damage by at least 50%. In this regard, the present proposal could well determine the optimal placement of a certain number of sensors. In this study, a range of appropriate sensor locations was proposed.

In closing, the results of this paper propose a complementary method for considering the importance of nodes in the process of sensor optimization and placement by using a data-based method. Further study of other uncertainties arising in the issue of placement of quality sensors, such as those corresponding to the type of pollutant and duration of the contamination infusion is suggested for future research.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Bahadur, R., Samuels, W. B., Grayman, W., Amstutz, D. & Pickus, J. 2003 PipelineNet: a model for monitoring introduced contaminants in a distribution system. In: *Proceedings of the World Water and Environmental Resources Congress 2003 and Related Symposia*. ASCE, Reston, VA.

Berry, J., Fleischer, L., Hart, W., Phillips, C. & Watson, J. 2005 Sensor placement in municipal water networks. *J. Water Resour. Plan. Manag.* **131** (3), 237–243.

Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G. & Watson, J. P. 2006 Sensor placement in municipal water networks with temporal integer programming models. *J. Water Resour. Plan. Manag.* **132** (4), 218–224.

Chastain, J. R. 2006 Methodology for locating monitoring stations to detect contamination in potable water distribution systems. *J. Infrastruct. Syst.* 252–259. doi:10.1061/(ASCE)1076-0342(2006)12:4(252).

Comboul, M. & Ghanem, R. 2013 Value of information in the design of resilient water distribution sensor networks. *J. Water Resour. Plan. Manag.* **139**, 449–455.

Cozzolino, L., Mucherino, C., Pianese, D. & Pirozzi, F. 2006 Positioning, within water distribution networks, of monitoring stations aiming at an early detection of intentional contamination. *Civ. Eng. Environ. Syst.* **23** (3), 161–174. doi:10.1080/10286600600789359.

He, G., Zhang, T., Zheng, F. & Zhang, Q. 2018 An efficient multi-objective optimization method for water quality sensor placement within water distribution systems considering contamination probability variations. *Water Res.* **143**, 165–175.

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 Comput.* doi:10.1007/s10586-017-0838-z.

Janke, R., Murray, R., Haxton, T. M., Taxon, T., Bahadur, R., Samuels, W., Berry, J., Boman, E., Hart, W., Riesen, L. & Uber, J. 2017 Threat ensemble vulnerability assessment sensor placement optimization tool (TEVA-SPOT) graphical user interface user's manual (Vol. 147). EPA/600/R-08.

Khorshidi, M. S., Nikoo, M. R. & Sadegh, M. 2018 Optimal and objective placement of sensors in water distribution systems using information theory. *Water Res.* **143**, 218–228.

Krause, A., Leskovec, J., Guestrin, C., Van Briesen, J. & Faloutsos, C. 2008 Efficient sensor placement optimization for securing large water distribution networks. *J. Water Resour. Plann. Manage.* **134**, 516–526. doi:10.1061/(ASCE)0733-9496(2008)134:6(516).

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

Kumar, A., Kansal, M. & Arora, G. 1999 Discussion of detecting accidental contaminations in municipal water networks. *J. Water Resour. Plan. Manag.* **125** (5), 308–309.

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

Nasirzade, S. S., Nikoo, M. R. & Montaseri, H. 2018 A risk-based multi-objective model for optimal placement of sensors in water distribution system. *J. Hydrol.* **557**, 147–159.

Nazempour, R., Saniee Monfared, M. A. & Zio, E. 2018 A complex network theory approach for optimizing contamination warning sensor location in water distribution networks. *Int. J. Disaster Risk Reduct.* **30**, 225–234.

Propato, M. 2006 Contamination warning in water networks: general mixed-integer linear models for sensor location design. *J. Water Resour. Plan. Manag.* **132** (4), 225–233.

Rathi, S. & Gupta, R. 2013 Monitoring stations in water distribution systems to detect contamination events. *ISH J. Hydraul. Eng.* **20** (2), 142–150.

Rathi, S. & Gupta, R. 2015 A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks. *Environ. Eng.* **20**, 597–608.

Resende, M. G. C. & Werneck, R. F. 2004 A hybrid heuristic for the *p*-median problem. *J. Heuristics* **10**, 59–88.

Rossman, L. A. 2000 *EPANET 2 User's Manual*. U.S. Environmental Protection Agency, Cincinnati, OH.

Shafiee, M. E. & Zechman, E. M. 2013 An agent-based modeling framework for sociotechnical simulation of water distribution contamination events. *J. Hydro. Inf.* **15** (3), 862–880.

Shastri, Y. & Diwekar, U. 2006 Sensor placement in water networks: a stochastic programing approach. *J. Water Resour. Plan. Manag.* **132** (3), 192–203.

Watson, J.-P., Hart, W. E. & Berry, J. G. 2005 Scalable high performance heuristics for sensor placement in water distribution networks. In: *Proceedings of the World Water and Environment Resources Conference*. ASCE, Reston, VA.

Winter, C., Reddy Palleti, V., Worm, D. & Kooij, R. 2018 Optimal placement of imperfect water quality sensors in water distribution networks. *Comput. Chem. Eng.* **121**, 200–211.

Young, H. P. 1993 An evolutionary model of bargaining. *J. Econ. Theory* **59**, 145–116.

Yousefi, M., Ghoochani, M. & Mahvi, A. H. 2018 Health risk assessment to fluoride in drinking water of rural residents living in the Poldasht city, Northwest of Iran. *Ecotoxicol. Environ. Saf.* **148**, 426–430.

Yousefi, M., Ghalehaskar, S., Baghal Asghari, F., Ghaderpoury, A., Dehghani, M., Ghaderpoori, M. & Mohammadi, A. 2019 Distribution of fluoride contamination in drinking water resources and health risk assessment using geographic information system, northwest Iran. *Regul. Toxicol. Pharmacol.* **107**, 104408.

First received 30 April 2020; accepted in revised form 21 August 2020. Available online 24 September 2020