A critical review of sensor location methods for contamination detection in water distribution networks
Shweta Rathi and Rajesh Gupta

ABSTRACT
Water distribution networks (WDNs) are vulnerable to various types of contamination events that may have impacts on human health and the environment. Therefore, there is a growing need to design an effective monitoring system. Due to the cost of both placing and maintaining the sensors, their numbers must be limited. This constraint makes the sensor deployment locations crucial in water monitoring systems. Several methodologies have been suggested in the past two decades by different researchers for placement of sensors in WDNs. These methodologies differ in many ways depending on the number of objectives, solution methodology, concentration level of contaminant considered, type of simulation, and so on. In this paper, various methodologies have been broadly classified based on the number of performance objectives as single and multi-objective sensor location problems. Some of the features of these methodologies are also mentioned to help understand the advantages of a particular method over other methods. A critical review of literature is presented. Some of the issues on which a consensus is being developed amongst researchers are discussed and recommendations are made with a view to suggest future research needs for sensor network design of large WDNs.

Key words | critical review, sensor location, sensor placement, water distribution networks, water security

ACRONYMS

ACO  Ant colony optimization
CMC  Contaminated mass consumed
CSDL Contamination source detection likelihood
DC  Demand coverage
DL  Detection likelihood
EC  Extent of contamination
GA  Genetic algorithm
HA  Heuristic algorithms
LOS Level of service
MIP  Mixed integer programing
MOGA Multi-objective genetic algorithm
MS  Monitoring station
NFD  Number of failed detection
NSGA Non-dominated sorted genetic algorithm
PE  Population exposed to contamination
PFD Probability of failed detection
PGA  Progressive genetic algorithm
SA  Simulated annealing
SDR  Sensor detection redundancy
SRT  Sensor response time
TD  Time to detection
T-hr LOS  T-hour level of service
VC  Volume of water consumed
WDN  Water distribution network

INTRODUCTION
The main objective of a water distribution network (WDN) is to distribute safe water to consumers. Various standards have been established for maintaining the physical and chemical quality of water. Mostly, these regulations require
that water quality standards must be satisfied at the consumer taps as well as at the source or at the treatment plant. Water quality may deteriorate substantially during transport from the treatment plant to the consumer due to: (1) chemical and biological quality of source water; (2) effectiveness and efficiency of treatment processes; (3) integrity of the treatment plant, storage facilities, and distribution system; (4) age, type, design, and maintenance of the distribution network; (5) quality of treated water; and (6) mixing of water from different sources and other hydraulic conditions (Clark 1995). Therefore, a strategic selection of water quality parameters at various locations in a public water supply network is routinely monitored. Berry et al. (2006b) and Eliades et al. (2011) suggested methodologies for scheduling manual sampling to test the water quality in WDNs. Water samples are collected through these monitoring stations (MSs), and field and laboratory tests are carried out to ensure quality water is being supplied to consumers. However, during recent years, more emphasis is being provided to online monitoring of water quality using sensors. Janke et al. (2006) showed the importance of shorter response time during online monitoring over physical sampling for drinking water security. They observed that online monitoring is a cost-effective technique.

The main objectives for the location of sensors/monitoring points in a WDN are related to: (i) early detection of any contamination events, such as time to detection (TD) defined as the time elapsed between contaminant intrusion and its detection by the first sensor; and (ii) minimizing the impact or consequences of a contamination event, such as volume of water consumed (VC) defined as the amount of contaminated water consumed by the population before detection of contamination event, population exposed to contamination (PE) defined as the number of people exposed to the contamination event before its detection, and extent of contamination (EC) defined as the length of pipe contaminated by a contamination event. Theoretically, contaminants may intrude at any point at any time in a WDN, which requires monitoring at each of the nodes for complete protection of public health, but this is practically not feasible. Therefore, attempts have also been made to cover maximum population with a limited number of sensors considering (i) associated risk (Risk) as an objective for sensor placement and/or (ii) maximizing detection likelihood (DL) of contamination events. Risk can be defined in terms of percentage of population exposed, VC, length of pipe contaminated, or in terms of number of failed detections (NFDs). DL is the probability that a contamination is detected by at least one of the sensors. A sensor may detect a contamination event falsely or may not detect a contamination event if the concentration is below detection limits. Further, there may be a delay in response from the deployed sensors. Therefore, additional objectives such as sensor response time (SRT), NFDs/probability of failed detection (PFD), and sensor detection redundancy (SDR) have also been considered. The SRT is measured by the time elapsed between registration of a contamination event at a sensor and the response provided by it. The PFD is defined as the proportion of undetected events by all sensors. The SDR is defined as the probability of detection of a contamination event by a specified number of sensors within a specified time.

Concerning security in WDNs, perhaps, the major thrust areas for research in the last decade are: impact assessment of contamination events (Murray et al. 2006; Davis & Janke 2008, 2009, 2011); sensor placement for contamination detection (Lee & Deininger 1992; Ostfeld & Salomons 2004; Trachtman 2006); contaminant source identification (Guan et al. 2006; Cristo & Leopardi 2008; Tryby et al. 2010); and flushing of networks after contamination events (Preis & Ostfeld 2008b; Alfonso et al. 2010; Poulin et al. 2010; Shafiee & Zechman 2013).

The main objective of this paper is to review the methodologies related to locations of monitoring points in WDNs for all the above listed objectives. Several methodologies have been developed to tackle the sensor placement problem with single or multiple objectives. Hart & Murray (2010) comprehensively reviewed available literature up to 2008–2009 on sensor placement strategies and identified key issues that need to be addressed in future work. An updated comprehensive review is provided in this paper to include the research work reported thereafter along with critiques. Further, some of the issues for future consideration are presented.

**LITERATURE SURVEY**

Research to optimize location of sensors/MSs is usually classified as single objective or multi-objective sensor
location problems. Performance objectives are those, which provide performance characteristics of monitoring systems. All the works based on any single objective are presented in Table 1, and those considering two or more objectives are compiled in Table 2. General parameters considered for comparison of different works are: (i) need of hydraulic and water quality simulation; (ii) optimization methodology used for solving problems; (iii) network(s) considered for illustration; and (iv) fixed/variable number of sensors.

There are few methods for location of monitoring points that do not require results of hydraulic or water quality simulation. Some are based only on hydraulic simulation results; while some require both hydraulic and water quality simulation. Further, hydraulic and water quality simulation may be for one loading (static) or for multiple loadings in a day (dynamic). Mostly, perfect mixing of streams is assumed at the junctions, where two or more inflow streams meet. However, the mixing may not be perfect and the concentrations in the outgoing pipes may not be the same (Austin et al. 2009; Kim et al. 2010). The data, time, and effort required for solving optimal sensor location problems increase with requirement of simulation and its type.

Mixed integer programing (MIP), evolutionary techniques such as genetic algorithm (GA) (and its variants like non-dominated sorted genetic algorithm (NSGA), multi-objective genetic algorithm (MOGA), and progressive genetic algorithm (PGA)), ant colony optimization (ACO), simulated annealing (SA), and several heuristic algorithms have been suggested to solve sensor location problems. The runtime and computer memory requirement increases in MIP solvers as the number of constraints and variables increases with the increase in the size of networks. Berry et al. (2005b) and Hart et al. (2008a) suggested minimizing the constraints and decision variable to reduce the size of MIP formulation in large WDNs. Evolutionary techniques are simple but computationally intensive, especially when dynamic water quality and hydraulic simulation are required. Heuristic methods are easy, however, they may not provide an optimal number of sensors and their locations for given objectives.

Application of some of the methodologies to large network problems with several pumps, tanks, and loading conditions may be difficult. The example network considered by a researcher gives just an idea about applicability of a method to large network problems with many complexities.

The problem for sensor placement has been formulated by different researchers in two different ways. One type of problem formulation provides locations for a given number of sensors to optimize performance objectives called herein fixed sensor problems; and the other type of problem formulation determines the number of sensors and their locations to meet defined objectives, called herein variable sensor problems. In fixed sensor problems, sensor numbers are limited by the budget. However, in variable sensor problems, importance is given to fulfillment of desired objectives. The fixed sensor problem can be solved repetitively by varying the number of sensors to develop a Pareto front between number of sensors and cost for choosing the most appropriate solution. This methodology has been categorized as variable sensor problems.

In practice, sensor deployment budgets are not fixed in priori, therefore, sensors are incremented over time depending upon cost and availability of resources. Looking to this aspect, Watson & Hart (2010) developed a multi-stage sensor placement model for sensor placement considering fixed cost available for sensors and uncertainty in the cost.

Recent developments in water quality monitoring include the use of mobile sensors (Perelman & Ostfeld 2013b; Rasekh et al. 2014). Sensors are introduced in the pipeline and they move along with flow and pass on the desired information to receptors fitted outside the pipe and connected to computers. Moving sensors along with fixed sensors are observed to increase the efficiency of the sensor system in case of uncertainty of contamination events (Perelman & Ostfeld 2013b).

CRITICAL REVIEW OF SENSOR LOCATION PROBLEMS

Single objective problems

Water quality levels below the drinking water guideline at a monitoring node ensure the delivery of safe water at that node and also at any upstream node, when a major portion of the total quantity received at the monitoring node passed through the upstream node. The demand coverage (DC) of a
MS is defined by Lee & Deininger (1992) as the total demand of all those nodes, which can be assumed to be safe if the quality of water at the monitoring node is below the guideline. Lee & Deininger (1992), Al-Zahrani & Moied (2005), and Afshar & Marino (2012) used MIP, GA, and ACO, respectively, to locate a given number of MSs to maximize the DC based on selected coverage criteria. Woo et al. (2001) emphasized use of water quality simulation to develop a coverage matrix. Kumar et al. (1997) and Kansal et al. (2012) proposed a more systematic way of preparing a

Table 1 | Single objective sensor location methodologies

<table>
<thead>
<tr>
<th>Objective</th>
<th>Citation</th>
<th>Hydraulic/water quality simulation</th>
<th>Methodology used/optimization solver</th>
<th>Network as case study/illustration</th>
<th>Fixed/variable number of sensors</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>Lee &amp; Deininger (1992)</td>
<td>SHS</td>
<td>IP</td>
<td>Hypothetical two loop network; network of Flint, Michigan; Cheshire, Connecticut</td>
<td>Fixed</td>
<td>Regular or routine monitoring</td>
</tr>
<tr>
<td></td>
<td>Kumar et al. (1997)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Hypothetical network with two source nodes</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>Woo et al. (2001)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Hypothetical network with two source nodes</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Zahrani &amp; Moied (2005)</td>
<td>SHS</td>
<td>GA</td>
<td>A hypothetical WDN with three source nodes</td>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghimire &amp; Barkdoll (2006)</td>
<td>Not required</td>
<td>Heuristic</td>
<td>BWSN net. 1 &amp; net. 2</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansal et al. (2002a)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Manendragarh town network, India</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rathi &amp; Gupta (2014)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Network of part of Nagpur city, India</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afshar &amp; Marino (2012)</td>
<td>SHS</td>
<td>ACO</td>
<td>Network of city of Babol, Iran</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>Kumar et al. (1999)</td>
<td>DHS</td>
<td>Heuristic</td>
<td>Any town, USA</td>
<td>Variable</td>
<td>Accidental</td>
</tr>
<tr>
<td>Kansal et al. (2002b)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Network</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chastain (2000)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Network</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rathi &amp; Gupta (2015)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Networks</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cozzolino et al. (2006)</td>
<td>SHS</td>
<td>Heuristic</td>
<td>Network</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC</td>
<td>Kessler et al. (1998)</td>
<td>DHS</td>
<td>Heuristic</td>
<td>EPANET Ex.1 and network</td>
<td>Variable</td>
<td>Accidental</td>
</tr>
<tr>
<td>Ostfeld &amp; Salomons (2004)</td>
<td>DH &amp; WQS</td>
<td>GA</td>
<td>EPANET Ex.1 &amp; network</td>
<td>Fixed</td>
<td>Intentional</td>
<td></td>
</tr>
<tr>
<td>Ostfeld &amp; Salomons (2005a, b)</td>
<td>DH &amp; WQS</td>
<td>GA</td>
<td>EPANET Ex.1 &amp; network</td>
<td>Fixed</td>
<td>Intentional</td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>Berry et al. (2005a)</td>
<td>SHS</td>
<td>MIP</td>
<td>Network, EPANET 2.0 Ex.1, network of 470 nodes &amp; 621 pipes</td>
<td>Fixed</td>
<td>(Intentional) deterministic</td>
</tr>
<tr>
<td>Shastri &amp; Diwekar (2006)</td>
<td>DHS</td>
<td>L-shaped BONUS algorithm</td>
<td>EPANET 1.0 Ex.1 &amp; network</td>
<td>Variable</td>
<td>Demand uncertainty</td>
<td></td>
</tr>
<tr>
<td>Rico-Romizoz et al. (2007)</td>
<td>DHS</td>
<td>Stochastic decomposition</td>
<td>EPANET 2.0 Ex.1 &amp; network</td>
<td>Fixed</td>
<td>Uncertainty in AR and PD</td>
<td></td>
</tr>
<tr>
<td>Uber et al. (2004)</td>
<td>DH &amp; WQS</td>
<td>Heuristics</td>
<td>EPANET 1.0 Ex.1 &amp; network</td>
<td>Fixed</td>
<td>Intentional</td>
<td></td>
</tr>
<tr>
<td>Schwartz et al. (2014)</td>
<td>DH &amp; WQS</td>
<td>GA</td>
<td>EPANET 2.0 Ex.1 &amp; network</td>
<td>Fixed</td>
<td>Intentional</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective</th>
<th>Citation</th>
<th>Hydrologic/water quality simulation</th>
<th>Methodology used/optimization solver</th>
<th>Network as case study/illustration</th>
<th>Fixed/variable number of sensors</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE and MC</td>
<td>Carr et al. (2004)</td>
<td>SHS</td>
<td>Branch and bound</td>
<td>Application not shown</td>
<td>Fixed</td>
<td>Uncertainty in AR and PD</td>
</tr>
<tr>
<td>PE and EC</td>
<td>Propato (2006)</td>
<td>DH &amp; WQS</td>
<td>MIP and heuristic</td>
<td>Networks(^{18}) &amp; case study(^{19})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>TD, VC, PE, EC, and PFD</td>
<td>Berry et al. (2006a)</td>
<td>DH &amp; WQS</td>
<td>Branch and bound</td>
<td>SNL-(1(^{20}), 2(^{21}), 3(^{22}))</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>PE and MC</td>
<td>Berry et al. (2009)</td>
<td>DH &amp; WQS</td>
<td>IP, local search &amp; NLP</td>
<td>Networks(^{20,21,22})</td>
<td>Fixed</td>
<td>Imperfect sensors</td>
</tr>
<tr>
<td>TD, PE, VC, and DL</td>
<td>Dorini et al. (2006)</td>
<td>DH &amp; WQS</td>
<td>Modified cross-entropy algorithm</td>
<td>Networks(^{7,8})</td>
<td>Fixed</td>
<td>Reduced network(^{7})</td>
</tr>
<tr>
<td>TD, PE, VC, and DL</td>
<td>Propato &amp; Piller (2006)</td>
<td>DH &amp; WQS</td>
<td>MIP</td>
<td>Networks(^{7})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>TD, VC, and DL</td>
<td>Wu &amp; Walski (2006)</td>
<td>DH &amp; WQS</td>
<td>GA</td>
<td>Networks(^{7,8})</td>
<td>Fixed</td>
<td>Reduced network(^{7})</td>
</tr>
<tr>
<td>TD, PE, and DL</td>
<td>Huang et al. (2006)</td>
<td>DH &amp; WQS</td>
<td>GA</td>
<td>Networks(^{7,8}) &amp; case study(^{23})</td>
<td>Fixed</td>
<td>Reduced network(^{22})</td>
</tr>
<tr>
<td>DC, TD, VC, PE, and DL</td>
<td>Eliades &amp; Polycarpou (2006)</td>
<td>DH &amp; WQS</td>
<td>Heuristic</td>
<td>Networks(^{7,8})</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>TD, DL, SDR, and CSDL</td>
<td>Preis &amp; Ostfeld (2006)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>Networks(^{7,8})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>TD, DL, and SDR</td>
<td>Preis &amp; Ostfeld (2008a)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>Networks(^{15}) &amp; Richmond water system(^{24})</td>
<td>Fixed</td>
<td>Heuristic method for CES</td>
</tr>
<tr>
<td></td>
<td>Austin et al. (2009)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>Networks(^{15})</td>
<td>Fixed</td>
<td>Imperfect mixing at nodes</td>
</tr>
<tr>
<td>TD and DL</td>
<td>Kim et al. (2010)</td>
<td>DH &amp; WQS</td>
<td>MOGA</td>
<td>Networks(^{7})</td>
<td>Variable</td>
<td>Imperfect mixing at nodes</td>
</tr>
<tr>
<td>TD, VC, and DL</td>
<td>Aral et al. (2010)</td>
<td>DH &amp; WQS</td>
<td>PGA</td>
<td>Networks(^{7,8})</td>
<td>Fixed</td>
<td>Reduced network(^{6,7})</td>
</tr>
<tr>
<td>TD and PE</td>
<td>Krause et al. (2008)</td>
<td>SH &amp; WQS</td>
<td>Greedy and SA</td>
<td>Networks(^{7,8}), large network (21,000 nodes)(^{25})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comboul &amp; Ghanem (2013)</td>
<td>DH &amp; WQS</td>
<td>Greedy algorithm</td>
<td>Networks(^{26})</td>
<td>Fixed</td>
<td>Nodal demand uncertainty</td>
</tr>
<tr>
<td>VC and NFD</td>
<td>Guidorzi et al. (2009)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>City of Ferrara(^{27})</td>
<td>Fixed</td>
<td>Valves &amp; hydrants operation to reduce VC</td>
</tr>
<tr>
<td>Risk in terms of VC and NFD</td>
<td>Weickgenannt et al. (2010)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>Almelo(^{28}) (The Netherlands)</td>
<td>Variable</td>
<td>IBSM developed to CES</td>
</tr>
<tr>
<td>DL and PE</td>
<td>Dorini et al. (2010)</td>
<td>SH &amp; WQS</td>
<td>Heuristic method</td>
<td>Networks(^{7,8})</td>
<td>Fixed</td>
<td>Heuristics SLOT algorithm</td>
</tr>
</tbody>
</table>

(continued)
coverage matrix and suggested a heuristic method to locate MSs one by one, by selecting the best location first and modifying the coverage matrix by eliminating nodes already covered to select the next station. Ghimire & Barkdoll (2006) and Rathi & Gupta (2014) also suggested heuristic methods to simplify the problem. Methods based on DC give importance to coverage and therefore try to locate MSs as far away as possible from source based on coverage criteria. They are good for the location of MSs required for regular monitoring against accidental contamination.

Early detection of a contamination event is desirable. Kumar et al. (1999) and Kansal et al. (2012b) used TD, defined as the time elapsed between the entry of contaminant and its detection by any of the MSs and used it as a measure for level of service (LOS) to consumers. They identified best monitoring locations one by one by constructing and using a travel-time matrix for desired LOS. Chastain (2006) developed a heuristic methodology considering extended period water quality analysis for creating a database of water system responses by injecting contaminant at each node. The method searched for the best locations of MSs one by one to maximize the covered nodes with the condition of time to detect. Rathi & Gupta (2013) also suggested a heuristic method that works on an appended shortest travel time tree to identify best locations of MSs to achieve desired T-hr LOS, defined as LOS in which a contamination event at any node is detected by at least one of the sensors within T-hours of its intrusion. The number of MSs increases as the desired LOS increases. With constraints on number of sensors, the desired LOS may not be achieved. Most of the multi-objective-based sensor location methodologies consider TD as one of the objectives.

Kessler et al. (1998) suggested using total VC before detection of any event to quantify the impact of contamination event. The LOS is decided by a pre-specified value of VC. They developed a pollution matrix for a given LOS and identified an optimal set of MSs, which covered all the contamination events. Ostfeld & Salomons (2004) considered random multiple contamination events to decide locations of MSs.

A random pollution matrix was generated by considering LOS in terms of VC, and GA was used to identify the location of MSs. Ostfeld & Salomons (2005a, b) extended the methodology to consider the randomness of the flow rate of the injected pollutant, randomness in consumers’ demands, and the detection sensitivity and response time of MSs. Schwartz et al. (2014) used GA to minimize the population exposed using the concept of the randomized pollution matrix of Ostfeld & Salomons (2004) for selection of sensor locations.

Berry et al. (2005a) used a simple model in which impact of contamination is considered by the population exposed to contaminated water and identified optimal sensors locations using MIP. Berry et al. (2006a) and Propato (2006)
formulated MIP models in such a way that a wide range of design objectives can be accommodated in the formulation, either individually or jointly. Berry et al. (2006a) quantified the impact of each contamination event by multiplying (1) the probability of events, (2) a binary (1, 0) contamination indicator, and (3) impact value evaluated from dynamic water quality simulation. Propato (2006) considered the minimization of impacts associated with contamination scenarios in terms of TD, PE, VC, contaminated mass consumed, and probability/percentage of failed detection (PFD). Since projected nodal demand has inherent uncertainty, several researchers incorporated it in deciding sensor locations with the objective of minimizing the expected population exposed (Shastri & Diwekar 2006; Rico-Ramirez et al. 2007) or detection time (Cozzolino et al. 2006).

Each objective has its own advantage and provides a different set of locations. Watson et al. (2004) demonstrated that sensor placement using one objective provides greater risk. They observed that the majority of objectives are uncorrelated, and an optimal solution associated with one objective function could be highly sub-optimal with respect to another design objective. Murray et al. (2008) compared sensor placement solutions considering three objectives (PE, EC, and DL) using eight example networks of varying size for illustrations and showed that the number of sensors needed for various objectives depends upon the marginal benefit achieved or acceptable risk defined by water utilities. Bahadur et al. (2010) assessed the impact of both spatial and temporal population variability on sensor network design and observed it to be significant. Pinzinger et al. (2011) compared approaches based on integer programing (IP) and two greedy heuristics for identifying the best locations of MSs for large networks. They maximized the coverage of network nodes within the allowable time limit considering for fixed numbers and observed that the method based on IP can be used for up to 5,000 nodes efficiently and for more than 5,000 nodes greedy-based approaches performed well. Hart et al. (2011) showed that sensor placement using optimization-based methods depends upon how precisely the objective function is modeled. They showed that even though minimization of impact is equivalent to maximization of complement of impact reduction, the choice between the two depends on the algorithm, quantity of impact when no sensors are available, and the number of sensors available. Since limited sensors are to be used, this makes sensor location crucial. Therefore, in order to provide a balanced solution using a limited number of sensors, researchers suggested different algorithms based on multiple objectives.

Multi-objective problems

Two types of multi-objective approaches have been used by various researchers. Some researchers considered the approaches in which the objective functions remain mutually distinct, and the result is expressed in the form of a Pareto front (Dorini et al. 2006; Eliades & Polycarpou 2006; Huang et al. 2006; Preis & Ostfeld 2006, 2008a; Propato 2006; Austin et al. 2009; Guidorzi et al. 2009; Ehsani & Afshar 2010; Weickgenannt et al. 2010; Shen & McBean 2011) and other approaches in which the different objectives considered are grouped together in a single objective function, which is then solved using an optimization solver (Propato & Piller 2006; Wu & Walski 2006; Krause et al. 2008; Aral et al. 2010; Dorini et al. 2010; Ehsani & Afshar 2010; Xu et al. 2010a). The available multi-objective methodologies are given in Table 2 in the chronological order of their development.

It can be observed from Table 2 that few additional objectives such as contamination source detection likelihood (CSDL) as used by Preis & Ostfeld (2006), Risk as used by Weickgenannt et al. (2010), DL and response delay of sensors as used by Ostfeld & Salomons (2005a, b), SDR as used by Preis & Ostfeld (2006), etc. are suggested as performance indicators of monitoring systems. Bristow & Brumbelow (2006) modeled the response time in various phases and defined it as the time between initial detection of a contamination event and the time an individual user stops using contaminated water.

As mentioned earlier, TD is the most preferred objective in multi-objective problems as it forces early detection of contamination events. It is coupled with one or more complementary objectives that quantify impact of contamination events with one or more competing objectives such as DL or coverage. Ostfeld et al. (2008) compared solutions provided by several algorithms based on four objectives: (1) TD; (2) PE; (3) VC; and (4) DL. The solutions provided by different algorithms were quite varied. Several researchers
had ignored the impact of undetected events in calculation of objectives (Pries & Ostfeld 2006; Ostfeld et al. 2008).

Grayman et al. (2006) considered VC, PE, EC, and percentage node impacted (defined as nodes where concentration of contaminant is observed above a threshold value) to check the effectiveness of a given system of MSs under a given set of contamination events. McKenna et al. (2006) observed the relationship between relative detection limit and the number of sensors for different performance measures (TD, PE, and EC). Their results demonstrated that detection of contamination events is improved by improving the detection limit of sensors. Isovitsch & VanBriesen (2008) observed that placement order is the same for multiple sensor locations with different intrusion scenarios and optimization criteria. They observed that sensor locations selected by minimizing the VC or maximizing the population affected are likely to coincide with network nodes with a high reachable average demand. Alternatively, sensor locations selected by maximization of DL are likely to coincide with network nodes with a low reachable average demand. Hart & Murray (2010) reviewed various sensor placement strategies for contamination warning systems adopted by different researchers and observed that many objectives have been identified in the literature, but a realistic multi-objective analysis can handle only a few.

GA is the most preferred solver for multi-objective optimization problems. The computational requirement in GA increases with the size of network and number of contamination scenarios required to be considered, which poses restrictions on its application to large network problems. Heuristic algorithms have their own limitations. Another problem observed especially in developing countries is that no well-calibrated model for the network is available, which is considered to be the most important requirement of most methodologies. Therefore, researchers have recently developed methodologies for real-world problems to tackle the complexity of the network and reduce the computer runtime using some graph theoretic and heuristic approaches (Xu et al. 2008; Chung et al. 2010; Chang et al. 2011; Chang et al. 2012; Klise et al. 2013; Perelman & Ostfeld 2013a). Hart et al. (2008a) and Berry et al. (2008a) introduced limited memory techniques by developing exact and heuristics algorithms for sensor placement for application to large networks with the aim of minimizing computer memory and computational speed while maintaining the accuracy compared to other methods available in literature. Depending upon the purpose, whether minimizing computer memory and/or computational speed, the choice of optimization heuristics is selected.

Berger-Wolf et al. (2005) considered the problem of sensor placement with the objective of identification of the source of contamination. Austin et al. (2009) considered imperfect mixing conditions at pipe junctions for optimizing sensor placement. Xu et al. (2010b) considered sensor placement along with the probability of contamination source detection.

It can be observed from the reported literature that some consensus is being developed amongst researchers on issues related to sensor location problems, for example: they are considered as multi-objective problem with early detection as one of the objectives; skeletonization of a large size network to reduce the number of candidate nodes; restriction on number of contamination scenarios; type of simulation; and type of solution methodology. Suggestions by different researchers on the above issues are discussed in the next section when considering the applicability of algorithms developed in the future to large WDNs.

SUGGESTIONS BY DIFFERENT RESEARCHERS ON SOME COMMON ISSUES AND RECOMMENDATION

Original or reduced network

A large WDN may involve thousands of pipes and nodes, and the number of sensors are restricted. Considering each and every node of the original network as a possible sensor location may unnecessarily increase computational burden, which can be significantly reduced by suitably eliminating some of the nodes from the list of candidate nodes such that sensor placement accuracy is not affected. Simple methods such as considering only large diameter pipes for sensor locations (Dorini et al. 2006; Huang et al. 2006) or identifying higher demand nodes (Aral et al. 2010) reduce the search process, however, the entire network is required to be considered during analysis. Wu & Walski (2006) suggested reducing networks by considering
hydraulic equivalence. However, Perelman et al. (2008) suggested a model to reduce a network by considering both hydraulic and water quality equivalence and showed its superiority over other models that consider skeletonization using hydraulic equivalence. Skeletonization of networks considering hydraulic and water quality is difficult for multi-source networks in which flow direction in some of the links changes with time due to changes in operating conditions such as pump on–off. Perelman & Ostfeld (2022a) considered changes in flow direction with time and suggested a methodology for clustering network nodes for various purposes. They suggested considering the location of sensors cluster-wise and deciding on their location within a cluster through optimization.

The basic purpose of reducing the size of a network is to reduce the number of candidate nodes for location of sensors without compromising on the hydraulic and water quality equivalences between the complete and reduced network. Therefore, clustering of nodes and placing sensors cluster-wise can be considered as the best solution for network reduction.

**Type of analysis**

Both nodal demands and the water levels in the reservoir change throughout the day, therefore flow scenarios within a network practically change continuously. This requires consideration of dynamic analysis at a suitable interval. From the reported literature, it was observed that different types of loading are used by different researchers. Some used static whereas some used dynamic, as described in Tables 1 and 2. Berry et al. (2005c) compared the static and dynamic sensor placement formulations and observed that static simulations allow a relatively fast determination of the optimal solution and hence are applied to large-scale problems whereas dynamic simulations provide more realistic scenarios and hence are more accurate in determination of contaminant assessment. Consideration of each and every scenario results in a high computational cost in terms of computer memory and computational speed; hence, the computational requirement can be substantially reduced by identifying a few important scenarios with respect to nodal demands, pump on and off situations, and flow in and out conditions from tanks.

**Number of contamination events and their locations**

A contaminant may enter into a network from any point and at any time. Further, there could be contaminant intrusions at multiple locations. The severity of damage due to contaminant intrusion depends on several factors such as location of node, direction of flow, type, mass rate and injection duration of contaminant, time (day/night), etc. Instead of considering each and every node as a possible location of entry of contaminant, a few vulnerable locations can be selected to reduce the computational work. Ostfeld & Salomons (2004) considered two randomly selected nodes and times for contamination injection. Some researchers generated random contamination events using a Monte Carlo simulation technique for optimization (Elliades & Polycarpou 2006; Wu & Walski 2006). Preis & Ostfeld (2008a) selected a few representative contamination events whose geographic coordinates are similar to that of the entire set of possible contamination events by developing an equation of a reduced size contamination matrix and provided similar results when whole/full contamination events are simulated for optimization. Berry et al. (2006a, 2009) considered contamination events only at non-zero demand nodes to reduce the number of contamination events. Krause & Guestrin (2009) developed an algorithm for solving large robust sensor placement problems and provided better results under both worst- and average-case scenarios. Krause et al. (2008) and Krause & Guestrin (2009) selected those contamination scenarios or events that affect at least 10% of the WDN nodes for simulation for large networks (the Battle of the Water Sensor Network (BWSN) network 2). Watson et al. (2009), Carr et al. (2006), Xu et al. (2004), and Davis et al. (2013) concentrated on high consequence contamination events. Shen & McBean (2011) selected the contamination events that occur at peak demand hours and events that occur at dead ends are ignored for optimization. Perelman & Ostfeld (2010, 2012b) suggested a crossed entropy approach to select critical contamination events that have low probability of occurrence but have an extreme impact. Weickgenannt et al. (2010) developed heuristic methods to choose a working set of contamination events for optimizing sensor placement by defining the importance of each scenario in terms of contaminated water volume consumed and applied more weight to those scenarios, which generate more polluted water. Their method showed superiority over other methods,
which considered a Monte Carlo simulation technique, by providing more importance to more important scenarios. The consideration of only high consequence events, which can be identified based on importance of locality or infrastructures, can be considered to reduce the number of contamination events.

**Number of performance objectives**

One objective is certainly not enough and multiple objectives are necessary to obtain a balanced solution. Selection of too many objectives at a time increases the computational requirement. Selection of objectives should be such that a balanced design could be obtained with respect to different objectives. This requires selection of competitive type of objectives, and some complementary type of objectives can be dropped.

**Type of solution methodologies and software**

The MIP and evolutionary techniques have been mostly suggested for multi-objective sensor location problems. While MIP requires handling of a large number of variables and constraints at a time, evolutionary techniques require more computational time. The heuristic methods along with optimization methods are independently used to reduce the size of the problem and obtain near-optimal solutions. The preferred solution methodology is one that can prioritize a selection with respect to different objectives considered in the sensor location problems and can be easily applied to large WDNs. Prioritization selection helps in future extension of monitoring locations. Software like TEVA-SPOT (Berry et al. 2008b, 2010; Hart et al. 2008b; Murray et al. 2010) and S-PLACE tool kit (Eliades et al. 2014) are required to handle large size networks, select flow distributions, restrict contamination events, and limit objectives.

**SUMMARY AND CONCLUSIONS**

The issue of water system security in recent years has motivated several researchers to develop a methodology for sensor location to prevent the public from negative impacts due to possible contaminations. Several performance objectives for evaluating monitoring systems have been considered. In spite of much research, there is no consensus amongst researchers on several issues related to sensor location problems. A critical review of available methodologies is presented in this paper, and some of the common issues on which consensus is being developed amongst researchers are discussed. The research work pertaining to these issues is highlighted, and some recommendations are made for researchers for future work on sensor location problems.

**REFERENCES**


In: Proc. 8th Annual Water Distribution Systems Analysis Symp. ASCE, Reston, VA.


