Multiobjective Contaminant Sensor Network Design for Water Distribution Systems

Ami Preis and Avi Ostfeld, M.ASCE

Abstract: A contaminant intentional intrusion into a water distribution system is one of the most difficult threats to address. This is because of the uncertainty of the type of the injected contaminant and its consequences, and the uncertainty of the location and intrusion time. An online contaminant sensor network is the main constituent to enhance the security of a water distribution system against such a threat. In this study a multiobjective model for water distribution system optimal sensor placement using the nondominated sorted genetic algorithm II is developed and demonstrated using two water distribution systems of increasing complexity. Tradeoffs between three objectives are explored: (1) sensor detection likelihood; (2) sensor detection redundancy; and (3) sensor expected detection time. Pareto fronts are plotted for pairs of conflicting objectives, and simultaneously for all three. A contamination event heuristic sampling methodology is developed for overcoming the problem of contamination event sampling.


CE Database subject headings: Sensors; Water distribution systems; Water pollution; Sampling; Network design.

Introduction

Water distribution systems, typically, are comprised of tanks, pipes, and pumps delivering treated water from treatment plants to consumers. Even a moderate network might have hundreds of kilometers of pipes and numerous delivery points, making it inherently vulnerable.

Threats on a water distribution system can be partitioned into three major groups according to their resulted enhanced security: (1) a direct attack on main infrastructure: dams, treatment plants, storage reservoirs, pipelines, etc.; (2) a cyber attack disabling the functionality of the water supervisory control and data acquisition (SCADA) system, taking over control of key components that might result in water outages or insufficiently treated water, changing or overriding protocol codes, etc.; and (3) a deliberate chemical or biological contaminant injection at one of the system’s nodes.

The threat of a direct attack can be minimized by improving the system’s physical security (e.g., additional alarms, locks, fencing, surveillance cameras, guarding, etc.), while a cyber attack can be minimized by implementing computerized hardware and software (e.g., an optical isolator between communication networks, routers to restrict data transfer, etc.).

Of the above threats, a deliberate chemical or biological contaminant injection is the most difficult to address. This is because of the uncertainty of the type of the injected contaminant and its effects, and the uncertainty of the location and injection time. Principally, a contaminant can be injected at any water distribution system connection (node) using a pump or a mobile pressurized tank. Although backflow preventers provide an obstacle, they do not exist at all connections, and at some might not be functional.

The main course to enhance the security of a water distribution system against a deliberate contamination intrusion is through placing a sensor system (ASCE 2004; AWWA 2004).

This paper extends previous studies on sensor placement by developing a multiobjective model for water distribution system optimal sensor layout. The developed model employs the non-dominated sorted genetic Algorithm–II (NSGA-II) (Deb et al. 2000), and is demonstrated using two water distribution systems of increasing size. Pareto fronts are explored for three objectives: (1) sensor detection likelihood; (2) sensor detection redundancy; and (3) sensor expected detection time.

This work also suggests a heuristic scheme for establishing a representative contamination matrix for sensor placement. A contamination matrix (Ostfeld and Salomons 2004) is an $N \times M$ matrix of 0–1 coefficients, where $N$=number of possible injection locations (nodes); $M$=number of contamination events considered; and “0” and “1” correspond to noncontaminated/contaminated nodes, respectively. The $i$th row of the matrix lists all contamination events that contaminated node $i$, while the $j$th column lists all nodes contaminated by the $j$th contamination event. From a detection perspective, the $i$th row corresponds to the domain of detection of a sensor placed at node $i$, while the $j$th column corresponds to the domain of pollution of a contamination event. As a water distribution system becomes more complex in terms of its size, loading conditions, controls, etc., the difficulty of establishing a representative contamination matrix also becomes more difficult (i.e., a matrix of reduced size that provides similar results as a full event enumerated contamination matrix).

The development of a reduced contamination matrix enables...
the implementation of the proposed methodology to more complex water distribution systems.

**Literature Review**

In recent years there has been growing interest in the development of sensor systems with the majority of models using a single objective approach. The employment of multiobjective optimization for sensor placement started recently. This section reviews models for sensor placement using the two approaches.

**Single Objective Sensor Placement Models**


**Multiobjective Sensor Placement Models**

Watson et al. (2004) were the first to introduce a multiobjective formulation to sensor placement by employing a mixed-integer linear programming model over a range of design objectives. Recently, the battle of the water sensors (Ostfeld et al. 2006) highlighted the multiobjective nature of sensor placement, with the following multiobjective models: Dorini et al. (2006) developed a constrained multiobjective optimization framework entitled the noisy cross-entropy sensor locator (nCESL) algorithm based on the cross-entropy methodology proposed by Rubinstein (1999); Eliades and Polycarpou (2006) proposed a multiobjective solution using an “iterative deepening of Pareto solutions” algorithm; Gueli (2006) suggested a predator–prey model applied to multiobjective optimization, based on an evolution process; Huang et al. (2006) proposed a multiobjective genetic algorithm framework coupled with data mining; Ostfeld and Salomons (2006) and Preis and Ostfeld (2006) used the multiobjective NSGA-II (Deb et al. 2000) scheme; Wu and Walski (2006) used a multiobjective optimization formulation, which was solved using a genetic algorithm, with the contamination events randomly generated using a Monte Carlo procedure.

**Methodology**

Multiobjective optimization can be defined as the problem of finding the vector of decision variables that satisfies a set of constraints and optimizes a vector function whose elements represent the objective functions. Consequently, a multiobjective optimization problem can be formalized as

\[
\begin{align*}
\text{Optimize} & \quad \mathbf{f}(\mathbf{x}) = [f_1(x), f_2(x), \ldots, f_M(x)]^T \\
\text{Subject to} & \quad \mathbf{g}_i(x) > 0, \quad i = 1, 2, \ldots, k \quad k \text{ inequality constraints} \\
& \quad \mathbf{e}_j(x) = 0, \quad j = 1, 2, \ldots, l \quad l \text{ equality constraints}
\end{align*}
\]

where \( \mathbf{x} = (x_1, x_2, \ldots, x_n)^T \) is a vector of decision variables.

The goal is to find, from all the sets of solutions that satisfy Eqs. (2) and (3), the set of solutions that yield optimal values with respect to all the objective functions. This set of solutions is entitled the Pareto optimal solution set or the nondominated solution set. Each solution \( x \) in the Pareto optimal set is optimal in the sense that it is not possible to improve one objective without making at least one of the others worse.

Any two solutions \( x^{(1)} \) and \( x^{(2)} \) are compared based on domination, where a solution \( x^{(1)} \) is said to dominate \( x^{(2)} \) if the following conditions hold:

1. \( x^{(1)} \) is no worse than \( x^{(2)} \) in all objectives:
   \[ f_j(x^{(1)}) \leq f_j(x^{(2)}) \forall j, j = 1, \ldots, M \]  
2. \( x^{(1)} \) is strictly better than \( x^{(2)} \) in at least one objective:
   \[ f_j(x^{(1)}) < f_j(x^{(2)}) \exists j, j = 1, \ldots, M \]

where “\( \leq \)” indicates a better performance evaluation of an objective function; and \( M \) is the number of objective functions.

There are two interconnected conceptual targets in multiobjective optimization. These are to find a set of solutions as close as possible to the Pareto optimal set, and to guarantee that the set of solutions is as diverse as possible.

In recent years several evolutionary methods have been developed that extend single objective evolutionary schemes to multiobjective algorithms. Three of the more employed algorithms are: (1) the multiobjective genetic algorithm (MOGA) (Fonseca and Fleming 1995); (2) the NSGA-II (Deb et al. 2000); and (3) the strength Pareto evolutionary Algorithm II (SPEA-II) (Zitzler et al. 2001).

In this study the NSGA-II (Deb et al. 2000) was selected following its general implementation success (Deb et al. 2002), and, in particular, its success for multiobjective water distribution systems management problems. Prasad and Park (2004) solved the optimal design problem of a water distribution network by minimizing the network cost versus maximizing a network resilience measure; Prasad et al. (2004) minimized the total disinfectant dose of booster chlorination stations versus the maximization of the volumetric demand; Vanvakeridou-Lyroudia et al. (2005) maximized the benefits of a water distribution system design problem versus cost, with the benefits evaluated using fuzzy logic reasoning. Other multiobjective algorithms like MOGA (Fonseca and Fleming 1995), or SPEA-II (Zitzler et al. 2001) could be applied as well in this study.

NSGA-II is a multiobjective evolutionary algorithm that alleviates the difficulties faced by previous nonelitist multiobjective evolutionary algorithms: (1) computational complexity; (2) non-elitism; and (3) need for identifying a sharing parameter.

The algorithm employs a nondominated sorting approach using a selection operator that creates a mating pool by combining the parent and offspring populations, and by selecting solutions with respect to fitness and spread. Generations are populated
starting with the best nondominated front and succeeding until the specified population size is reached. If at the final stage there are more individuals in the nondominated front than the available space, the crowded distance-based niching strategy is invoked to choose which individuals of that front will enter into the next population. A flowchart of the NSGA-II scheme implemented in this study is described in Fig. 1.

**Objective Functions**

Three objectives for sensor placement were utilized: (1) sensor detection likelihood; (2) sensor detection redundancy; and (3) sensor expected detection time. Their definitions are provided below.

**Sensor Detection Likelihood**

Given a sensor network (i.e., number and locations), the detection likelihood (i.e., the probability of detection) is estimated by

\[ f_1 = \frac{1}{S} \sum_{r=1}^{S} d_r \]

where \( d_r = 1 \) if contamination event (i.e., an intrusion through a node of a contamination at a specific flow rate, concentration, and duration) \( r \) is detected, and zero otherwise; and \( S \) denotes the total number of contamination events considered. \( f_1 \) is to be maximized.

**Sensor Detection Redundancy**

The development of sensors aimed at revealing in real time the presence of contaminations is an ongoing effort. However, it is already clear that uncertainty will be involved in sensor detections. Thus, to minimize false positive indications and to increase the reliability of the sensor network there is a need to maximize a redundancy measure of the sensor system.

In this study a triply redundant measure is assumed where at least three sensors are required to reveal the presence of a nonzero contaminant concentration at a duration of a maximum of 30 min between the first and third detections. For each contamination event the redundancy of a sensor network design is equal to 1 if this condition holds, and zero otherwise.

Thirty minutes and three detections are parameters that were set heuristically to represent a level of required performance. Increasing the number of detections and/or decreasing the value of thirty will obviously impose a larger and denser sensor network requirement.

The redundancy \( f_2 \) for a given sensor network design is

\[ f_2 = \frac{1}{S} \sum_{r=1}^{S} R_r \]

where \( R_r = \text{sensor network redundancy for the } r\text{th contamination event, zero or one, as defined above.} \]

\( f_2 \) is to be maximized.

**Sensor Expected Detection Time**

For a given contamination event, the time of detection by a sensor is the elapsed time from the start of the contamination event, to the first identified presence of a nonzero contaminant concentration. The time of first detection, \( t_i \), refers to the \( i\text{th sensor location. The time of detection for the sensor system for a particular contamination event, } t_d = \text{minimum among all sensors present in the design} \)

\[ t_d = \min_i t_i \]

The sensor expected detection time is computed over an assumed uniform probability distribution of contamination events

\[ f_3 = E(t_d) \]

where \( E(t_d) \) denotes the mathematical expectation of the minimum detection time \( t_d \). \( f_3 \) is to be minimized.

**Observations**

1. The three selected objectives characterize three main requirements from a sensor network: a measure of the probability of detection expressed by the detection likelihood \( f_1 \), a measure of the sensor network reliability expressed by the redundancy \( f_2 \), and a measure of contamination exposure—represented by the expected detection time \( f_3 \). Other surrogate objectives could be similarly used. For example: a different methodology other than the employment of a contamination matrix for...
evaluating \( f_1 \), a reliability index incorporating fuzziness of the detection ability of the sensors for \( f_2 \), and expected population affected prior to detection for \( f_3 \).

2. Objectives are evaluated only against the events that are detected. Krause et al. (2006) noted and addressed this limitation.

3. It is anticipated that the three objectives will mutually compete. If only the detection likelihood would have been maximized then the sensors would probably be placed at the most downstream network nodes (i.e., the nodes located at the most downstream flow paths, usually close to or at the deadends of the distribution network). However, to minimize the expected time of detection the sensors need to be placed as close as possible, to the contaminant injection locations, thus competing against the detection likelihood. To maximize the sensor network redundancy the sensor layout is required to be as dense as possible, thus competing against the detection likelihood, and consequently against the expected time of detection. Verification of these arguments is provided through the example applications by analyzing the tradeoff curves among pairs of objectives and simultaneously for all three.

**Example Applications**

The methodology is demonstrated using two example applications of increasing complexity: EPANET (USEPA 2002) Example 3 (Fig. 2), and the Richmond Water System (CWS 2001), shown in Fig. 3.

EPANET Example 3 consists of two constant head sources: a lake and a river, three elevated storage tanks, 120 pipes, 94 nodes (consumers and internal nodes) and two pumping stations. The system is subject to a demand flow pattern of 24 h. The Richmond Water System is subject to a demand cycle of 72 h and is comprised of one source, 949 pipes, 865 nodes (consumers and internal nodes), seven pumping stations, and six tanks. Full EPANET input files can be downloaded for both systems from the corresponding cited references.

**Contamination Matrix Construction**

To evaluate the fitness of the sensor network detection likelihood \( (f_1) \), the sensor network detection redundancy \( (f_2) \), and the sensor network expected time of detection \( (f_3) \), a contamination matrix (Ostfeld and Salomons 2004) needs to be constructed. To establish a contamination matrix, assumptions are to be made for the number, location, starting time, mass rate, and duration of the injection events.

The following assumptions were made in constructing the con-
The contamination matrix is established by solving between 60 and 600 min. The range of 50–500 gr contamination events participated to provide similar results as if using the full enumerated nation events. Developed for sampling a small yet representative set of contamination injection mass rate, and an injection duration of 60–600 min, are assumed model parameters for this study.

Since contamination events can occur at any node at any time, the possible number of contamination events grows significantly with system size. To cope with this difficulty a heuristic procedure was developed for sampling a small yet representative set of contamination events (i.e., a set of contamination events that is anticipated to provide similar results as if using the full enumerated event contamination matrix).

The set of contamination events comprising the reduced sized contamination matrix is established by solving

\[
\text{Minimize } \sum_{i=1}^{5} |AS_i - AN_i| + |\sigma S_i - \sigma N_i| \quad (10)
\]

Subject to: \( q_j > 0 \), \( j = 1, \ldots, N \) \quad (11)

where \( AS_i \), \( \sigma S_i \) = average and standard deviation values, respectively, of the geographical \( x \) coordinate of a sampled contamination events set (e.g., if four contamination events were randomly sampled with geographical \( x \) coordinates of 312.2, 145.6, 241.5, and 74.8, then \( AS_i = 193.5 \), \( \sigma S_i = 104.5 \)); \( AN_i \), \( \sigma N_i \) = average and standard deviation values, respectively, of the geographical \( x \) coordinate of the water distribution system nodes; \( q_j \) = outgoing flow from node \( j \), and \( N \) = total number of system nodes. Similarly, \( i = 2 \) refers to the sampled contamination events geographical \( y \) coordinate; \( i = 3 \) to the injection mass rate; \( i = 4 \) to the injection starting time; and \( i = 5 \) to the injection duration. Note that all \( AN_i \) and \( \sigma N_i \) values are known as the layout of the system is given, and an assumption is made on the probability distributions of the injection mass rates, the injection starting times, and the injection durations.

Eq. (10) represents a minimization of the sum of the average and standard deviation characteristics of the sought sample compared to the average and standard deviation characteristics of the full contamination events matrix [i.e., Eq. (10) is a measure of the sample representativeness, with zero corresponding to a full representativeness that coincides with using a sample equal to the entire set of possible contamination events]. Eq. (11) eliminates from sampling nodes that are located at deadends of the system, as \( q_j \) refers to flows outgoing of node \( j \). The optimization problem defined in Eqs. (10) and (11) is solved using a genetic algorithm.

Fig. 4 shows the locations of five sensors for EPANET Example 3, obtained by maximizing the detection likelihood using: (1) a reduced matrix of 1,000 contamination events constructed by the above heuristic procedure; (2) 80,640 contamination events established by injecting at 84 nodes (i.e., all system nodes excluding deadends) every 30 min four mass injection rates of 50, 200, 350, and 500 gr/min, at five injection durations of 60, 220, 380, 500, and 600 min [i.e., \( 84 \times (2 \times 24) \times 4 \times 5 = 80,640 \) events]; and (3) a random contamination matrix of 1,000 events. It can be seen from Fig. 4 that the resulting sensor locations using the proposed reduced matrix closely matched the sensor locations based on the 80,640 events, while the results from the 1,000 random contamination events did not perform as well.

It should be noted that the results shown in Fig. 4 are not a general proof of the representativeness ability of the proposed

![Fig. 4. Comparison of sensor placement for EPANET Example 3 using different contamination matrices](image-url)
reduced contamination matrix scheme. The quality of results will differ in accordance with the water distribution system properties, and the injection events probabilistic characteristics.

Model Results

A base run (BR) and two sensitivity analysis runs are presented: Figs. 5 and 6 summarize BR for EPANET Example 3, and for the Richmond Water System, respectively, for placing five sensors; Figs. 7 and 8 describe the results of Sensitivity Analysis 1 (SA1); and Figs. 9 and 10 show the outcome of Sensitivity Analysis 2 (SA2). Tables 1–3 provide complementary results to the base run and the sensitivity analysis, respectively. Fig. 11 summarizes all solutions.

One thousand and 3,000 random contamination events were used for establishing the reduced contamination matrices for EPANET Example 3, and for the Richmond Water System, respectively. Genetic algorithm parameters of crossover probability of 0.75 and mutation probability of 0.1 were employed following Prasad et al. (2004) and Prasad and Park (2004), who showed good performance results for water distribution system optimization problems with these values. The maximum number of generations was set to 1,000 with each generation having a population of 640 strings. The NSGA-II algorithm was terminated if either the upper bound of 1,000 generations was attained or if at 10 successive generations no new nondominated solutions were found. The average computational time of a single generation for EPANET Example 3 was about 1 min, and for the Richmond Water System it was about 10 min, on an IBM PC 3.6 GHz, 1 GHz of RAM (i.e., a maximum running time of about 17 h for EPANET Example 3, and about a week for the Richmond Water System).

Base Run Results

Figs. 5 and 6 summarize a BR for EPANET Example 3, and for the Richmond Water System, respectively, for placing five sensors. In Figs. 5 and 6, optimal Pareto fronts are plotted for: minimizing \( f_3 \) versus maximizing \( f_1 \) (top-left corner); maximizing \( f_2 \) versus maximizing \( f_1 \); and (3) maximizing \( f_2 \) versus mini-
Fig. 6. Richmond Water System base run Pareto fronts, and selected sensor location solutions, where $f_1 =$ detection likelihood (%), $f_2 =$ sensor detection redundancy (%), and $f_3 =$ expected time of detection (minutes).

Fig. 7. Sensitivity analysis 1 for EPANET Example 3
Fig. 8. Sensitivity analysis 1 for Richmond Water System

Fig. 9. Sensitivity analysis 2 for EPANET Example 3

Fig. 10. Sensitivity analysis 2 for Richmond Water System
Table 1. NSGA-II Solution Characteristics

<table>
<thead>
<tr>
<th>NSGA II repetition</th>
<th>Number of generations to convergence</th>
<th>Number of nondominated solutions at Pareto front</th>
<th>Solution dominance strengtha</th>
<th>Euclidean normalized generational distance to true Pareto frontb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>290</td>
<td>101</td>
<td>101</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>2</td>
<td>287</td>
<td>107</td>
<td>107</td>
<td>6.00E-03</td>
</tr>
<tr>
<td>3</td>
<td>401</td>
<td>122</td>
<td>109</td>
<td>1.70E-02</td>
</tr>
<tr>
<td>4</td>
<td>315</td>
<td>110</td>
<td>110</td>
<td>1.00E-03</td>
</tr>
<tr>
<td>5</td>
<td>330</td>
<td>117</td>
<td>101</td>
<td>2.00E-04</td>
</tr>
<tr>
<td>6</td>
<td>370</td>
<td>115</td>
<td>115</td>
<td>1.00E-04</td>
</tr>
<tr>
<td>7</td>
<td>292</td>
<td>97</td>
<td>95</td>
<td>1.40E-02</td>
</tr>
<tr>
<td>8</td>
<td>367</td>
<td>111</td>
<td>111</td>
<td>2.70E-04</td>
</tr>
<tr>
<td>9</td>
<td>355</td>
<td>110</td>
<td>105</td>
<td>5.00E-04</td>
</tr>
<tr>
<td>10</td>
<td>311</td>
<td>114</td>
<td>100</td>
<td>7.26E-04</td>
</tr>
<tr>
<td>11</td>
<td>372</td>
<td>109</td>
<td>109</td>
<td>4.00E-04</td>
</tr>
<tr>
<td>12</td>
<td>300</td>
<td>95</td>
<td>92</td>
<td>1.72E-02</td>
</tr>
<tr>
<td>13</td>
<td>393</td>
<td>128</td>
<td>118</td>
<td>2.03E-04</td>
</tr>
<tr>
<td>14</td>
<td>355</td>
<td>109</td>
<td>109</td>
<td>1.90E-04</td>
</tr>
<tr>
<td>15</td>
<td>370</td>
<td>115</td>
<td>115</td>
<td>1.00E-04</td>
</tr>
<tr>
<td>16</td>
<td>347</td>
<td>108</td>
<td>108</td>
<td>1.70E-04</td>
</tr>
<tr>
<td>17</td>
<td>380</td>
<td>120</td>
<td>110</td>
<td>9.00E-04</td>
</tr>
<tr>
<td>18</td>
<td>400</td>
<td>125</td>
<td>100</td>
<td>1.30E-03</td>
</tr>
<tr>
<td>19</td>
<td>358</td>
<td>110</td>
<td>110</td>
<td>3.10E-04</td>
</tr>
<tr>
<td>20</td>
<td>373</td>
<td>112</td>
<td>112</td>
<td>1.10E-04</td>
</tr>
<tr>
<td>21</td>
<td>345</td>
<td>116</td>
<td>109</td>
<td>7.30E-04</td>
</tr>
<tr>
<td>22</td>
<td>376</td>
<td>129</td>
<td>117</td>
<td>1.60E-04</td>
</tr>
<tr>
<td>23</td>
<td>320</td>
<td>103</td>
<td>103</td>
<td>2.40E-03</td>
</tr>
<tr>
<td>24</td>
<td>337</td>
<td>117</td>
<td>114</td>
<td>8.80E-04</td>
</tr>
<tr>
<td>25</td>
<td>328</td>
<td>109</td>
<td>109</td>
<td>1.90E-03</td>
</tr>
<tr>
<td>Average</td>
<td>346.9</td>
<td>112.4</td>
<td>107.6</td>
<td>3.11E-03</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>34.6</td>
<td>8.56</td>
<td>6.58</td>
<td>5.44E-03</td>
</tr>
<tr>
<td>Base run comparisonc</td>
<td>370</td>
<td>115</td>
<td>115</td>
<td>1.00E-04</td>
</tr>
</tbody>
</table>

aSolution dominance strength=number of nondominated solutions present at the true Pareto front.
bDeb (2001, p. 326), following Veldhuizen (1999); 200 min served as the reference detection time for sensor detection time normalization (i.e., corresponding to a value of 1.0 for the detection likelihood).

cSee Figs. 5: sensor detection time versus sensor detection likelihood (top-left corner).

mizing ($f_3$) versus maximizing ($f_1$). At the bottom-right corner in Figs. 5 and 6, selected sensor solution locations from the Pareto fronts are provided.

Table 1 summarizes statistics of 25 NSGA-II repetitions for establishing the Pareto front for the case of sensor detection time versus sensor detection likelihood (Fig. 5, top-left corner).

Table 2. Base Run Selected Solutions

<table>
<thead>
<tr>
<th>Solution IDa</th>
<th>Node number</th>
<th>Detection likelihood ($f_1$) (%)</th>
<th>Sensor detection redundancy ($f_2$) (%)</th>
<th>Expected detection time ($f_3$) (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>15, 35, 147, 225, 255</td>
<td>84</td>
<td>25</td>
<td>162</td>
</tr>
<tr>
<td>1.1.2</td>
<td>10, 61, 117, 119, 265</td>
<td>17</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>1.1.3</td>
<td>204, 207, 209, 211, 217</td>
<td>54</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td>1.1.4</td>
<td>35, 149, 183, 215, 255</td>
<td>72</td>
<td>46</td>
<td>114</td>
</tr>
<tr>
<td>1.2.1</td>
<td>93, 352, 428, 600, 672</td>
<td>55</td>
<td>15</td>
<td>206</td>
</tr>
<tr>
<td>1.2.2</td>
<td>64, 71, 184, 197, 318</td>
<td>23</td>
<td>51</td>
<td>65</td>
</tr>
<tr>
<td>1.2.3</td>
<td>532, 542, 561, 563, 567</td>
<td>20</td>
<td>99.8</td>
<td>174</td>
</tr>
<tr>
<td>1.2.4</td>
<td>162, 300, 564, 604, 670</td>
<td>46</td>
<td>21</td>
<td>177</td>
</tr>
</tbody>
</table>

aSee Fig. 5 and 6.

In Table 1: (1) the true Pareto front is defined as the nondominated solution set obtained by all the 25 NSGA-II runs; (2) the solution dominance strength as the number of nondominated solutions of a single NSGA-II run present at the true Pareto front; and (3) a measure of distance between Pareto fronts as the Euclidean normalized generational distance measure (Deb 2001, p. 326), following Veldhuizen (1999); 200 min served as the reference detection time for sensor detection time normalization (i.e., corresponding to a value of 1.0 for the detection likelihood).
Sensitivity Analysis Results

Two sensitivity analysis runs are explored. Figs. 7 and 8 describe the results of SA1, and Figs. 9 and 10 describe the outcome of SA2. Table 3 provides numerical detailed solutions for SA1 and SA2.

In SA1 a sensor network with 10 sensors was considered for each of the systems. It can be seen from Figs. 7 and 8 that the Pareto fronts generated with a system of 10 sensors dominated the Pareto fronts established with five sensors.

In SA2 two types of modifications to the sensor detection ability compared to the base run are explored: (1) setting a minimum detection concentration of 1 mg/L [i.e., a sensor will detect a contaminant only if its concentration is above 1 mg/L (zero at the base run)] (denoted as SA2.1), and (2) taking into consideration a sensor malfunction probability of 0.1 (i.e., out of order of 10% of the sensors) (SA2.2). In SA2.2 malfunctioning sensors are selected randomly using a uniform probability distribution. This evaluation is thus an expectation over all random malfunctions. For both cases Pareto optimal fronts are plotted and selected sensor solution locations are presented, as described in Figs. 9 and 10.

It can be seen from Figs. 9 and 10 that for both a reduction in the sensor detection ability (SA2.1) or taking into consideration a

Table 3. Sensitivity Analysis Selected Solutions

<table>
<thead>
<tr>
<th>Solution ID</th>
<th>Node number</th>
<th>Detection likelihood ($f_1$)</th>
<th>Expected detection time ($f_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1</td>
<td>15, 35, 61, 131, 151, 166, 171, 225, 239, 255</td>
<td>91</td>
<td>98</td>
</tr>
<tr>
<td>2.2.1</td>
<td>43, 90, 93, 250, 453, 477, 600, 672, 701, 746</td>
<td>68</td>
<td>97</td>
</tr>
<tr>
<td>3.1.1</td>
<td>40, 50, 141, 179, 217</td>
<td>51</td>
<td>172</td>
</tr>
<tr>
<td>3.1.2</td>
<td>40, 50, 141, 179, 239</td>
<td>51</td>
<td>139</td>
</tr>
<tr>
<td>3.2.1</td>
<td>250, 428, 664, 701, 1305</td>
<td>36</td>
<td>300</td>
</tr>
<tr>
<td>3.2.2</td>
<td>250, 288, 425, 653, 1305</td>
<td>34</td>
<td>148</td>
</tr>
</tbody>
</table>

See Figs. 7–10.

Table 3. Sensitivity Analysis Selected Solutions

<table>
<thead>
<tr>
<th>Solution ID</th>
<th>System</th>
<th>Analysis</th>
<th>Objective function</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood</td>
</tr>
<tr>
<td>1.1.2</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ max $f_2$</td>
<td>best sensor detection redundancy</td>
</tr>
<tr>
<td>1.1.3</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ min $f_2$</td>
<td>a possible compromised solution</td>
</tr>
<tr>
<td>1.1.4</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood</td>
</tr>
<tr>
<td>1.2.1</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ min $f_2$</td>
<td>best expected time of detection</td>
</tr>
<tr>
<td>1.2.2</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ max $f_2$</td>
<td>best sensor detection redundancy</td>
</tr>
<tr>
<td>1.2.3</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ min $f_2$</td>
<td>a possible compromised solution</td>
</tr>
<tr>
<td>1.2.4</td>
<td>1</td>
<td>BR</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for ten sensors</td>
</tr>
<tr>
<td>2.1.1</td>
<td>2</td>
<td>SA1</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for imperfect sensors</td>
</tr>
<tr>
<td>2.2.1</td>
<td>2</td>
<td>SA2.1</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for imperfect sensors</td>
</tr>
<tr>
<td>3.1.1</td>
<td>1</td>
<td>SA2.1</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for ten sensors</td>
</tr>
<tr>
<td>3.1.2</td>
<td>1</td>
<td>SA2.2</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for imperfect sensors</td>
</tr>
<tr>
<td>3.2.1</td>
<td>2</td>
<td>SA2.1</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for imperfect sensors</td>
</tr>
<tr>
<td>3.2.2</td>
<td>2</td>
<td>SA2.2</td>
<td>max $f_1$ min $f_2$</td>
<td>best detection likelihood for imperfect sensors</td>
</tr>
</tbody>
</table>

Legend: 1 = EPANET Example 3, 2 = Richmond Water System, $f_1 = $ sensor detection likelihood, $f_1 = $ sensor detection redundancy, $f_2 = $ sensor expected detection time, min = minimization, max = maximization, BR = base run, SA1 = sensitivity analysis 1, SA2 = sensitivity analysis 2 partitioned into SA2.1, and SA2.2

Fig. 11. Summary of solution characteristics
sensor malfunction probability (SA2.2), results in Pareto fronts that are dominated by the perfect sensor case (i.e., the base run).

Fig. 11 summarizes the solution characteristics obtained for the base run and sensitivity analysis for both examples.

Conclusions

Real world problems require the simultaneous optimization of multiple, possibly conflicting, objectives that should be optimized simultaneously. In such problems there is no single optimal solution, but rather a set of compromised solutions known as non-dominated, or Pareto optimal. This non-dominated set describes the tradeoffs among different objectives, and may help the designer understand the options available for selecting a solution for implementation.

Following the September 11, 2001 event in the United States, there has been an increasing effort to enhance the security of water distribution systems through modeling of sensor networks, with the majority of attempts utilizing a single optimization objective approach. This study extended these efforts through developing and demonstrating a multiobjective scheme for sensor network design.

Sensor network detection likelihood, redundancy, and expected detection time, were selected to characterize sensor placement design objectives. Tradeoff curves using the NSGA-II ( Deb et al. 2000 ) algorithm were explored for pairs of objectives and simultaneously for all three using two example applications. A heuristic procedure was developed for sampling a small representative set of contamination events, making the employment of a multiobjective algorithm computationally feasible. The analysis provided explanatory results, and thus confirmed the methodology ability to supply a multidimensional tool for sensor placement decision making.

The model, however, has few limitations: (1) since a multiobjective evolutionary procedure is employed the computational effort needed for establishing Pareto fronts is highly intensive; (2) it is assumed that the sensor network provides real time data, where in practice sensors have a delay in transmitting detections; and (3) in reality only partial data of the system’s hydraulics are available (i.e., flows, consumptions, pressures), thus an inherent uncertainty exists in the resulted Pareto fronts.

Future research directions for extending this study can address: (1) explicit integration of the nondetected events in the estimates of the design performance metrics; (2) inclusion of uncertainty in the design objectives; (3) incorporation of nonuniform spatial probabilities of injection events; (4) consideration of more than one injection intrusion at a single contamination event; (5) development of rigorous aggregation methodologies to cope with growing system size and complexities; and (6) implementation of other multiobjective algorithms other than NSGA-II.

Acknowledgments

This research was supported by the Henri Gutwirth Promotion of Research Fund, by the Institute for Future Defense Technologies Research, named for The Medvedi, Shwartzman, and Gensler families; by the Grand Water Research Institute; and by NATO [Science for Peace (SIP) Project No. CBD.MD.SFP 981456]. The valuable reviewer’s suggestions and comments are highly acknowledged.

Notation

The following symbols are used in this paper:

\[ AN_1 = \text{average value of the geographical } x \text{ coordinate of the water distribution system; } \]
\[ AS_1 = \text{average value of the geographical } x \text{ coordinates of the evaluated sample; } \]
\[ d_r = \text{coefficient receiving the value of 1 if contamination event (i.e., an intrusion through a node of a contamination at a specific flow rate, concentration, and duration)} \]
\[ E(t_p) = \text{mathematical expectation of the minimum detection time } t_p; \]
\[ e_j(x) = \text{equality expectation constraint; } \]
\[ F(x) = [f_1(x), f_2(x), \ldots, f_M(x)]^T = \text{vector of objective functions; } \]
\[ f_i = \text{detection likelihood; } \]
\[ f_2 = \text{detection instrumentation (sensors) redundancy; } \]
\[ f_3 = \text{expected time of detection; } \]
\[ g_j(x) = \text{inequality constraint; } \]
\[ M = \text{number of objective functions; } \]
\[ N = \text{total number of system nodes; } \]
\[ q_j = \text{flows outgoing of node } j; \]
\[ R = \text{redundancy; } \]
\[ R_r = \text{sensors redundancy for the } r \text{th contamination event, receiving the value of 1 if the conditions as in Eq. (2) are met, and zero otherwise; } \]
\[ S = \text{total number of contamination events considered; } \]
\[ x^{(1)}, x^{(2)} = \text{possible two solutions; } \]
\[ x = (x_1, x_2, \ldots, x_n)^T = \text{vector of decision variables; } \]
\[ \sigma N_1 = \text{standard deviation of the geographical } x \text{ coordinates of the water distribution system (known value); } \]
\[ \sigma S_1 = \text{standard deviation of the geographical } x \text{ coordinates of the evaluated sample; and } \]
\[ \bigtriangleup = \text{better performance evaluation of an objective function. } \]

References


