A review of sensor placement objective metrics for contamination detection in water distribution networks

Shweta Rathi, Rajesh Gupta and Lindell Ormsbee

ABSTRACT

Security of water distribution networks (WDNs) through early warning systems is one of the topics of research over the last two decades to safeguard human health and the environment against accidental and intentional contamination. Several methodologies have been suggested by different researchers for placement of sensors in WDNs which can provide early warning. Owing to the cost of both placing and maintaining the sensors, the number of these must be limited. This constraint makes the sensor deployment locations crucial in early warning systems. In this paper various methodologies suggested for sensor placements have been broadly classified based on the number of performance objectives as single and multi-objective sensor location problems. Various objectives and their significance in the context of sensor placement strategy have been discussed. A complete review of literature is presented to understand complexities in the sensor design problems and to determine the directions in which research is heading. Several challenges for the design of robust sensor network for large WDNs with less computational efforts are discussed and recommendations are presented.

Key words | early warning systems, monitoring system, sensor location, sensor placement, water distribution networks, water security

LIST OF ACRONYMS

ACO | ant colony optimization
AR | attack risk
BWSN | the battle of the water sensor network
CES | contamination event sampling
CSDL | contamination source detection likelihood
DC | demand coverage
DHS | dynamic hydraulic simulation
DH & WQS | dynamic hydraulic and water quality simulation
DL | detection likelihood
EC | extent of contamination
F | penalty based VC
GA | genetic algorithm
HA | heuristic algorithms
IBSM | importance based sampling method
IP | integer programming
LOS | level of service
MC | mass consumed
MIP | mixed integer programming
MOGA | multi-objective genetic algorithm
MS | monitoring station
NFD | number of failed detections
NSGA | non-dominated sorted genetic algorithm
PD | population density
PE | population exposed to contamination
PFD | probability of failed detection
PGA | progressive genetic algorithm
PR | expected penalty reduction
R | risk of contamination
RDS | response delay of sensor
SA | simulated annealing
SDR | sensor detection redundancy
SHS | static hydraulic simulation
SH & WQS | static hydraulic and water quality simulation

Shweta Rathi
Rajesh Gupta (corresponding author)
Civil Engineering Department,
Visvesvaraya National Institute of Technology,
Nagpur,
Maharashtra 440 010,
India
E-mail: drrajeshgupta123@hotmail.com

Lindell Ormsbee
Kentucky Water Resources Research Institute,
University of Kentucky,
233 Mining & Mineral Resources Bldg,
Lexington,
KY 40506-0107,
USA

doi: 10.2166/ws.2015.077

© IWA Publishing 2015
**INTRODUCTION**

Concerned with water security in distribution networks, the major thrust areas for research in the last decade include: impact assessment of contamination events (e.g. Murray et al. 2006; Davis & Janke 2008, 2009, 2011); sensor placement for contamination detection (Bahadur et al. 2003; Ostfeld & Salomons 2004; Trachtman 2006; Hart & Murray 2010); contamination source identification (e.g. Guan et al. 2006; Cristo & Leopardi 2008; Tryby et al. 2010); and flushing of network after contamination event (e.g. Preis & Ostfeld 2008b; Alfonso et al. 2010; Poulin et al. 2010; Shafiee & Zechman 2013). The problem of sensor placement in water distribution networks (WDNs) is mostly considered to minimize the impact of contamination events. However, contamination source identification, or response to a contamination event has recently been considered while placing sensors in the network (Guidorzi et al. 2009; Xu et al. 2008a). The main objective of this paper is to review the methodologies related to the location of sensors in WDNs. This paper thus builds on an earlier review by Hart & Murray (2010). Since their review, over 40 additional papers have been published on the topic. Hart & Murray (2010) classified the reviewed literature into nine different categories by jointly considering risk calculations with simulation, imperfect sensors, multiple objectives, optimization objectives and data uncertainties. In the current review, we examine the methods by considering the scope of the problem being considered (i.e. single vs. multi-objective). The review is further classified based on the type of operational objectives being considered (e.g. time to detect, number of customers impacted, etc.). The different methods are further classified based on: (i) type of simulation employed; (ii) methodology used for solving the optimization problem; (iii) network(s) considered for illustration; and (iv) fixed/variable number of sensors. Some challenges for the water supply community and recommendations of various researchers are provided. This paper is an extension of the work (Rathi & Gupta 2014c) presented in the 16th International Symposium on Water Distribution System Analysis (WDSA 2014) held at Bari, Italy in 2014.

Sensor placement problems can be broadly classified into problems involving a single objective or multiple objectives. In a single objective problem, sensor placement is formulated in two different ways: (1) minimization of the sensors (number or cost) subject to satisfying the design criteria of a selected performance objective; and (2) maximization of the performance objective subject to a known number of sensors. A trade-off can be seen between the number of sensors and the objective function value for solving a second type of formulation with a different number of sensors.

More than one performance objective is considered at a time in multi-objective sensor location problems. The multi-objective problems are tackled in two ways: (1) different objectives are grouped together in a single objective function; or (2) objective functions remain mutually distinct and results are expressed in the form of a Pareto front. The multi-objective optimization problem provides a compromised solution by providing a maximum/minimum for possible values of different objectives under the given conditions.

Both integer programming (IP) and mixed integer programming (MIP) have been found to be well suited for the facility location problem and has been more commonly used by several researchers for sensor placement in WDNs. Further, several other methodologies, like the genetic algorithm (GA) (and variants like the non-dominated sorting genetic algorithm (NSGA) and progressive genetic algorithm (PGA)), ant colony optimization (ACO), and simulated annealing (SA), have also been suggested. A MIP solver requires a lot of computer memory because as the size of network increases, their application to large WDNs is difficult (Berry et al. 2005b; Hart et al. 2008a). Evolutionary techniques are simple but computationally intensive, especially when dynamic hydraulic simulations are required. Several heuristic methods have been suggested which require less computational effort, however they may not provide an optimal solution.
VARIOUS OBJECTIVES AND THEIR SIGNIFICANCE

Several objectives have been proposed to decide location of sensors in a WDN. Including all of them in deciding sensor locations is a challenging task (Ostfeld et al. 2008). This section covers definitions of the various objectives and their significance in placement of sensors in the network.

Demand coverage

The ‘Demand Coverage’ (DC) objective is defined as the percentage of network demand monitored by a particular sensor or set of sensors located at different nodes in a WDN (Lee & Deininger 1992). If the water quality is good at a sensor node, then it can be presumed good at all upstream nodes through which a major portion of water received at the sensor node has passed. Thus, upstream nodes are assumed to be covered, if the desired fraction of flow received at the sensor node has passed through it. It can be mathematically expressed as:

$$DC = \frac{\sum_{p=1}^{P} \sum_{j=1}^{J} a_{jp} \times q_{jp}}{\sum_{p=1}^{P} \sum_{j=1}^{J} q_{jp}}$$  \hspace{1cm} (1)

where $a_{jp} = 1$, if node $j$ is covered by any of the sensor nodes during pattern $p$, else $a_{jp} = 0$. $q$ = nodal demand, $J$ = number of covered nodes, and $P$ = number of flow patterns. The sensor network design based on DC selects the sensor locations as far as possible from the source to include more nodes to maximize DC. Thus, a contamination event at a node may require more time for its detection than desired and may result in more negative consequences.

Time of detection

The detection time for a particular contamination scenario is the time elapsed between the start of a contamination event and its detection at the first sensor location. Thus, the detection time for any event $x$ at node $i$ which is first detected at the sensor $S_p$ placed at node $j$ would be the minimum travel time by a contaminant to reach node $j$ from node $i$ during flow pattern $p$, which is represented as $t_{xp}$. This would be different for different contamination scenarios. The ‘Time of Detection’ (TD) for the sensor network is considered in two different ways: (i) the maximum detection time of different contamination scenarios; and (ii) the average detection time of all contamination scenarios (Kumar et al. 1999).

The first definition is more useful when all contamination scenarios are detected. The second definition is used when detection of all events is not possible due to a limited number of sensors. Some researchers considered only detected events, while others included undetected events by considering detection time of undetected events based on time of simulation (say $\tau$) or when contamination is indirectly detected in public (Watson et al. 2004). In the simplest form, the TD for a sensor network is expressed as:

$$TD = \sum_{x \in J} \sum_{p=1}^{P} \alpha_{xp} DT(x, p)$$  \hspace{1cm} (2)

where $DT(x, p) = t_{xp}$ for detected events and $\tau$ for undetected events; $\alpha_{xp} =$ probability of event $x$ during pattern $p$ and it is calculated as per the expert’s knowledge of network vulnerability; and $x \in$ nodes 1,$\ldots$, $J$, one event at each node.

Optimizing TD by ignoring undetected events will minimize the TD for the detected events and will try to locate sensors near to the source. When both detected and undetected events are included in minimizing TD, the first priority would be to convert undetected events to detected events as TD decreases faster. Thus, sensors would be located away from the source to at least detect all events in whatever minimum time frame. The TD is an important parameter of a sensor network (early warning system). However, it does not explicitly quantify the impact of contamination events.

To understand the impact (negative consequences) of any contamination event, it is necessary to understand the event first (i.e. duration of event, strength of contaminant and its movement in water distribution network after intrusion). Contaminants which pollute the water during the period are introduced. A part of the contaminated water may be consumed and this would depend on how early an event is detected. The contaminant moves and mixes at junctions and finally comes out from the node. Thus, certain stretches of pipeline get polluted and consumers located...
on this stretch will be exposed to the contamination events. The next three objectives involve quantifying the impact of contamination.

Population exposed

The number of people exposed to a contamination event before detection is termed ‘Population Exposed’ (PE) (Berry et al. 2005a). When hydraulic simulation results are used, the PE during event \( x \) would be the addition of the population served at all nodes which get contaminated in time \( t_{xp} \). Therefore, the PE is given by:

\[
PE = \sum_{x=1}^{j} \sum_{p=1}^{\alpha_{xp}} \delta_{jp} \tag{3}
\]

where \( \delta_{jp} \) = population associated with contaminated node \( j \) during pattern \( p \). \( \sum_{x=1}^{j} \sum_{p=1}^{\alpha_{xp}} \alpha_{xp} = 1 \). However, when water quality simulation results are used, PE is mathematically expressed as:

\[
PE = \sum_{x=1}^{j} \sum_{p=1}^{\alpha_{xp}} \sum_{j \in \text{contaminated nodes}} C_{xjp} \delta_{jp} \tag{4}
\]

where \( C_{xjp} \) = contamination concentration indicator variable. \( C_{xjp} = 1 \), if the concentration is more than threshold concentration, 0 otherwise.

Extent of contamination

The total length of the contaminated pipeline is defined as the ‘Extent of Contamination’ (EC) (Watson et al. 2004). The length of pipe contaminated during contamination event \( x \) would be the addition of the length of all pipes contaminated in the period \( t_{xp} \) under the flow pattern \( p \). Mathematically, EC for the sensor network can be stated as:

\[
EC = \sum_{x=1}^{j} \sum_{p=1}^{\alpha_{xp}} \lambda_{jp} \tag{5}
\]

\[
\lambda_{jp} = \sum_{e \in \text{pipes between contaminated nodes}} L_e \tag{6}
\]

where \( L_e \) = length of pipe \( e \). An expression, similar to Equation (4), can be written for a water quality simulation.

Volume consumed

The amount of contaminated water consumed by the population before detection of an event by a sensor is defined as ‘Volume Consumed’ (VC) (Kessler et al. 1998; Watson et al. 2004). Mathematically it can be expressed as:

\[
VC = \sum_{x=1}^{j} \sum_{p=1}^{\alpha_{xp}} \sum_{j \in \text{contaminated nodes}} q_{jp}(DT(x, p) - t_{ijp}) \tag{7}
\]

where \( (DT(x, p) - t_{ijp}) \) is the consumption time at node \( j \). Consumption time is the duration when the contaminated node \( j \) consumes contaminated water before detection at a sensor node. Thus, \( DT(x, p) \) is the total detection time at sensor node, say \( i \), and \( t_{ijp} \) = travel time between nodes \( j \) and \( i \) for pattern \( p \).

The amount of contaminant consumed is measured in terms of total mass and termed as mass consumed (MC) prior to detection of an event.

It can be noted that PE, EC, and VC are related to TD. However, individual minimization of TD, PE, EC, and VC results in different locations of sensors (Watson et al. 2004). Minimization of PE or VC selects sensor locations at or near higher population (or higher demand) nodes, while EC minimization results in sensor locations to restrict the spread of contamination. The above quantification of impacts in the form of PE, EC, and VC assumes that there is quick action by a water authority after detection of a contaminant to locate the source of detection and to flush the contaminated water. The undetected events may have high negative consequences as the contaminated water will travel in the network until withdrawn at consumer locations.

Detection likelihood

The proportion of attacks that is detected by a sensor network, or the probability of detection of a contamination event by a sensor network, is defined as ‘Detection Likelihood’ (DL) (Ostfeld & Salomons 2008). Mathematically, it
can be expressed as:

\[ DL = \sum_{x=1}^{I} \sum_{p=1}^{P} \alpha_{xp} b_{xp} \]  

(8)

where \( b_{xp} = 1 \), if contamination event \( x \) during pattern \( p \) is detected; else 0. The complement of DL is defined as ‘Number of failed detections’ (NFD) or the ‘Probability of failed detection’ (PFD) and is expressed as:

\[ NFD = 1 - DL \]  

(9)

The placement of sensors to maximize DL results in the location of sensors away from the source to reduce the number of undetected events. Kansal et al. (2012b) defined DL considering only those events in the list of detected events which have DT less than some acceptable time. Thus, any event detected beyond a specific time interval is not included while computing DL. Thus, DL can be termed as time-constrained DL. Obviously, maximization of time-constrained DL would restrict sensor locations to move away from the source. There are other performance objectives which combine the impact of two or more individual objectives as discussed in the following sections.

**Risk of contamination**

The impact of not detecting the contamination intrusion and the corresponding consequences were combined in a single objective by Weickgenannt et al. (2010). They defined the risk of contamination as the product of the probability of not detecting the contaminant intrusion and the corresponding consequences expressed as the average volume of water consumed prior to the network shutdown. It is expressed as:

\[ R = \left[ 1 - \frac{1}{J} \sum_{x=1}^{J} \xi(S_p, x) \right] \left[ \frac{1}{J} \sum_{x=1}^{J} \chi(S_p, x) \right] \]  

(10)

where \( R \) = associated contamination risk; \( \xi(S_p, x) \) = a binary function equal to 1 if the event is detected by any sensor node \( S_p \), else 0; \( \chi(S_p, x) \) = volume of contaminated water consumed prior to network shutdown following the intrusion detected at a specific sensor. It can be noted that the first summation in Equation (10) needs to be maximized and the second summation needs to be minimized for the minimum value of \( R \). Moreover, when \( R \) is minimized for a given number of sensors; it results in the maximum possible value of DL and the minimum possible value of VC. The basic assumption in evolving Equation (10) is that the network will be shutdown in a specific period following the detection of a contamination event.

**Unified normalized objective**

Various sensor location algorithms for WDNs were evaluated based on four objective: TD, PE, VC, and DL, for the Battle of the Water Sensor Networks Challenge (BWSN), held as part of the WDSA symposium, in Cincinnati in 2008. These four objectives were combined into a single objective to design the sensor network for WDNs (Wu & Walski 2006). A unified objective is formulated by normalizing and combining the four objectives as shown below:

\[ Z = \frac{TD}{H_{\text{time}}} + \frac{PE}{P_{\text{total}}} + \frac{VC}{V_{\text{max}}} + \frac{1}{DL} \]  

(11)

where \( H_{\text{time}} \) is the time of an undetected event elapsed from the contamination attack to the end of the hydraulic and water quality simulation, \( P_{\text{total}} \) is the equivalent total population; and \( V_{\text{max}} \) is the maximum contaminated demand of water volume over the simulation period. The minimization of \( Z \) will equivalently minimize TD, PE, and VC and maximize DL.

**Expected penalty reduction**

The penalty reduced by replacing one sensor location with another is termed as expected penalty reduction. For a contamination event \( x \) that remains undetected, let \( \pi(x) \) be the impact quantified through a certain parameter. Let the event be detected at certain sensor locations \( S_p \), and the penalty reduction would be \( P_R(S_p, x) = \pi(x) - \pi(S_p) \). The expected penalty reduction over possible scenarios would be

\[ P_R = \sum_{x=1}^{J} \alpha_x P_R(S, x) \]  

(12)

Maximizing \( P_R \) would minimize the expected penalty. It can be noted that the above objective would reduce the
number of undetected events and also the TD. Multiple objectives of a similar nature can be considered by maximizing the normalized weighted mean as \( \text{Max } \sum \lambda_i P_E(s, j) \) (Krause et al. 2008).

**Penalty based VC**

Aral et al. (2010) combined the effect of four performance objectives, TD, PE, VC, and DL in a single-objective formulation, \( F_i \) which needs to be minimized for optimal location of sensors in the network. Aral et al.’s formulation is expressed as:

\[
F = \left(1 - DL\right) \times E_x \left[ \sum_{j=1}^{J} \sum_{t=t_{\text{in}}^j}^{t_{\text{out}}^j} (t - t_{\text{in}}^j + 1)V_{ji}(t) \right]
\]  

(13)

where \( x \) = index of contamination event, \( E_x[\bullet] \) = expected volume of total water demand that exceeds a predefined threshold with a penalty based on time taken for detection, \( t_{\text{in}}^j \) = time when contamination intrusion begins; \( t_{\text{out}}^j \) = time when contamination is detected; \( V_{ji}(t) \) = volume of contaminated water consumed at node \( j \) during event \( x \) in time step \( t \) and calculated as \( q_j(t) \times \Delta t \). The minimization of \( F \) for the placement of a fixed number of sensors results in the minimum possible value of DL, the minimum possible value of VC and PE (considered as a function of VC), and the minimum value of TD. The key difference between Equations (10) and (13) is the penalty coefficient \( (t - t_{\text{in}}^j + 1) \) which increases as \( t \) increases. With a reduction in TD (i.e. \( t_{\text{in}}^j \) in Equation (13)), the number of time steps reduces and also the penalized volume \( V_j(t) \) reduces.

**Sensor detection redundancy**

A contamination event may be detected at more than one sensor location. ‘Sensor Detection Redundancy’ (SDR) is defined as the probability of detection of a contamination event by a specified number of sensors in a specified time (Preis & Ostfeld 2006). The maximizing of SDR will place the sensors in close proximity to one another. SDR alone is not maximized but considered as one of the objectives in multi-objective problems.

A couple more objectives included in multi-objective problems are contaminant source detection likelihood (CSDL) and response delay of sensor. The CSDL is the likelihood of detection of a contamination source and response delay by a sensor and is measured by time elapsed between registration of a contamination event at the sensor and the response provided by it.

**RESEARCH PERTAINING TO VARIOUS OBJECTIVES**

The problem of identifying optimal monitoring station (MS) locations for regular water quality monitoring in WDNs was considered in the initial research. Thus, the simple objective was to cover the maximum demand with a minimum number of MSs. In the last two decades, the vulnerability of a WDN to accidental and intentional contamination has been emphasized and several objectives have been developed, as discussed in previous sections. Single objective sensor location problems are straightforward in which a fixed number of sensors are to be located to maximize or minimize the selected objective. However, in multi-objective sensor location problems the question starts with selecting a group of objectives. Some of the objectives are complimentary to each other; an improvement in one improves the other. For example, a reduction in VC reduces PE (Wu & Walski 2006); or a reduction in TD reduces PE, VC, and EC. However, these objectives are observed to compete with each other for location of sensors. Watson et al. (2004) observed that an optimal solution associated with one objective function could be highly sub-optimal with respect to other design objectives. Hart et al. (2011) showed sensor placement using optimization-based methods depends upon how precisely the objective function is modelled. They showed that even though the minimization of impact is equivalent to the maximization of the 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, researchers suggested different algorithms based on multiple objectives in order to provide a balanced solution using a limited number of sensors.

Design objectives for sensor locations can be classified into four groups based on the following criteria: (i) likelihood of detection (including sensor and source location); (ii) TD;
(iii) impact metrics (e.g. volume/MC, PE, spatial EC); and (iv) operational metrics (e.g. SDR, sensor response delay, etc.)

For simplicity, we will consider formulations which involve a single objective first, followed by a discussion of multi-objective formulations. For multi-objective problems, four different types of formulations are considered: (i) formulations involving multiple impact metrics; (ii) formulations involving a combination of two detection and/or impact metrics; (iii) formulations involving a combination of three or more detection and impact metrics; and (iv) formulations involving an operational objective and one or more detection and/or impact metrics.

**Single objective sensor location methodologies**

All the methodologies based on any single objective are presented in Table 1 and described briefly as follows.

**Maximizing DC**

Lee & Deininger (1992) were the first to suggest an IP based methodology for the identification of a given number of MSs to maximize the DC based on selected coverage criteria. Woo *et al.* (2001) emphasized the use of water quality simulations to develop a coverage matrix. Kumar *et al.* (1997) and Kansal *et al.* (2012a) suggested a heuristic method for one by one selection of MSs. Al-Zahrani & Moied (2003) applied a GA and Afshar & Marino (2012) used ACO to solve the same problem. Formulation of a water fraction matrix to develop a coverage matrix based on selected coverage criteria is a tedious job for large size networks even with a simplified procedure of Kumar *et al.* (1997). Ghimire & Barkdoll (2006) suggested locating MSs at the nodes of high demands. Rathi & Gupta (2014a) suggested a heuristic method which avoids coverage criteria and thus formulation of a coverage matrix by considering all nodes on the shortest path as covered. Starting from zero, MSs are located one by one based on the maximum increase in total coverage due to additions in MSs.

**Minimizing TD**

Kumar *et al.* (1999) suggested considering TD as a measure of the level of service (LOS) to consumers. They identified the best monitoring locations using a travel time matrix for a desired LOS. Chastain Jr (2006) developed a heuristic methodology considering an extended period of water quality analysis for creating a database of water system responses by injecting a contaminant at each node. The method searched the best locations of MSs one by one to maximize the covered nodes with the condition of detection time. Cozzolino *et al.* (2006) considered uncertainty in the user’s demand to locate MSs for detection of contamination events within a specific time. Considering different contamination scenarios, probable locations of MSs were identified which satisfied the criteria of TD within a desired LOS. Priority was given to a node which could identify more contamination events. Pinzinger *et al.* (2011) compared approaches based on IP and two greedy heuristics for identifying the best locations of MSs for large networks to maximize the coverage of network nodes within the allowable time limit with a fixed number of MSs. Rathi & Gupta (2014b) also suggested a heuristic method which works on appended shortest travel time tree to identify the best locations of MSs which to achieve a desired LOS. However, satisfying TD as LOS requires a large number of MSs may not be possible, from a practical standpoint. Therefore, TD is considered as one of the objectives by defining it as the average or expected TD and minimizing it in multi-objective problems as shown later.

**Minimizing impact of contamination**

Kessler *et al.* (1998) suggested consideration of a pre-specified value of VC as LOS. 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 the location of MSs. Ostfeld & Salomons (2005a, b) extended the methodology to consider the randomness of flow rate of the injected pollutant, randomness in consumers’ demands and the detection sensitivity and response time of sensors.

Berry *et al.* (2005a) identified optimal sensor locations using MIP to maximize PE. Berry *et al.* (2006) and Propato (2006) formulated MIP models in such a way that a wide range of design objectives (TD, PE, VC, MC, and PFD) can be accommodated in the formulation, either individually or jointly. Several researchers incorporated uncertainty in
projected nodal demand to decide sensor locations with the objective of minimizing expected PE (Shastri & Diwekar 2006; Rico-Ramirez et al. 2007). Carr et al. (2004, 2006) considered uncertainty in contamination events (attack probability) and populations. Different MIP formulations were suggested to separately minimize the EC and PE under

Table 1 | Single objective sensor location methodologies

<table>
<thead>
<tr>
<th>Objective</th>
<th>Citation</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 IP</td>
<td>Hypothetical two loop network(^1) Network of Flint, Michigan(^2) and Cheshire, Connecticut(^3)</td>
<td>Fixed</td>
<td>Regular or routine monitoring</td>
</tr>
<tr>
<td>Kumar et al. (1997)</td>
<td>SHS Heuristic</td>
<td>Hypothetical network with two source nodes(^4)</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woo et al. (2001)</td>
<td>DH and WQS IP</td>
<td>Small hypothetical network(^5)</td>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Zahrani &amp; Moied (2003)</td>
<td>SHS GA</td>
<td>A hypothetical WDN with three source nodes(^6)</td>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansal et al. (2022a)</td>
<td>SHS Heuristic</td>
<td>Manendragarh Town Network(^7), India</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afshar &amp; Marino (2011)</td>
<td>SHS ACO</td>
<td>Network of City of Babol, Iran(^8)</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rathi &amp; Gupta (2014a, b, c)</td>
<td>SHS Heuristic</td>
<td>Networks(^9), (^10), Network of part of Nagpur city, India(^11)</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD Kumar et al. (1999)</td>
<td>DHS Heuristic</td>
<td>Any town(^12), USA Network(^13)</td>
<td>Variable</td>
<td>Accidental</td>
<td></td>
</tr>
<tr>
<td>Cozzolino et al. (2006)</td>
<td>DH &amp; WQS Heuristic</td>
<td>Any town(^12), USA</td>
<td>Variable</td>
<td>Demand uncertainty</td>
<td></td>
</tr>
<tr>
<td>Chastain (2006)</td>
<td>DH &amp; WQS Heuristic</td>
<td>Network(^12)</td>
<td>Variable</td>
<td>Intentional/ Accidental</td>
<td></td>
</tr>
<tr>
<td>Pinzinger et al. (2011)</td>
<td>DHS IP and heuristic</td>
<td>Network</td>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansal et al. (2012b)</td>
<td>SHS Heuristic</td>
<td>Network(^6)</td>
<td>Variable</td>
<td>Accidental</td>
<td></td>
</tr>
<tr>
<td>Kansal et al. (2014b)</td>
<td>SHS Heuristic</td>
<td>Networks(^9), (^10)</td>
<td>Variable</td>
<td>Accidental</td>
<td></td>
</tr>
<tr>
<td>VC Kessler et al. (1998)</td>
<td>DHS Heuristic</td>
<td>EPANET Ex.1(^14) and Network(^12)</td>
<td>Variable</td>
<td>Accidental</td>
<td></td>
</tr>
<tr>
<td>Ostfeld &amp; Salomons (2004)</td>
<td>DH &amp; WQS GA</td>
<td>EPANET Ex.3(^15) (2005a) Networks(^14), (^12) (2005b)</td>
<td>Fixed</td>
<td>Intentional</td>
<td></td>
</tr>
<tr>
<td>Ostfeld &amp; Salomons (2005a, b)</td>
<td>DH &amp; WQS GA</td>
<td>EPANET Ex.3(^15) (2005a) Networks(^14), (^12) (2005b)</td>
<td>Fixed</td>
<td>Accidental</td>
<td></td>
</tr>
<tr>
<td>PE Carr et al. (2004)</td>
<td>SHS Branch and bound Heuristics</td>
<td>Application not shown</td>
<td>Fixed</td>
<td>Also considered EC</td>
<td></td>
</tr>
<tr>
<td>Uber et al. (2004)</td>
<td>DH &amp; WQS MIP</td>
<td>Network(^20) Network(^21), EPANET 2.0 Ex.2(^16), Network of 470 nodes &amp; 621 pipes(^17)</td>
<td>Fixed</td>
<td>Intentional (Deterministic)</td>
<td></td>
</tr>
<tr>
<td>Berry et al. (2005a)</td>
<td>DHS L-shaped BONUS algorithm Stochastic decomposition</td>
<td>Network(^18), EPANET Ex.1(^19)</td>
<td>Variable</td>
<td>Uncertainty in AR and PD</td>
<td></td>
</tr>
<tr>
<td>Shastri &amp; Diwekar (2006)</td>
<td>DHS GA</td>
<td>Network(^14)</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rico-Ramirez et al. (2007)</td>
<td>DHS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwartz et al. (2014)</td>
<td>DHS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACO, ant colony optimization; AR, attack risk; BWSN, The Battle of the Water Sensor Network; DC, demand coverage; DHS, dynamic hydraulic simulation; DH & WQS, dynamic hydraulic and water quality simulation; GA, genetic algorithm; IP, integer programming; MIP, mixed integer programming; PD, population density; PE, population exposed to contamination; SHS, static hydraulic simulation; TD, time to detection; VC, volume of water consumed.

Superscript used with networks indicates networks considered by different researchers. These are the same if their numbers are the same.
different types of uncertainty considered as random uncertainty and interval uncertainty.

Murray et al. (2008) compared sensor placement solutions considering three objectives (PE, EC, and DL) using eight example networks of varying size for illustration and showed that the number of sensors needed for various objectives depends upon 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.

**Multi-objective sensor location methodologies**

All the methodologies based on multiple objectives are presented in Table 2 and described briefly as follows.

**Formulations involving multiple impact metrics**

As mentioned earlier, the MIP formulation of Berry et al. (2006) and Propato (2006) can accommodate several design objectives. Berry et al. (2009) showed how the imperfectness of sensors in not detecting an event can be modelled and included in a MIP formulation. Grayman et al. (2006) considered VC, PE, EC, and percentage node impacted (defined as nodes where the concentration of a contaminant is observed more than a threshold value) to check the effectiveness of a given sensor network under a set of contamination events.

**Formulations involving a combination of two detection and/or impact metrics**

Ehsani & Afshar (2010) used NSGA to minimize the expected detection time of detected events and maximize sensor DL in a multi-objective sensor location problem to generate a Pareto front and select a suitable solution. They examined minimizing expected detection time by considering a penalty value for undetected events along with the detection time of a detected event (Equation (2)) through GA. They observed that both the approaches provided good performance. GA results in faster solutions but provided a single solution, whereas NSGA is computationally more expensive but provided an optimal Pareto front. Kim et al. (2010) compared sensor locations by considering perfect and imperfect mixing at a node. They used a multi-objective GA to develop a Pareto front between DL and TD for both detected and undetected events. TD for undetected events was considered as twice the hydraulic simulation time. The perfect mixing condition was considered using EPANET and AZRED, which were used to model imperfect mixing. They observed that water quality solution was not considerably different for the example network. Also, limitations of ARZED were noted.

Krause et al. (2008) suggested single and weighted multi-objective formulations (Equation (12)), the maximization of which resulted in the reduction of total system-wide impacts associated with contamination scenarios. They suggested a greedy heuristic approach for step-wise placement of sensors. Further, the initial solution generated by the greedy algorithm was improved using SA. Comboul & Ghanem (2013) considered uncertainties in nodal demands for imperfect sensors in the design of resilient sensor networks. The formulation of Krause et al. (2008) was modified and solved using a greedy algorithm. Cozzolino et al. (2011) considered uncertainty in nodal demand and uncertainty in contamination attacks to develop the impact probability distribution. The approach based on minimization of a given percentile of impact probability distribution is compared with the usual approach of minimization of damages considered in terms of VC, MC, and PE for optimal location of sensors. Both of these approaches are found feasible for small to medium WDSs.

Dorini et al. (2010) considered minimizing PE along with maximizing DL in the formulation of Krause et al. (2008) to obtain optimal sensor locations using a heuristic method. They compared their algorithm with other greedy algorithms and also carried out sensitivity analysis by using different inputs to demonstrate the superiority and robustness of their algorithm. Xu et al. (2010a) suggested two robust models which minimize PE and maximize DL by converting them into a single objective using relative weights. A heuristic method was suggested to solve the models.

Guidorzi et al. (2009) considered the process of contamination detection and its flushing simultaneously by means of two consecutive optimization processes. The first process minimizes VC and NFD using NSGA-II. The second optimization process minimizes operations like hydrant opening and valve closing to reduce the expected volume of
<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>TD, VC, PE, EC and PFD</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>Considered objectives individually and jointly</td>
</tr>
<tr>
<td>PE, MC</td>
<td>Berry et al. (2009)</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>TD, DL</td>
<td>Ehsani &amp; Afshar (2010)</td>
<td>DH &amp; WQS</td>
<td>IP, local search and NLP</td>
<td>Networks(^{20,21,22})</td>
<td>Fixed</td>
<td>Imperfect sensors</td>
</tr>
<tr>
<td>TD, DL</td>
<td>Kim et al. (2010)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II &amp; GA</td>
<td>Networks(^{15})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>TD, PE, VC, 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</td>
<td>Krause et al. (2008)</td>
<td>DH &amp; WQS</td>
<td>Greedy and SA</td>
<td>Networks(^{7,8}), Large network (21000 nodes)(^{25})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>Comboul &amp; Ghanem (2015)</td>
<td>DH &amp; WQS</td>
<td>Greedy algorithm</td>
<td>Networks(^{26})</td>
<td>Fixed</td>
<td>Nodal demand uncertainty</td>
<td></td>
</tr>
<tr>
<td>TD, VC, PE, MC</td>
<td>Cozzolino et al. (2011)</td>
<td>DH &amp; WQS</td>
<td>GA</td>
<td>Networks(^{13})</td>
<td>Fixed</td>
<td></td>
</tr>
<tr>
<td>DL, 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>
<tr>
<td>DL, PE</td>
<td>Xu et al. (2010a)</td>
<td>DH &amp; WQS</td>
<td>Heuristic</td>
<td>Networks(^{7})</td>
<td>Fixed</td>
<td>Robust model for accidental contamination</td>
</tr>
<tr>
<td>VC, 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 and 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>TD, PE, VC, 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, 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, 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>TD, VC, DL</td>
<td>Aral et al. (2010)</td>
<td>DH &amp; WQS</td>
<td>PGA</td>
<td>Networks(^{7,8})</td>
<td>Fixed and variable</td>
<td>Reduced network(^{6,7})</td>
</tr>
<tr>
<td>TD, DL, SDR, 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>
</tbody>
</table>

(continued)
contaminated water consumed between the beginning of response operations and the disappearance of the contaminant. Weickgenannt et al. (2013) combined VC and NFD in a single objective, \( R \) (Equation (10)) and minimized the resulting single objective for placement of sensors.

Formulations involving a combination of three or more detection and impact metrics

Propato & Piller (2006) developed a weighted single objective formulation incorporating four objectives: TD, PE, VC, and DL. They investigated two main aspects: (i) identification of the contamination event ensemble size so as to estimate a meaningful expected function; and (ii) maximize the report time step without deteriorating design results. The concept of uncertainty was taken into account. Uncertainty is generated due to the report time step. Evaluation of uncertainty reduces the computational complexity and helps in evaluating how bad a suboptimal solution is with respect to the optimal solution.

Wu & Walski (2006) combined TD, VC, and DL in a single objective (Equation (11), without second term containing PE) to optimize using a GA. The TD for undetected events was taken as the length of a simulation duration and contamination events were randomly generated using a Monte Carlo simulation. Huang et al. (2010) considered three objectives (TD, PE, and DL) for determining optimal sensor locations using a GA. They initially developed a contamination detection database and ranked the nodes which have higher detection probability and used them as initial solutions for the GA. The results demonstrated computational effectiveness over the methodology of Preis & Ostfeld (2006). Aral et al. (2008) considered junctions with higher water demand values at potential sensor locations in order to reduce the search space. NSGA-II was used to work over the subset of junctions by considering three objectives: TD, PE, and DL. The results demonstrated computational effectiveness over the methodology of Preis & Ostfeld (2006). Aral et al. (2010) proposed a single objective formulation (Equation (13)) which combines the effect of four objectives of BWSN and PGA is used to solve the problem. Their results demonstrated that their model provides robust solutions as compared to others in BWSN.

Formulations involving an operational objective and one or more detection and/or impact metrics

Berger-Wolf et al. (2005) used a graph theoretic approach to solve the problem of sensor placement with the

Table 2 | continued

<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>TD, DL, SDR</td>
<td>Preis et al. (2008a)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>Networks(^{12}) &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(^{12})</td>
<td>Fixed</td>
<td>Imperfect mixing at nodes</td>
</tr>
<tr>
<td>DC, TD, VC, PE, 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>Time delay, SDR</td>
<td>Shen &amp; McBean (2011)</td>
<td>DH &amp; WQS</td>
<td>NSGA-II</td>
<td>City of Guelph(^{20})</td>
<td>Variable</td>
<td>Select specific incident list event</td>
</tr>
<tr>
<td>DL or CSDL</td>
<td>Xu et al. (2010b)</td>
<td>DH &amp; WQS</td>
<td>MIP solver</td>
<td>Networks(^{7})</td>
<td>Fixed</td>
<td>Imperfect sensors</td>
</tr>
</tbody>
</table>

CES, contamination event sampling; CSDL, contamination source detection likelihood; DC, demand coverage; DH & WQS, dynamic hydraulic and water quality simulation; DL, detection likelihood; EC, extent of contamination; IBSM, importance based sampling method; MC, mass consumed; MIP, mixed integer programming; MOGA, multi-objective genetic algorithm; NFD, number of failed detections; NLP, non-linear programming; NSGA, non-dominated sorted genetic algorithm; PD, population density; PE, population exposed to contamination; PFD, probability of failed detection; PGA, progressive genetic algorithm; SA, simulated annealing; SHS – static hydraulic simulation; SH & WQS, static hydraulic and water quality simulation; SLOT, sensor local optimal transformation system; TD, time to detection; VC, volume of water consumed.

Superscript used with networks indicates networks considered by different researchers. These are the same if their numbers are the same.
objective of minimization of TD along with sensor cost and identification of the source of contamination. Preis & Ostfeld (2006) developed a multi-objective model considering objectives as TD, DL, SDR, and CSDL for optimal sensor placement using the NSGA-II and trade-off between various objectives are explored. To reduce the computational requirements as the problem grows in size, Preis & Ostfeld (2008a) developed a heuristic method for identifying important contamination events for simulation and selects representative events which have similar geographic coordinates with the ones of the array of events of a WDS. They compared the sensor location based on contamination injection at all nodes excluding dead ends and sensor location and their results showed that sensor locations using their reduced sized matrix closely matched the sensor location based on full enumerated contamination matrix.

Austin et al. (2009) considered imperfect mixing conditions at pipe junctions for optimizing sensor placement. Their results demonstrated that sensor network design based on an imperfect mixing model is more reliable when compared to a perfect mixing model at pipe junctions.

Eliades & Polycarpou (2006) considered DC in addition to the four objectives used in BWSN and suggested a heuristic method in which Pareto optimal solutions were obtained for the first location by evaluating objective function values. The solutions are sorted based on higher DC. Then all solutions for the first location are used to identify Pareto solutions for the first two locations. The process of obtaining Pareto solutions continued for the desired number of sensor locations and the best solution is selected.

Shen & McBean (2011) considered two objectives: time delay (defined in terms of TD of both detected and undetected events by providing the largest time to undetected events) and SDR. A Pareto front was developed for determination of impact of increasing the number of sensors using NSGA-II. Xu et al. (2010b) developed a methodology for optimal sensor location considering two aspects: (i) maximizing coverage within allowable time limit; and (ii) inferring the probability of the occurrence of a contamination event and the possible contamination source. Based on a sensor signal they divide the system into subareas that is further linked to a Bayesian network to evaluate the overall detection probability of the sensor network system, the overall false-positive rates of the sensor network system and the contaminant source detection.

**CHALLENGES FOR RESEARCHERS AND RECOMMENDATIONS**

The application of MIP and genetic algorithms to sensor network design problems involving multiple objectives can present significant computational challenges. Some of the issues and strategies for addressing this challenge are summarized below.

**Original or reduced network**

A large water distribution network may involve thousands of pipes and nodes, and the numbers 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) reduces the search process, however, the entire network is required to be considered during analysis. Wu & Walski (2006) suggested reducing networks 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 which consider skeletonization using hydraulic equivalence. Skeletonization of networks considering hydraulics and water quality is difficult for multisource networks in which flow direction in some of the links changes with time due to changes in operating conditions such as pump on–off controls. Perelman & Ostfeld (2012b) considered changes in flow direction with time and suggested a methodology for clustering network nodes for various purposes. They suggested considering location of sensors cluster-wise and deciding its location within a cluster through optimization.

The basic purpose of reducing the size of network is to reduce the number of candidate nodes for location of...
sensors without compromising 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.

**Selection of contamination events**

Designing a robust sensor network to detect accidental or intentional contaminants is a difficult task because of the uncertain nature of contamination events (i.e., what, how much, when, where, and for how long an event will occur). A contaminant may enter a network from any point and at any time. Furthermore, 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 contaminant entry, a few vulnerable locations can be selected to reduce the computational work. The size of the problem is reduced by restricting number of injection locations and limiting injection time, by random sampling, by considering importance based sampling, or by designing a system based on consideration of high impact scenarios.

Chastain Jr (2006) considered a small network with a possible contamination event at every node in the system. Berry et al. (2006, 2009) considered only those nodes having demands as possible locations of injection and eliminated junction nodes with no demand at one fixed time to reduce the number of events to be considered. Furthermore, Berry et al. (2006) considered the period of contamination during the four periods of higher demands in a day. Xu et al. (2010a, b) considered the possibility of injection at any node and in any 5-minute interval. Dorini et al. (2010) considered possible events at each node and in each 5-minute interval for the first 24 hours of simulation. Krause & Guestrin (2009) developed an algorithm for solving large robust sensor placement problems and provided better results under both worst-case and average-case scenarios. Krause et al. (2008) and Krause & Guestrin (2009) selected contamination scenarios or events which affect at least 10% of the water distribution network nodes for simulation for large networks (BWSN network 2). Shen & McBean (2011) considered non-terminal nodes as possible injection locations with injection occurring at peak water demand. However, for large networks, approaches which are based on restricting number of injection locations and limiting injection time can still encounter computational problems, especially in the case where many injections are considered. Considering these aspects, researchers selected contamination events based on random sampling by generating random events using a Monte Carlo simulation technique for optimization (Wu & Walski 2006; Eliades & Polycarpou 2006) or by constructing the few representative samples whose geographic coordinates are similar to those of the entire set of possible samples by developing an equation of a reduced size contamination matrix which provided similar results when whole/full contamination events are simulated for optimization (Preis & Ostfeld 2008a). Furthermore, the possibility of simultaneous injections at two nodes which are selected randomly with different times are considered by Ostfeld & Salomons (2004) and Ostfeld et al. (2008). The sensor locations to reduce expected impact of contamination events which are calculated based on average of different impacts may result in high value at concentrated on high consequence contamination events. Watson et al. (2009) explore the nature of robust versus expected-case sensor placements and observed that robust sensor placements can yield large reductions in the number and magnitude of high-consequence events, with only modest increases in expected impact.

Consideration of importance-based sampling is provided by Perelman & Ostfeld (2010, 2012b) and Weickgenannt et al. (2010). Perelman & Ostfeld (2010, 2012b) suggested a crossed entropy approach to select critical contamination events which 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 VC 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 consequences events which can be identified based on importance of locality or infrastructures can be considered to reduce number of contamination events. Robustness of the solution can be checked with low consequences events.

**Event modelling**

Both static and dynamic simulation for event modelling have been considered by various researchers. Berry et al. (2005c) compared the static and dynamic sensor placement formulations and observed that static simulation allows a relatively fast determination of optimal solution and hence are applied to large scale problems. On the other hand, dynamic simulations provide more realistic scenarios and hence are more accurate in the determination of contaminant assessment. Consideration of each and every scenario requires 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. Furthermore, most of the studies consider contaminants as conservative (Dorini et al. 2010; Weickgenannt et al. 2010; Xu et al. 2010a). Some studies consider changes in the contaminant due to chemical reaction (Cozzolino et al. 2005) and recently Thienen (2014) and Ohar et al. (2015) used EPANETmx for event modelling with contaminant reactions.

**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 objectives and some complementary objectives can be dropped. Unified single objective formulation (Aral et al. 2010) jointly considers various parameters affecting sensor locations which make the problem formulation simple and provide better sensor locations as compared to other methodologies.

**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 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 which can prioritize a selection with respect to different objectives considered in the sensor location problems and can be easily applied to large water distribution networks. Prioritization selection helps in future extension of monitoring locations. Software like TEVA-SPOT (Berry et al. 2008, 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.

**Uncertainty consideration in sensor design**

Researchers have considered uncertainty in various factors such as consumer demand or changing the population density at various junctions, sensor sensitivity and response time (i.e. sensor accuracy) to design sensor network. Berry et al. (2005a) considered restricted nature of uncertainty in population density and models it by considering various noise levels along with uncertain attacks. Rico-Ramirez et al. (2007) extended the model by considering uncertainty in occurrence of contamination event and population density. Shastri & Diwekar (2006) suggested stochastic optimization formulation and considers the effect of demand uncertainty. Cozzolino et al. (2006) considered nodal demand uncertainty. The problem of deploying increasing set of sensors was approached by considering the allocation of sensors one at a time. Ostfeld & Salomans (2005a, b) extended their previous methodology (Ostfeld & Salomons 2004) to consider the randomness of flow rate of injected pollutant, randomness in consumers’ demands, and the detection sensitivity and response time of MSs. Comboul & Ghanem (2015) obtained the sensor location regarding the uncertainty in water demands and the sensor accuracy.
GENERAL DISCUSSION

While extensive research has been conducted in the development of algorithms for use in sensor placement (Hart & Murray 2010; Rathi & Gupta 2014), widespread application of such algorithms to actual systems appears to be somewhat limited. To date, USEPA (2015a) has designed and deployed pilot systems in Cincinnati, San Francisco, New York City, Philadelphia, and Dallas as part of its Water Security Initiative. Each of these installations has been fairly expensive. For example, in 2008, USEPA (2015b) provided New York City with $12 million to help it ‘develop and evaluate a drinking water contamination warning systems for its drinking water supply.’ In addition to these efforts, USEPA has worked to replicate such implementations by working with other utilities in several states (e.g. Colorado, Florida, Michigan, New York, North Carolina, Oklahoma, and Washington). Skadsen et al. (2008) provide an overview of one such implementation for the city of Ann Arbor, Michigan. It is likely that adoption of such technology by other utilities has been somewhat limited due to several factors, including the installation and operating cost of such systems and possible lingering concerns about either the accuracy or ultimate need for such systems. This paper has focused primarily on the perceived accuracy of such systems as impacted by the ability to properly formulate the actual problem to be solved and the resulting computational challenges associated with solving that problem.

The sensor location problem for large practical networks considering all its complexities is observed to be a challenging task. The complexities in sensor network design are due to several reasons. Some of them include: flow reversal in some links (requiring consideration of either multiple or dynamic simulations), uncertainty in attack in terms of location, time, type, strength and duration of contaminant intrusion (requiring consideration of a large number of contamination scenarios and water quality simulations), uncertainty in nodal demands, sensor detection limits, delay by sensor in registering an event and raising alarm, imperfectness of sensors resulting in false alarms, imperfect mixing at nodes, uncertainty involved in quantification of impacts on human health and other impacts, etc. Sensor network design to select optimal locations of a limited number of sensors involving all the complexities are computationally expensive. Further, it requires a calibrated hydraulic model. Thus, one school of thought is to combine various objectives, limit probable location of attacks, limit probable location of sensors, reduce the size of the network itself, consider sensor imperfectness, and suggest a method which is not computational intensive (may be heuristic) to obtain an optimal solution and check the robustness of solution (Perelman et al. 2008; Watson et al. 2009; Krause & Guestrin 2009; Hart & Murray 2010; Xu et al. 2010a; Perelman & Ostfeld 2012b, 2013; Davis et al. 2013; Eliades et al. 2014; Schwartz et al. 2014). Chang et al. (2011, 2012) suggested a sensor deployment strategy with less computational requirements. A new rule-based expert system, based on accessibility rule and complexity rule, was developed to tackle the complexity of the network and reduce the computer runtime while achieving the same level of robustness in planning and design. The accessibility rule ranks different nodes based on the number of downstream nodes that can be protected by a node; while complexity rule ranks different nodes based on the number of inner nodes. Klise et al. (2015) suggested a two-stage approach, in which sensors are placed at nodes of a reduced network in the first-stage and transferred to nodes of the original network in the second stage.

The other school of thought is to develop methodologies for real world problems to tackle the complexity of the network through simple parameters using graph theoretic and heuristic approaches. Xu et al. (2008) suggested a procedure for identifying a set of key nodes for placing sensors based on ‘time-constrained receivability’. Receivability is defined by the existence of one or more flow paths. If flow from any node i could reach to node j, then node i is considered to be receivable at node j. When the condition of time is imposed for flow to reach from node i to node j, the receivability becomes time-constrained. Chung et al. (2010) suggested a methodology based on ‘betweenness centrality’ defined as the centrality of a node in terms of degrees to which the node falls on the shortest path between other pairs of nodes. Deuerlein et al. (2010) suggested capturing water hammer pressure conditions in the network due to pumping of water for intrusion of contaminant to reduce...
long detection times. Perelman & Ostfeld (2011, 2012b) suggested connectivity analysis for topological clustering of nodes. Besides betweenness centrality, Perelman & Ostfeld (2013) considered various other measures of graphs like ‘Eigenvector Centrality’, ‘Closeness Centrality’, ‘Page Rank’, ‘Edge Betweenness Community’, and ‘Kmeans’ for the location of sensors. Schal et al. (2013) studied optimally designed sensor locations (obtained through TEVA-SPOT (Berry et al. 2008; Hart et al. 2008b; Murray et al. 2010) on different types of network (Branch, Loop and Grid) for various graph parameters such as betweenness centrality, receivability, proximity to source or filling station with a view to develop general guidelines for the design of sensor networks for a small utility in case a calibrated network model is not available.

**SUMMARY AND CONCLUSIONS**

A complete review of methodologies for sensor location is provided. Methodologies are grouped based on the number of objectives as single and multi-objective techniques. Multi-objective techniques are further classified based on objectives related to the contamination event. A full consideration of all the complexities of sensor placement, such as those arising due to uncertainty of the contamination event, network characteristics, demand variability, sensor characteristics, and limited budget, is computationally intensive when applying such optimization methodologies to large WDNs. Thus, the main issue in the sensor location problem to develop a methodology which can be deployed easily for optimal sensor placement in large water supply systems with all the involved complexities. The other important aspect is the unavailability of a calibrated network model for most water utilities which is the prime necessity for these methodologies. Hence, research is needed to develop methodologies based on some topological parameters and available hydraulic data by testing its robustness under different conditions. Finally, it is quite possible that future technologies may ultimately eliminate the need for the location and design of static MSs. Recent research in the use of in-line mobile sensors (Perelman & Ostfeld 2022a) for such monitoring may hold the prospect for more dynamic solutions to this problem. However, as with static sensors, it is likely that additional technological advances will be necessary before such technology can be reliably implemented. As with static sensors, such technologies will likely bring their own set of computational challenges (Rasekh et al. 2015).

**REFERENCES**


US Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.


First received 20 December 2014; accepted in revised form 26 May 2015. Available online 10 June 2015