Multi-Objective Optimization of Sensor Placement Water Distribution Systems

Article · January 2006

DOI: 10.1061/40941(247)105

2 authors:

Zheng Yi Wu
Bentley Systems

Tom Walski
Independent Researcher

MULTI OBJECTIVE OPTIMIZATION OF SENSOR PLACEMENT IN WATER DISTRIBUTION SYSTEMS

Zheng Yi Wu
Zheng.wu@bentley.com
Tom Walski
Tom.walski@bentley.com
Haestad Methods Solution Center
Bentley Systems, Incorporated
27 Siemon Co Dr, Suite 200W
Watertown, CT06795, USA

Abstract

Placement of water quality sensor has received an increasing concern for timely providing the warning of possible contamination in a water system. Due to the large dimension of water distribution network and the difficulty for predicting where a contamination event occurs, it is a great challenge for engineers to come up with good sensor locations with any confidence to effectively detect possible contamination events. The problem is complicated by the fact that sensor location is evaluated against a number of objective criteria that may include the detection likelihood, the expected detection time, affected population and contaminated water consumption. A design that improves one objective may deteriorate another. In this paper, sensor placement is formulated as a multi objective optimization problem that is solved by using a competent genetic algorithm while the contamination events are simulated by the latest development of Monte Carlo method.

Key Words

Water distribution system, sensor design, sensor placement, optimization, genetic algorithm, water quality

INTRODUCTION

Placing reliable and effective water sensors is of great importance for provision of secure water service and early warning system for detecting a contamination event. However, finding the best location for water quality sensors is a challenging problem because of the large number of possible contamination events and sensor locations. For a typical system, there is no guarantee that a global optimum solution can be achieved. However, using optimization, it should be possible to find better locations than those that would be selected randomly.

Over last decade, a number of methods have been developed and applied to optimizing water sensor placement. Lee and Deininger (1992) developed an integer programming optimization model to maximize the sensor coverage with minimum number of sensors. The model is based upon steady state simulation and network connectivity. The same model was solved by Kumar et al. (1997) using a heuristic-based optimization method and by Al-Zahrami and Moised (2001) using genetic algorithm.

Berry et al. (2005) proposed a sensor placement model that minimizes the contamination risk using sensors that are placed on edges or pipes between two nodes. The risk for a node is evaluated as the multiplication of population density, contamination indicator (1 for contaminated or 0 for noncontaminated node) and the probability of an attack. The overall risk is the sum of individual node risk over a number of flow patterns over an extended period of time. Thus water quality model is incorporated into the model to simulate the response of a certain attack to come up with a correct contamination indicator for a node. The problem is solved by using integer programming. Shashtri and Diwekar (2006) extended Berry's work by considering demand uncertainty.

In the meantime, Ostfeld and Salomons (2005) uncovered a number of shortcomings of Lee and Deininger's method based upon network connectivity and steady state simulation. The pollution matrix approach originally proposed by Kessler et al (1998) was extended by incorporating an EPS water quality simulation and the optimization sensor placement was solved by using a genetic algorithm (Ostfeld and Salmons 2005). This method is to construct a matrix relation to contain information of whether or not a node is contaminated by a contamination event, which can be deliberate injection at a random node during a random time period. Each node is perceived as a possible sensor location. The more contamination events are observed or recorded at one node, the greater the detection likelihood is counted, the better sensor placement the node can be. A combination of certain number of nodes represents a sensor placement solution. A genetic algorithm is applied to optimize the sensor placement under a single optimization objective of maximizing detection likelihood.

However, the sensor placement is governed by multiple criteria such as detection time, contaminated water volume and affected population. In this paper, the pollution matrix is generalized to contain more information for evaluating multi criteria for sensor allocations. The optimization problem is solved by using the competent genetic algorithm (Wu and Simpson 2001). Two benchmark design network are demonstrated for the application of the method.

DESIGN SPECIFICATION

A sensor placement consists of identifying the number of sensors and their locations in a network in order to detect accidental and deliberate contamination events in water distribution systems and thereby facilitate corrective action or public notification. An optimization model and solution algorithm is developed to locate a limited number of sensors that minimize the impact of such contamination events. The sensor design solutions are evaluated by a number of multiple objective criteria as follows.

A sensor network design will be evaluated using four quantitative design objectives including

a. Expected time of detection (Z1);
b. Expected population affected prior to detection (Z2);
c. Expected contaminated water demand prior to detection (Z3);
d. Expected likelihood of detection (Z4).

It is expected that objective Z1, Z2 and Z3 are to be minimized while Z4 is to be maximized. Calculation procedures for each of the four design objectives are given by Ostfeld et al. (2006).

In particular, the objective function Z2 is based upon the population allocated according to daily demand at a node. Computing Z2 for each design requires a significant amount of calculation to determine the affected population based upon a set of prescribed parameters. Furthermore, the affected population is proportional to the amount of polluted water, the more contaminated water volume is recorded, the more the affected population is expected. Therefore, it is believed that the objective of the affected population Z2 is coincides with the expected demand of the contaminated water Z3. Thus, only Z1, Z3 and Z4 are taken into account in this study for optimizing the location of sensor network placement.

OPTIMIZATION FORMULATION

Three objectives, including the modified detection time (Z1), expected demand of contaminated water volume (Z3) and expected detection likelihood (Z4) are considered for optimizing the sensor location. Sensors are all designed to be located at model nodes that can be junctions, tanks and reservoirs in a system. To optimize the location of the sensors is then formulated as an optimization problem.

Search for: $S = (s_1, s_2, \ldots s_n)$, $s_i \in SN$ (1)

Minimize: *Z1* (2)

Minimize: *Z3* (3)

Maximize: *Z4* (4)

Where s_i is the location of the ***i***-th sensor, *SN* is the complete set of nodes in a system. A multi objective optimization algorithm is required for solving the optimization problem.

Not all the objective functions coincide in search for the optimal sensor design. For instance, maximizing the expected detection likelihood (Z4) will favor the sensor locations that cover as many contamination events as possible (i.e. at the far end of the system), thus it may take longer time to detect the events, which deteriorate the objectives of the expected detection time (Z1) and contaminated water demand (Z3). In contrary, a sensor design that only minimizes the time of detection (Z1) (i.e. near likely contamination sources) may miss the contamination events that take more time to be detected, thus reduces the likelihood of detection (Z4). Therefore objective function Z4 appears to be conflict with Z1 and Z3. Optimizing (maximizing) the detection likelihood may not consistently minimize the objectives of the detection time as well as the contaminated demand as originally formulated.

To enable consistent optimization of all the design objectives, Z1 is reformulated as follows.

Detection Time: Z1

The detection time for a contamination event is the minimum time among all sensors that detect the event. Assuming $t_{i,j}$ is the time of sensor *i* when contamination event *j* is detected. Detection time of contamination event *j* is originally given as:

$$T_j = \min_{i=1}^{S} t_{i,j} \tag{5}$$

Where *S* is the number of sensors designed for a network. Thus expected detection time is given as:

$$Z1 = \frac{1}{M} \sum_{j=1}^{M} T_j \tag{6}$$

Where *M* is the number of randomized contamination events.

While this formulation is valid for the expected detection time for the randomized contamination, for the detection time for undetected event or successful contamination attack, it is uncertain what the detection time will be recorded. If it is recorded as zero, then minimizing the objective Z1 will definitely favor the sensor locations where contamination events are mostly likely undetected. To avoid this undesirable landscape characteristic of objective Z1, detection time for each event is reformulated as follows.

$$T_j = \begin{cases} \min\limits_{i=1}^{S} t_{i,j} & \text{If event j is detected.} \\ Htime & \text{If event j is not detected.} \end{cases} \qquad (7)$$

Where Htime is used to denote the time of an undetected event elapsed from the contamination attack to the end of the hydraulic and water quality simulation. It represents the largest time period computable and elapsed from the contamination. This way, a significant time will be recorded for the undetected event. Therefore, it enables an optimization process to search for the sensor locations that facilitate the contamination detection, that is, maximize the detection likelihood Z4 while minimizing the detection time.

The other objectives Z3 and Z4 are the same as formulated by Ostfeld et al. (2006). The expected demand of contaminated water volume is calculated and summed up over time when the concentration is greater than a prescribed concentration level. The longer time is taken to detect a contamination event, the more contaminate water volume is expected to be accumulated. The more events are detected, the less contaminated water is to occur. Thus minimizing Z3 is consistent with maximizing the detection likelihood Z4.

A unified objective is formulated by optimizing just a combined objective.

$$Z = \frac{Z1}{Htime} + \frac{Z2}{Ptotal} + \frac{Z3}{Vmax} + \frac{1}{Z4} \qquad (8)$$

Where *Vmax* is the maximum contaminated demand of water volume over the simulation period, *Ptotal* is the equivalent total population for a system. This unifies three objectives Z1, Z2, Z3 and Z4. Minimizing Z will equivalently minimize Z1, Z2, Z3, and maximize Z4.

The multi objective optimization problem (1) – (4) can be equivalently reformulated as below.

Search for: $S = (s_1, s_2, \ldots s_n)$, $s_i \in SN$ (9)

Minimize: Z (10)

This transforms the multi-objective optimization problem into a single objective optimization problem, a competent genetic algorithm (Wu and Simpson 2001) is applied to solving for the network sensor design.

SOLUTION METHODOLOGY

The problem is solved in two phases. Phase one is to construct a database of contamination responses of randomized contamination events. The second phase is to optimize the sensor location based upon the contamination response database.

Contamination Response Database

Randomized pollution events are generated by a Monte Carlo method. Each of pollution events is represented by a prescribed number of contamination injection locations and also the start time when the injection is taking place. The contamination pollution is simulated by mass booster water quality model. The pollution concentration is analyzed throughout the network and recorded in a contamination response database that includes the following response attributes for each node:

i. Contamination status, whether or not a node is contaminated;
ii. Time of concentration reaching the detectable level;
iii. Elapsed time until it is detected;
iv. Time of concentration reaching a harmful threshold to customer.

For a given contamination event *i*, the simulated response at node *j* is noted as:

$$p_{i,j} = (cs_{i,j}, tc_{i,j}, td_{i,j}, th_{i,j}) \tag{11}$$

Where $cs_{i,j}$ is the contamination status, taking a binary value of 1 (contaminated) or 0 (not contaminated), $tc_{i,j}$ is the time in seconds that the simulated pollution reaches the minimum concentration level that can be detected by a sensor instrument, $td_{i,j}$ is the time elapsed of a contamination event until it is detected by at least one sensor, and $th_{i,j}$ is the calculated pollution concentration that is perceived as harmful to customers. Therefore, a contamination response database or matrix can be noted as:

$$P = (p_{i,j}) \qquad where\ i = 1, 2, \ldots, N;\ j = 1, 2, \ldots M. \tag{12}$$

Where *N* is the number of nodes and *M* is the number of pollution events. The more contamination events are simulated and recorded in the database, the better representation the database is for evaluating the sensor design, the longer it takes to generate the response database and sensor design as well. The contamination response database can be reused many times as desired once it is generated.

Sensor Placement Optimization

A sensor design solution is to determine where the possible sensors can be located while the number of sensors is predetermined for each design. The design is evaluated by the criteria (Z1, Z3, Z4 and Z) as defined above. A competent genetic algorithm is applied to optimize the sensor location. All the nodes are considered as possible sensor locations. Therefore, the GA is effectively expected to search for a set of nodes as sensor locations. Each sensor substring is created to encode all the candidates of possible sensor locations.

The GA optimization of sensor design can be undertaken as separately after a contamination response database is constructed. Sensor design solutions and discussion are presented for two challenge networks.

DESIGN EXAMPLES

Two challenge sensor design examples are presented to test the solution method. Each network example is tested for one base scenario and three derived scenarios, which are outlined in Table 1 below.

Each injection event is simulated as the water quality source of mass booster type at a node that is randomly selected. To ensure the contamination to travel throughout the system and to be observed at nodes, the water quality analysis was conducted over 96 hours with the time step of 5 minutes.

Sensor design solutions are optimized for both challenge networks under 4 scenarios.

Table 1 Example sensor design scenarios

Base Case A			Derived Cases		
Component	Characteristics	Description	Case B	Case C	Case D
System	Injection location	Nodes			
	Existing sensors	None			
	Demand	Deterministic			
Injection	Probability	Even			
	No. of injections	1			2
	Duration	2 hours	10 hours		
	Flow	125 L/h			
	Concentration	23000 mg/L			
	Constituent Type	Conservative			
Detection	Time delay	0 hour		3 hours	
	Sensitivity	Ideal (above zero)			
Design	Sensor number	5 ; 20			

Network 1

This sample network is relatively small system, consisted of about 130 nodes, one reservoir, one pump station and one tank. Optimal sensor placement runs have been conducted for the scenarios in Table 1. Using the competent GA optimization (Wu and Simpson 2001), all the nodes are considered as possible sensor locations. For each design scenario, a pollution matrix, as given as Eq.(11) and (12), is generated by simulating a total of 300 random contamination events. The optimal solutions are given in Table 2 and Table 3 respectively for 5-sensor designs and 20-sensor designs.

Table 2 Network 1optimal design solutions of five-sensor locations for different scenarios

	Optimal Solutions of Five-Sensor Placement for Network 1			
Sensors	N1A5*	N1B5	N1C5	N1D5
1	45**	17	17	68
2	68	46	45	83
3	83	83	83	101
4	100	101	101	118
5	118	126	126	122
Detection Likelihood	0.82	0.82	0.82	0.95
Detection Time (hours)	21.71	22.11	25.08	9.47
Polluted Water (gallons)	41822	39671	29648	34567
Unified Fitness	1.87	1.84	1.78	1.50

*The notation means the solution of 5 sensor placement case A for network 1.
**45 means JUNCTION-45.

As shown in Table 2 and 3, the solutions for both 5-sensor and 20 sensor design are slightly different from one scenario to another. For 5-sensor design, only one location JUNCTION-83 is selected as one of sensor location for all four design scenarios, the other sensor nodes are different but they are fairly close each other. For instance, N1A5 scenario favors JUNTION-100 while N1B5 selects JUNCTION-101, both junctions are adjacent nodes, so are the JUNCTION-118 and JUNCTION-126, as shown in Figure 1. Due to the variation of the designed injection events, the design solutions are selected differently. However, the optimized solutions as annotated in Figure 1 will be a good set of sensor locations for further tuning the final design.

Table 3 Network 1 optimal design solutions of twenty-sensor locations for different scenarios

	Optimal Solutions of Twenty-Sensor Placement for Network 1			
Sensors	N1A20	N1B20	N1C20	N1D20
1	10	10	10	4
2	12	12	11	11
3	19	19	19	14
4	21	21	21	19
5	34	34	34	21
6	35	35	35	34
7	40	45	40	35
8	45	68	45	40
9	68	74	68	45
10	75	76	74	68
11	80	83	80	74
12	83	89	83	83
13	98	100	100	90
14	100	101	101	98
15	102	102	114	100
16	114	114	118	102
17	118	118	122	117
18	123	123	123	123
19	124	124	124	124
20	126	126	126	126
Detection Likelihood	0.92	0.94	0.93	0.99
Detection Time (hours)	10.86	9.01	16.09	2.8
Polluted Water (gallons)	13144	14276	14647	15180
Unified Fitness	1.34	1.31	1.39	1.19

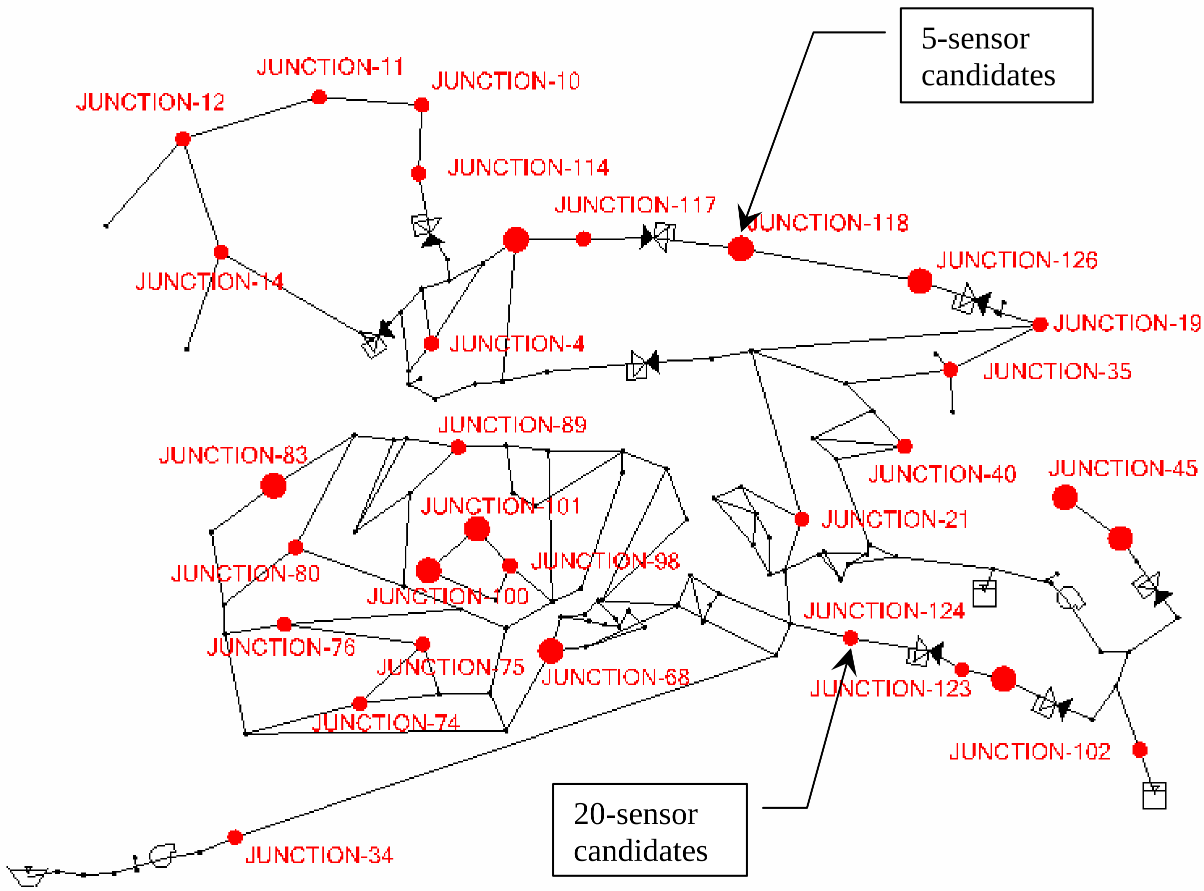

Figure 1 Optimal sensor locations for Network 1

Network 2

This network consists of 14822 pipes, 12523 nodes, 2 tanks, 4 pumps and 2 reservoirs. Considering every node as a possible sensor placement location will results in astronomically large number of possible sensor placement solutions.

In general, to allocate k sensors within a network of n nodes, the total number of possible sensor placement combinations is given as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{13}$$

There are 12523 nodes in the model. If every node is considered a possible sensor location, it would give a total number of $2.55\text{x}10^{18}$ and $3.64\text{x}10^{63}$ for 5 and 20 sensor placement combinations respectively. However, in real system, it is not possible to access every node location and use it as sensor placement location. A careful selection of possible sensor locations will be extremely helpful to reduce the sensor placement solution space.

Merging the pipes in series will also help to reduce the number of nodes that are connected the pipes of same physical parameters. To maintain the equivalent hydraulic characteristics, the network is simplified by merging the series pipes of the same diameters and materials, while maintaining the network geometry and assigning the demand to the adjacent nodes (Bentley 2006). It helps to reduce the number of nodes from 12523 to 7789, and consequently reduce the number of sensor allocation combinations from 2.55×10^{18} to 2.38×10^{17} for 5 sensor placement, and from 3.64×10^{63} to 2.71×10^{59} for 20 sensor placement. The reduction of the solution space represents more than 90% of the original solution combinations. The sensor placement optimization is conducted by applying the competent GA-based approach as described above. The pollution matrix is generated via simulating 15,000 randomized events for the remaining set of junctions that are also considered as possible sensor locations.

The optimal sensor placement solutions are given as in Table 4 and 5. The solutions for design scenario A are annotated in Figure 2.

Table 4 Network 2 optimal solutions for five-sensor placement of different scenarios

	Optimal Solutions of Five-Sensor Placement for Network 2			
Sensors	N2A5	N2B5	N2C5	N2D5
1	3709	4654	1486	4654
2	4957	4957	4664	4957
3	6583	6734	4957	6583
4	8357	8578	7953	7953
5	9364	9364	9364	9364
Detection Likelihood	0.25	0.25	0.20	0.38
Detection Time (hours)	37.70	37.59	38.06	32.09
Polluted Water (gallons)	394799	347642	379106	577777
Unified Fitness	4.46	4.45	4.55	2.98

CONCLUSIONS

Sensor placement approach developed in this paper has been demonstrated for optimizing the sensor locations for two benchmark examples. Although truly optimal sensor locations are unlikely justified by the optimization model alone due to the randomness of the deliberate contamination injections and multiple evaluation criteria, the solutions produced by the optimization will serve as a good basis for further fine tuning the final sensor locations.

REFERENCES

Al-Zahrani, M. A. and Moied. K. (2001). "Locating optimal water quality monitoring stations in water distribution system" *Bridging the gap: Meeting the world water and environmental challenges, State of practice – Proceeding of World Water and Environmental Resources Congress.* (CD-ROM), ASCE, Reston, Va.

Bentley Systems, Incorporated, *WaterGEMS XM v8 User Manual,* Haestad Method Solution Center, 27 Siemon Co Dr., Suite200W, Watertown, CT06795, USA.

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

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

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

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

Ostfeld A. and Salomons E. (2005) "Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security", *J. Water Resour. Plan. Manage.*, 130(5), 377-385.

Ostfeld A., Uber J. and Salomons E. (2006) "Battle of the Water Sensor Networks: A summary of Contributed Design", *Proceeding of the Conference on Water Distribution Systems Analysis, Cincinnati, Ohio, U.S.A.*

Shastri Y. and Diwekar U. (2006) "Sensor Placement in water Networks: A Stochastic Programming Approach", *J. Water Resour. Plan. Manage.*, 132(3), 192-203.

Wu, Z. Y. and Simpson A. R. (2001) "Competent Genetic Algorithm Optimization of Water Distribution Systems." *J. Computing in Civil Engineering*, 15(2), 89-101.

Table 5 Network 2 optimal design solutions of twenty-sensor locations for different scenarios

	Optimal design solutions of twenty-sensors			
Sensors	N2A20	N2B20	N2C20	N2D20
1	871	623	871	872
2	1334	872	1081	1081
3	2589	1081	2313	1422
4	3115	1486	3115	2595
5	3640	2313	3322	3115
6	3719	3115	3640	3318
7	4247	3640	3719	3640
8	4990	3718	3780	3782
9	5630	4375	4209	4375
10	6733	4621	4990	4990
11	7442	4990	5630	5630
12	7714	5481	6737	6734
13	8387	7443	7442	7442
14	8394	7711	7908	7713
15	9778	8294	8408	8394
16	10290	8392	9364	9778
17	10522	9778	9779	10522
18	10680	10190	10494	10680
19	11151	10522	10680	11151
20	11519	10680	12389	12389
Detection Likelihood	0.36	0.35	0.35	0.53
Detection Time (hours)	32.53	32.70	32.82	24.90
Polluted Water (gallons)	232482	212378	218892	358978
Unified Fitness	3.15	3.16	3.20	2.15

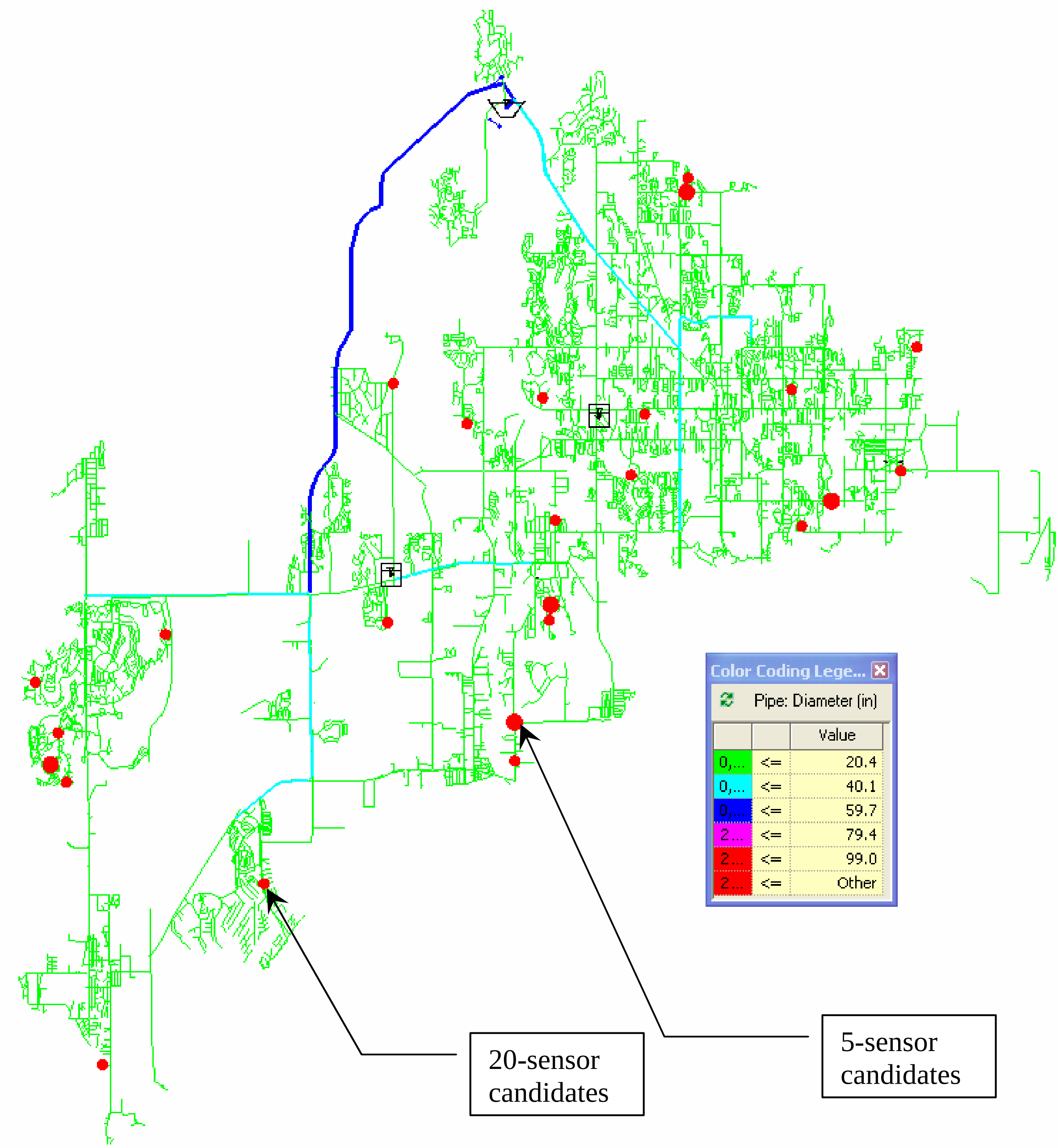

Figure 2 Optimal Sensor Placement for Design Scenario A of Network 2