Identifying Sets of Key Nodes for Placing Sensors in Dynamic Water Distribution Networks

Article *in* Journal of Water Resources Planning and Management · July 2008

Impact Factor: 2.68 · DOI: 10.1061/(ASCE)0733-9496(2008)134:4(378)

5 authors, including:

Paul Fischbeck
Carnegie Mellon University

Mitchell J. Small
Carnegie Mellon University

Jeanne Vanbriesen
Carnegie Mellon University

Elizabeth Casman
Carnegie Mellon University

Identifying Sets of Key Nodes for Placing Sensors in Dynamic Water Distribution Networks

Jianhua Xu, M.ASCE[1]; Paul S. Fischbeck[2]; Mitchell J. Small, M.ASCE[3]; Jeanne M. VanBriesen, M.ASCE[4]; and Elizabeth Casman[5]

Abstract: The design of a sensor-placement scheme capable of detecting all possible contamination events for a water distribution system before consumers are put at risk is essentially impossible given current technologies and budgets. It is, however, possible to design sensor-placement schemes that optimize related objectives (e.g., minimize expected volume of contaminated water consumed prior to detection), but this requires the availability of hydraulic and water quality models for the distribution network and significant computational power, which are the main obstacles to the identification of optimal sensor locations. This paper describes a different approach that reduces the problem's complexity by expressing a water distribution system as different graphs based on the information readily available from most, if not all, water utilities. The approach provides critical policy and decision support for utilities when hydraulic and water quality models are not available and/or when simulation-based techniques are computationally infeasible.

DOI: 10.1061/(ASCE)0733-9496(2008)134:4(378)

CE Database subject headings: Water distribution systems; Network analysis; Security; Water quality; Water pollution.

Introduction

Significant research has taken place in recent years to support the placement of sensors in water distribution systems to detect accidental or intentional contamination events (Kessler et al. 1998; Ostfeld and Salomons 2004; Berry et al. 2005, 2006; Grayman et al. 2006; Gueli 2006; Propato 2006; Shastri and Diwekar 2006). Placing sensors optimally requires an understanding of how water flows and contaminants behave in water distribution systems. These behaviors can be approximated by using a simulation-based analysis with calibrated hydraulic and water quality models for the system. However, many water utilities do not have water quality models for their systems because of the significant calibration requirements needed to build this type of model. Even if a calibrated water quality model does exist, the computational requirements for computing the contaminant concentrations resulting from all possible random contamination events are daunting. Without prior knowledge, a contamination event must be assumed to be possible at any time, any location, and for any duration, thus, requiring the simulation of a large number of contamination scenarios. As an example, using 30 parallel processors, it took 8 days to simulate random contamination events that could occur at 5 min intervals over a 24 h period from any of the 12,527 nodes in a modest-sized distribution network (Krause et al. 2006). The placement of sensors based solely on hydraulic models negates the necessity of extracting contaminant behavior, but still requires the formulation of optimization models to determine where to place sensors (Lee and Deininger 1992; Kessler et al. 1998; Berry et al. 2005). Solving these mathematical models to optimality for large networks is itself a difficult problem given that most facility location models for a general network are NP hard (Drezner and Hamacher 2002), which means that no polynomial-time algorithms for these problems have been found yet. These are the obstacles that must be faced when working on real-world problems.

In this work, we simplify the problem by applying a graph-theoretic (network analysis) approach for placing sensors in a water distribution system, eliminating the need for calibrated water quality models, and alleviating the computational requirements. The paper is outlined as follows: (1) we explain how network analysis can be adapted to work with water distribution networks; (2) we demonstrate how key sets of nodes can be identified for sensor placement based on network analysis; and (3) we evaluate the methods by comparing the performance of our network analysis-based sensor-placement schemes with that of simulation-based optimal sensor-placement schemes for an illustrative water distribution network.

[1]Graduate Research Assistant, Dept. of Engineering and Public Policy, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: xujianhua.cmu@gmail.com

[2]Professor, Dept. of Social and Decision Sciences, and Engineering and Public Policy, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: pf12@andrew.cmu.edu

[3]Professor, Dept. of Civil and Environmental Engineering, and Engineering and Public Policy, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: ms35@andrew.cmu.edu

[4]Associate Professor, Dept. of Civil and Environmental Engineering, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: jeanne@cmu.edu

[5]Associate Research Professor, Dept. of Engineering and Public Policy, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: casman@andrew.cmu.edu

Note. Discussion open until December 1, 2008. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on December 21, 2006; approved on October 18, 2007. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 134, No. 4, July 1, 2008. ©ASCE, ISSN 0733-9496/2008/4-378–385/$25.00.

Methods

Water distribution systems are physical networks where *nodes* represent sources, tanks, junctions (the connections between pipes), and points of water withdrawal, while *edges* represent pipes, valves, and pumps. In a graphical notation $G=(V,E)$, $V=v_1,\ldots,v_n$=node set, and $E=e_1,\ldots,e_m$=edge set. Network characteristics of water distribution systems have been exploited either explicitly or implicitly to perform hydraulic analysis (Todini and Pilati 1987), reliability analysis (Su et al. 1987; Wagner et al. 1988a,b), and sensor placement (Kessler et al. 1998; Berry et al. 2005; Shastri and Diwekar 2006; Watson et al. 2004). In our analysis, three types of graphs (undirected graphs, dynamic directed graphs, and weighted dynamic directed graphs) are used. An *undirected graph* represents the physical structure of a water distribution system. A *dynamic directed graph* represents the compilation of a series of snapshots of a water distribution system, with edge direction being the water flow direction. If the water flow direction changes over time in a pipe, then in the dynamic directed graph the pipe is represented as two edges with opposite direction between the pair of nodes incident with the edge. A *weighted dynamic directed graph* is defined based on the dynamic directed graph with the value of an edge being water travel time between the pair of nodes incident with the edge. Thus, the construction of an *undirected graph* requires only information on the physical structure of a distribution system, while the latter two types of graphs require hydraulic information for a distribution system.

Graph theory and network analysis have shed light on the properties of a variety of networks: The internet backbone (Faloutsos et al. 1999), electricity grids (Blumsack 2006), transportation networks (Banavar et al. 2000), urban streets (Crucitti et al. 2006a,b), and social networks (Wasserman and Faust 1994). In the context of water distribution systems, we are interested in knowing the structurally important nodes, which might have implications on where sensors should be placed. The following is the description of the network measures (betweenness centrality and receivability) that we adopted and how sensors are placed based on them.

Betweenness centrality defines the centrality of a node in terms of the degree to which the node falls on the shortest path between other pairs of nodes. If a node has a high betweenness centrality, then it lies on the path of many pairs of nodes. Eq. (1) quantifies the betweenness centrality for a node in an undirected graph (Freeman 1977). Gould (1987) proved that the same equation holds for directed graphs

$$C_i^B=\left[\sum_{i,j,k\in G,j\neq k\neq i} n_{jk}(i)/n_{jk}\right]\Big/[(N-1)(N-2)] \tag{1}$$

where C_i^B=betweenness centrality of node i; N=total number of nodes in the network; i, j, and k=indices of nodes; n_{jk}=number of geodesics (the shortest path between two nodes) linking node j and node k; and $n_{jk}(i)$=number of geodesics linking node j and node k that contain i. Fig. 1 depicts a hypothetical water distribution network and the betweenness centrality of each node. Nodes N3, N2, N4, N5, and N7 have the highest betweenness centrality, with the highest listed first.

In a water distribution network, a node with high betweenness centrality would be between many potential upstream contamination events and downstream receptor populations. Nodes with high betweenness centrality are potential locations for sensors. However, a systematic analysis of the betweenness centrality across a set of water distribution networks shows that nodes with the highest betweenness centrality tend to cluster. Fig. 2 shows such an example with Network 1, which is based on a real network and can be found in Ostfeld et al. (2006). This means that placing sensors in a water distribution network based only on the ranks of betweenness centrality might give substantial suboptimal solutions, as clustered sensors would likely provide redundant information. As shown by Borgatti (2006), multiple nodes, each with a high centrality value, might not provide the highest centrality value collectively. We improved the betweenness-centrality only-based sensor-placement scheme with two measures: (1) to widen the graph distance (the number of edges along the shortest path between two nodes) between nodes with high betweenness centrality, we divide a water distribution system into a set of exhaustive and mutually exclusive communities and then identify the node with the highest betweenness centrality within each community; and (2) to increase the detection likelihood, we force the sensor locations to be biased toward the downstream nodes.

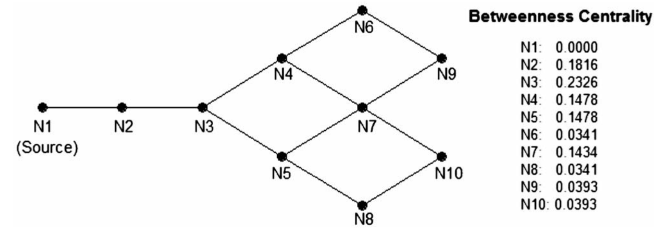

Fig. 1. Illustration of betweenness centrality with a hypothetical water distribution network

The procedure for identifying a set of key nodes for placing sensors based on betweenness centrality, community structure, and graph distance is as follows.

1. An *undirected graph* is used to express a given water distribution system. An adjacency matrix ($N\times N$) is created based on the physical structure of the system, with N being the number of nodes in the network. In the adjacency matrix cell, if there is an edge between node i and node j, then the cell value c_{ij} is equal to 1, and if there is no edge between node i and node j, then the cell value c_{ij} is equal to 0.

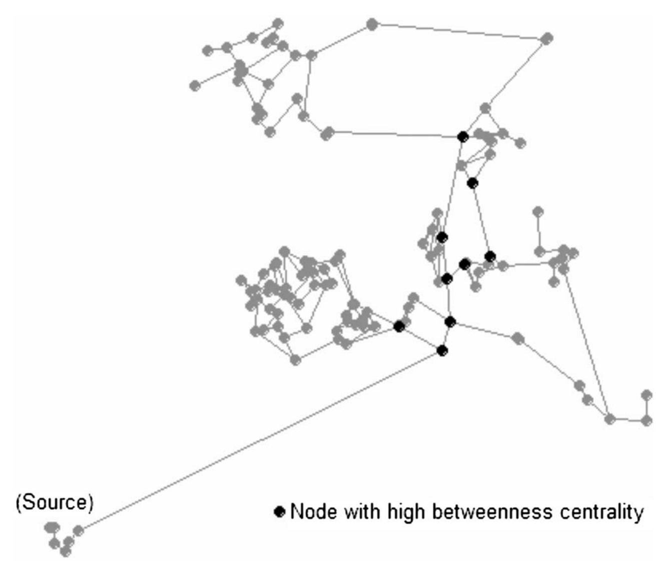

Fig. 2. Nodes with high betweenness centrality in Network 1

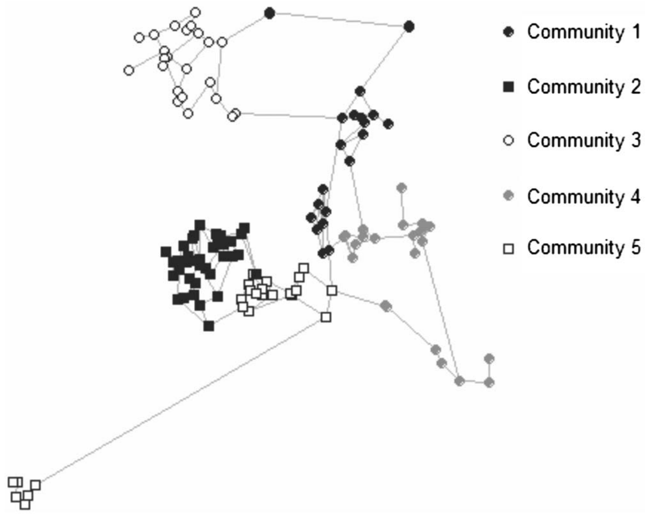

Fig. 3. Five communities detected for Network 1 using the Newman algorithm

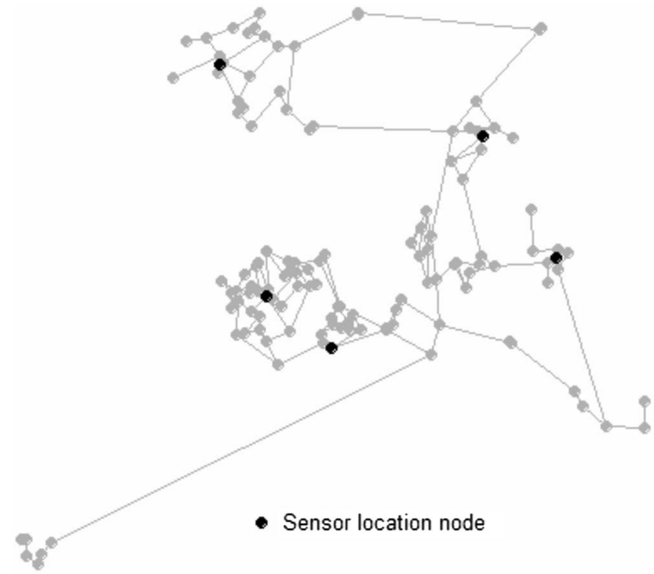

Fig. 4. Five sensor locations identified for Network 1 based on betweenness centrality, community structures, and graph distance

2. For each node in the network, the graph distance between the node and each water source is calculated. From these, the shortest distance is chosen. Various graph search algorithms can be used to do such a calculation (Cormen et al. 1997). In our research, we used a breadth-first search algorithm. For example, the graph distance between node N1 (water source) and node N10 in the hypothetical network shown in Fig. 1 is 5.
3. The number of communities to be detected is set to be equal to the number of sensors to be placed. The Newman algorithm (Newman 2004) is used to identify communities within the water distribution network. The rationale of the Newman algorithm is that the connections of nodes within a community are dense and the connections of nodes between communities are sparse. Water distribution systems reflect the community structure of the real world. Transmission mains transport water from the water source (e.g., treatment plant or reservoir) to municipalities/communities. Blocks in communities are connected to water mains. As a result, within communities, node connections are denser, while between communities node connections are sparser. For example, Fig. 3 shows the five communities identified for Network 1 with the Newman algorithm.
4. Within each community, a node with a high betweenness centrality and a long graph distance from water sources is chosen as the potential location. The trade-off between graph distance and betweenness centrality is left for decision makers.

We used a software (named ORA) developed by Carley and Reno (2006) to detect community structures, calculate the betweenness centrality of each node, and determine the shortest graph distance from each node to water sources. When we do the trade-off between the betweenness centrality and the shortest graph distance for each community, we select the five nodes with the highest betweenness centrality and then from those nodes, we choose the node with the greatest shortest path length value. The five sensor locations identified for Network 1 using this procedure are shown in Fig. 4.

Receivability is a concept that we formulate in this study to describe the set and number of nodes that have paths to the measured node in a graph. The concept is derived from the well-known concept of *reachability*. If there are one or more paths from node i to node j, then node j is *reachable* from node i and node i is receivable to node j. In our context, node i is the node where a contamination event originates, and node j is the node that the contaminant might reach. Receivability can be determined using a comprehensive set of contamination simulations with water quality models. However, network analysis provides a more efficient way. Receivability can be calculated for each node in a graph (of any type) by using graph search algorithms (Cormen et al. 1997). However, in our context, receivability for nodes in an undirected graph is trivial. The receivabilities of all nodes are the same in number for a connected undirected graph since each pair of nodes in such a graph can reach each other. The receivabilities for nodes in dynamic directed graphs and weighted dynamic directed graphs are particularly of interest. Based on a dynamic directed graph, we can get the set of nodes that if contaminated, could reach the studied node eventually if no intervention is implemented. Receivability calculated from a dynamic directed graph is called nontime-constrained receivability. Fig. 5 shows the sets of nodes that could reach Node 8 (a randomly-chosen node) in Network 1 (i.e., the receivability of Node 8 under nontime-constrained condition). Note that in Fig. 5 water is pumped into the system in the lower left-hand portion of the network, enters the main part of the network in the lower-central portion of the graph, and flows outwards to consumers from there. Based on a weighted dynamic directed graph, we can determine the set of nodes that if contaminated, could reach the studied node within a certain period of time. Receivability calculated from a weighted dynamic directed graph is called time-constrained receivability.

Both nontime-constrained receivability and time-constrained receivability measure the capability of a node with a sensor to detect contamination events. Sensors placed at nodes with higher receivability are expected to detect more contamination events, assuming that every node has an equal chance of being the source of the contamination. With this assumption, maximizing the detection likelihood is equivalent to maximizing the *coverage* of the sensors (i.e., the percentage of the nodes which, if contaminated, could be detected by the sensors placed in the system). The set of

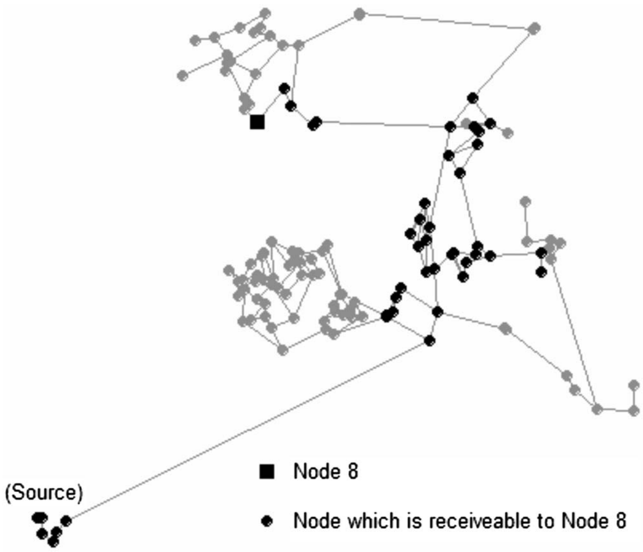

Fig. 5. Receivability of Node 8 under the nontime-constrained condition

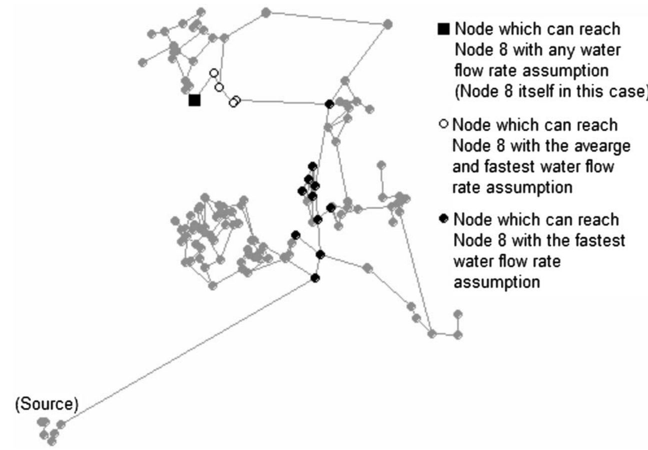

Fig. 6. Receivability for Node 8 under the time-constrained condition (150 min) with different water flow rate assumptions

nodes that collectively have the highest receivability would maximize the coverage. A greedy heuristic algorithm is used to find the optimal solution for sensor placement. Nemhauser et al. (1978) prove that at least 63% of the optimal value is guaranteed by using a greedy algorithm for submodular set functions. Krause et al. (2006) show that for submodular set functions for water distribution network structures, the greedy algorithm finds the optimal solutions in 99% of the cases.

The procedure for identifying a set of key nodes with nontime-constrained receivability is as follows:

1. A *dynamic directed graph* is used to describe a given water distribution system. An adjacency matrix ($N \times N$) is created based on the physical structure of the system, with N being the number of nodes in the network. In the adjacency matrix cell, if there is a pipe connection between node i and node j and water flows from node i to node j sometimes during the studied time period, then the cell value c_{ij} is set as 1; otherwise, the cell value c_{ij} is set as 0.
2. Based on the adjacency matrix, the receiveability of each node is calculated by using a breadth-first search algorithm (Cormen et al. 1997).
3. Following the greedy heuristic algorithm, the first sensor is placed at the node with the highest nontime-constrained receivability. The nodes covered by the first sensor are removed, and the second sensor is placed at the node with the highest nontime-constrained receivability among the remaining nodes. This continues until the nth sensor is placed or until all the nodes in the water distribution system are covered.

The procedure for identifying a set of key nodes for placing sensors based on time-constrained receivability is similar except for how the adjacency matrix is populated. The cell value c_{ij} in this case becomes water travel time instead of a binary value of 1 or 0. To derive the water travel time in each of the pipes, the water flow rate is required. In a dynamic water distribution system, water flow rates change with time. We record three flow rates for each pipe to derive a water travel time: The slowest flow rate, the average flow rate, and the fastest flow rate, which correspond to the longest, average, and shortest water travel time, respectively. The water travel time is then assigned as the edge value of the weighted dynamic directed network. Fig. 6 shows the time-constrained receivability of Node 8 under the three water flow rate assumptions.

The sensor location results for Network 1 under the two receivability conditions are shown in Table 1 with the columns for sensor location (Node ID), the number of nodes covered by each sensor, and the cumulative percentage of the nodes covered, respectively. As shown, based on the nontime-constrained receivability, the first five sensors cover 91% of the nodes in the network and 14 sensors are sufficient to cover the whole network. Based on the time-constrained receivability, 35 sensors are needed to cover the whole network when requiring that contamination events be detected within 10 h, assuming the average flow rate in the system. Table 1 shows the first 20 sensors to be placed. As expected, under the time-constrained condition, more sensors are needed to achieve the same coverage as with the nontime-constrained condition.

Comparison of Sensor-Placement Schemes

We evaluated the performance of sensor schemes derived using the graph-theoretic approach on four criteria: (1) the expected time to detect a contamination event; (2) the expected population at risk prior to detection; (3) the expected volume of water contaminated prior to detection; and (4) the detection likelihood. The choice of these four criteria is motivated by the ultimate purposes of placing sensors in a water distribution network, which are to (1) detect contamination events as quickly as possible to allow for a timely response; (2) capture as many contamination events as possible; and (3) protect as large a population as possible.

A given sensor-placement scheme is evaluated against each of the criteria with an exhaustive simulation of contamination scenarios. A contamination scenario is an event occurring at a given node and starting at a given point in time with the input of contamination lasting for a given duration. If contamination scenario i can be detected by any of the sensors, d_i (a binary indicator) is recorded as 1, the shortest time for contaminant traveling from the contamination origin to any of the sensors is recorded as t_i, the population at risk within t_i is calculated as p_i, and the volume of contaminated water within t_i is denoted as v_i. Readers are referred to Ostfeld et al. (2006) for a detailed description on how p_i and v_i are calculated. If contamination scenario i cannot be detected by

Table 1. Sensor Placement Using a Greedy Heuristics Algorithm for Network 1 Based on Nontime-Constrained Receivability and Time-Constrained Receivability

*i*th sensor	Nontime-constrained receivability: Location ID	Nontime-constrained receivability: Number of nodes covered	Nontime-constrained receivability: Cumulative percentage (%)	Time-constrained receivability (10 h): Location ID	Time-constrained receivability (10 h): Number of nodes covered	Time-constrained receivability (10 h): Cumulative percentage (%)
1st	83	81	63	118	39	30
2nd	126	16	75	122	17	43
3rd	45	10	83	68	10	51
4th	114	7	88	101	9	58
5th	100	4	91	83	8	64
6th	123	3	94	89	4	67
7th	125	1	95	85	3	70
8th	124	1	95	75	3	72
9th	113	1	96	72	3	74
10th	38	1	97	46	3	77
11th	36	1	98	14	3	79
12th	13	1	98	123	2	81
13th	7	1	99	76	2	82
14th	131	1	100	52	2	84
15th	—	—	—	126	1	84
16th	—	—	—	125	1	85
17th	—	—	—	124	1	86
18th	—	—	—	117	1	87
19th	—	—	—	116	1	88
20th	—	—	—	114	1	88

any of the sensors, d_i is recorded as 0, and the detection time t_i is recorded as a large value D (e.g., simulation duration), the population at risk p_i and the volume of contaminated water v_i are calculated accordingly (e.g., within D). The expected time to detect contamination events (Z_1), the expected population at risk prior to detection (Z_2), the expected volume of contaminated water consumed prior to detection (Z_3), and the detection likelihood (Z_4) are calculated using the following equations by assuming that the total number of the contamination scenarios is S

$$Z_1 = \left(\sum_{i=1}^{S} t_i \right) \Big/ S \quad Z_2 = \left(\sum_{i=1}^{S} p_i \right) \Big/ S$$

$$Z_3 = \left(\sum_{i=1}^{S} v_i \right) \Big/ S \quad Z_4 = \left(\sum_{i=1}^{S} d_i \right) \Big/ S$$

Even with the absolute values calculated for each of the criteria, we still cannot tell how good the performance of a sensor-placement scheme is without a benchmark. In our research, we use the performance of the optimal results obtained from a simulation-based analysis as the benchmark. Thus, to evaluate the effectiveness of sensor-placement schemes developed with our graph-theoretic approach, we compared their performances with the performances of the optimal results obtained from a simulation-based analysis for Network 1. The simulation-based analysis for Network 1 was originally conducted for the battle of the water sensor networks (BWSNs) at the 8th Annual International Water Distribution Systems Analysis Symposium, August 27–30, 2006, Cincinnati, Ohio. Steps of the simulation-based analysis are summarized here but can be found in detail in Krause et al. (2006). An exhaustive simulation of random contamination events was conducted with distinct contamination events beginning every 5 min for the first 24 h with each event lasting for 2 h. A suite of 288 ($24 \times 60/5$) contamination events was simulated for each of the 129 nodes in the network, yielding a total of 37,125 events. A submodular-function maximization model was applied for the four objectives (derived from the four criteria described above). To make the model results easy to interpret, the objective values were normalized in a way such that larger values are preferable (Krause, personal communication, October 16, 2006) and the normalized scores range from 0 to 1; i.e., 0 is the worst and 1 is the best. Thus, in the results, we do not show the actual effects (the expected population at risk, the expected volume of contaminated water, the expected detection time, or the detection likelihood). Instead, we only show the indicators based on the normalized scores for each objective. Table 2 shows the sensor-placement schemes for each of the four objectives and an equally weighted multiobjective based on simulation analysis for 20 sensors.

We have sensor-placement schemes based on four methods (betweenness centrality, nontime-constrained receivability, time-constrained receivability, and the exhaustive simulation-based analysis). For the exhaustive simulation-based analysis only, six sensor-placement schemes were derived based on different optimization objectives (each of the four objectives mentioned above, detection time and detection likelihood equally weighted, and all four objectives equally weighted). We evaluated different sensor-placement schemes with WaterSim, a software package that uses a database storing results for an exhaustive simulation of all combinations of contamination events (across time, location, and duration) for Network 1 (Krause, personal communication, October 16, 2006). WaterSim evaluates the performance of any set of sensor locations based on the indicator scores of expected detection time, expected population at risk, expected volume of contaminated water, and the detection likelihood. Figs. 7–10 show the performances of the different sensor-placement schemes for each

Table 2. Optimal Sensor Placement Based on a Simulation-Based Approach for Different Objectives and 1 to 20 Sensors

*i*th sensor	Minimizing detection time (Z_1)	Minimizing population at risk (Z_2)	Minimizing contaminated water consumed (Z_3)	Maximizing detection likelihood (Z_4)	Equally weighted (Z_1, Z_2, Z_3, Z_4)	Equally weighted (Z_1, Z_4)
1st	83	17	17	83	17	83
2nd	118	31	102	126	83	118
3rd	45	68	79	45	122	45
4th	100	80	68	10	31	100
5th	11	118	49	100	45	11
6th	123	21	118	123	100	123
7th	68	122	29	114	11	68
8th	35	96	97	124	126	35
9th	75	102	37	19	68	114
10th	90	77	83	34	90	124
11th	114	37	34	35	21	126
12th	124	46	31	11	35	10
13th	10	126	126	106	34	19
14th	126	34	46	110	118	34
15th	34	65	122	128	123	75
16th	19	4	74	129	114	90
17th	14	85	21	—	124	14
18th	21	26	5	—	76	21
19th	101	98	30	—	10	101
20th	72	117	94	—	19	72

of the criteria. Gauged across the four criteria, three sensor-placement schemes (i.e., those sensor-placement schemes based on equally weighted multiobjective, time-constrained receivability and betweenness centrality) give comparable performance, though differences are apparent. The sensor-placement scheme based on the full simulation and multiobjective optimization outperforms the sensor-placement scheme based on the time-constrained receivability, which outperforms the sensor-placement scheme based on betweenness centrality.

The sensor-placement scheme based on the nontime-constrained receivability, which targets detection likelihood, and the sensor-placement schemes based on simulation analysis with single objectives perform well on the corresponding single objective, but perform worse for the other objectives. The performance of sensor-placement scheme based on the nontime-constrained receivability is similar to the performance of sensor-placement scheme based on simulation analysis that only maximizes the detection likelihood. Among the top ten sensor locations in the sensor-placement schemes based on the nontime-constrained receivability and the single objective (detection likelihood), seven locations are the same. The sensor-placement scheme based on time-constrained receivability targets two objectives: Maximizing

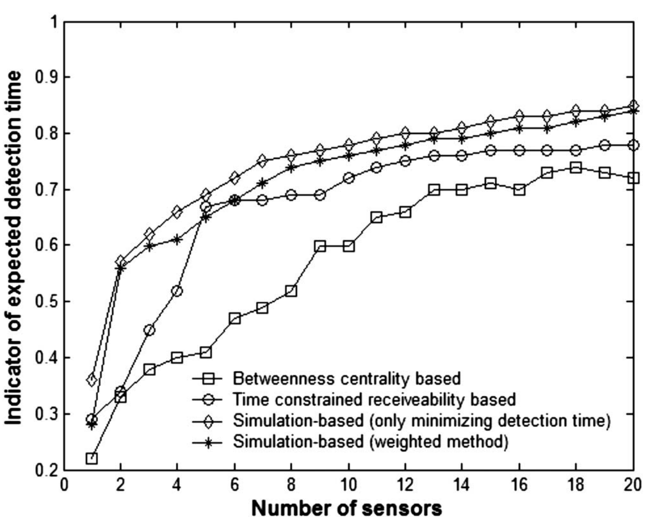

Fig. 7. Comparison of different sensor-placement schemes based on the expected-detection-time criterion

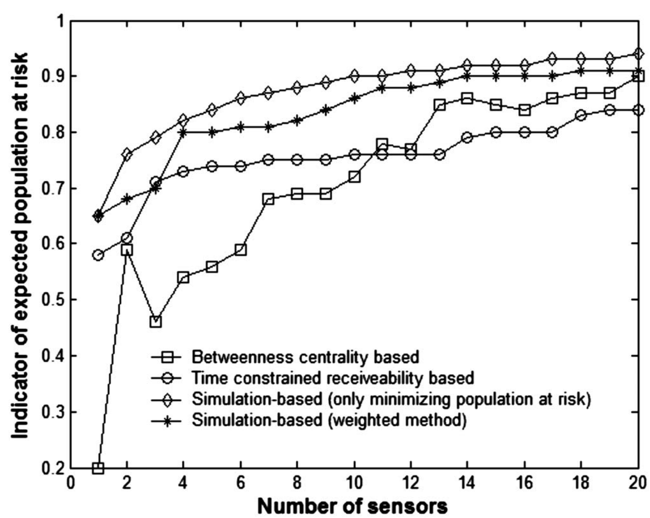

Fig. 8. Comparison of different sensor-placement schemes based on the expected-population-at-risk criterion

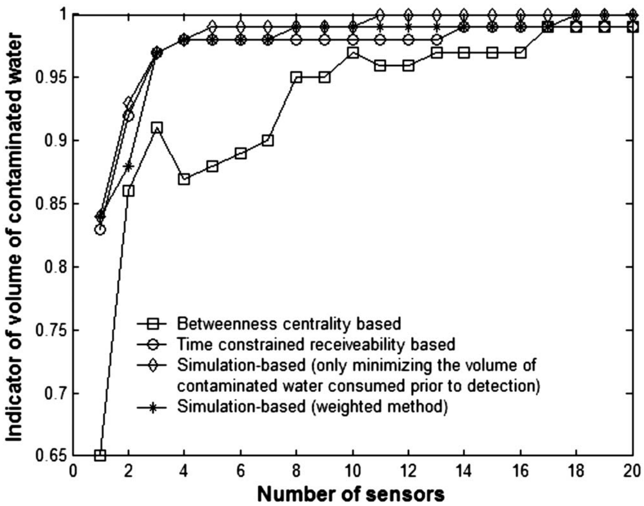

Fig. 9. Comparison of different sensor-placement schemes based on the expected-volume-of-contaminated-water-consumed criterion

detection likelihood and detecting contamination events as soon as possible (within the allotted time). The method based on time-constrained receivability gave comparable results with the equally weighted objectives of maximizing expected detection likelihood and minimizing expected detection time in the simulation-based method.

Conclusion

This paper provides a graph-theoretic approach to identify key sets of nodes for placing sensors in water distribution systems when the information available to a water utility is limited to the physical structure of the system or the hydraulic behavior of water in the distribution network. When a water utility only has the information of physical structure of their system, betweenness centrality, community structure, and graph distance can be used to identify a set of key nodes for placing sensors. The computational complexity for betweenness centrality, all pairs shortest path (for calculating graph distances), and the Newman algorithm (for identifying community structure) are $O[N(M+N)]$, $O(MN)$, and $O[DN\log(N)]$, respectively, with N=number of nodes in the graph; M=number of the edges in the graph; and D=height of the dendrogram tree (Boccaletti et al. 2006). When only a hydraulic model is available, nontime-constrained and time-constrained receivability can be calculated, and reasonable sensor locations can be identified using a greedy heuristic algorithm. These processes are all computationally affordable to any water utility.

Comparisons across the four criteria (the expected detection time, the expected population at risk prior to detection, the expected volume of water contaminated prior to detection, and the detection likelihood) were conducted for the sensor-placement schemes based on the graph-theoretic approach and sensor-placement schemes based on the simulation-based approach. The results showed that sensor-placement schemes based on the exhaustive simulation-based approach with equally weighted multi-objective performs better than that based on the time-constrained receivability analysis, which performs better than that based on the betweenness-based analysis. However, the differences between the three approaches are not large. The comparison between the nontime-constrained receivability analysis and the simulation-based analysis that maximizes detection likelihood shows that they yield very similar results. Simpler network-analysis methods are, thus, able to replicate more complex, simulation-based methods when their objectives are clearly aligned.

Furthermore, the receivability metrics are useful for demonstration purposes to educate water utility staff, to allow for the formulation of a mitigation plan in the case of a contamination event, and to inform scenario analysis for the purposes of developing preventative measures and strategies.

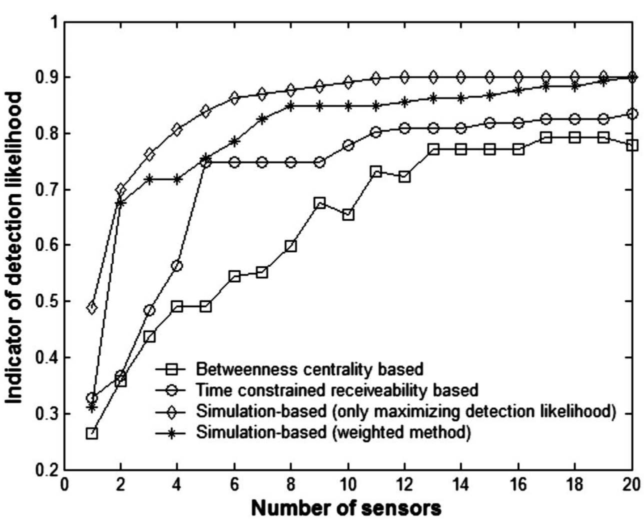

Fig. 10. Comparison of different sensor-placement schemes based on the detection-likelihood criterion

Acknowledgments

This work was funded by the National Science Foundation under Grant No. BES-0329549. The writers wish to thank Kathleen Carley and her group at Carnegie Mellon University for all their help in doing the network analysis, Andreas Krause for providing the WaterSim software package for evaluating sensor placement performance, and the BWSN organizers for providing the network.

References

Banavar, J. R., Colaiori, F., Flammini, A., Maritan, A., and Rinaldo, A. (2000). "Topology of the fittest transportation network." *Phys. Rev. Lett.*, 84(20), 4745–4748.

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

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

Blumsack, S. A. (2006). "Network topologies and transmission investment under electric-industry restructuring." Ph.D. thesis, Dept. of Engineering and Public Policy, Carnegie Mellon Univ., Pittsburgh.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. (2006). "Complex networks: Structure and dynamics." *Phys. Rep.*, 424(4-5), 175–308.

Borgatti, S. P. (2006). "Identifying sets of key players in a social network." *Comput. Math. Org. Theory*, 12(1), 21–34.

Carley, K., and De Reno, M. (2006). "ORA: Organization risk analyzer, ORA user's guide." ⟨http://www.casos.cs.cmu.edu/publications/papers.php⟩ (Nov. 12, 2006).

Cormen, T. H., Leiserson, C. E., and Rivest, R. L. (1997). *Introduction to algorithms*, 1st Ed., McGraw-Hill, New York.

Crucitti, P., Latora, V., and Porta, S. (2006a). "Centrality in networks of urban streets." *Chaos*, 16(1), 015113.

Crucitti, P., Latora, V., and Porta, S. (2006b). "Centrality measures in spatial networks of urban streets." *Phys. Rev. E*, 73(3), 036125.

Drezner, Z., and Hamacher, H. (2002). *Facility location: Applications and theory*, Springer, Berlin.

Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). "On power-law relationships of the internet topology." *Comput. Commun. Rev.*, 29(4), 251–262.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness." *Sociometry*, 40(1), 35–41.

Gould, R. V. (1987). "Measures of betweenness in nonsymmetric networks." *Soc. Networks*, 9, 277–282.

Grayman, W. M., Ostfeld, A., and Salomons, E. (2006). "Locating monitors in water distribution systems: Red team-blue team exercise." *J. Water Resour. Plann. Manage.*, 132(4), 300–304.

Gueli, R. (2006). "Predator-prey model for discrete sensor placement." *Proc., Conf. on Water Distribution Systems Analysis, the 8th Annual Int. Water Distribution Systems Analysis Symp.*

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

Krause, A., et al. (2006). "Optimizing sensor placements in water distribution systems using submodular function maximization." *Proc., Conf. on Water Distribution Systems Analysis, the 8th Annual Int. Water Distribution Systems Analysis Symp.*

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

Nemhauser, G. L., Wolsey, L., and Fisher, M. (1978). "An Analysis of the approximations for maximizing submodular set functions." *Math. Program.*, 14, 268–294.

Newman, M. E. J. (2004). "Fast algorithm for detecting community structure in networks." *Phys. Rev. E*, 69(6), 066133.

Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *J. Water Resour. Plann. Manage.*, 130(5), 377–385.

Ostfeld, A., Uber, J., Salomons, E. (2006). "Battle of the water sensor networks (BWSNs): A design challenge for engineers and algorithms." *Proc., Conf. on Water Distribution Systems Analysis, the 8th Annual Int. Water Distribution System Analysis Symp.*

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

Shastri, Y., and Diwekar, U. (2006). "Sensor placement in water networks: A stochastic programming approach." *J. Water Resour. Plann. Manage.*, 132(3), 192–203.

Su, Y., Mays, L. W., Duan, N., and Lansey, K. E. (1987). "Reliability-based optimization model for water distribution systems." *J. Hydraul. Eng.*, 113(12), 1539–1556.

Todini, E., and Pilati, S. (1987). "A gradient method for the analysis of pipe networks." *Proc., Int. Conf. on Computer Application for Water Supply and Distribution.*

Wagner, J. M., Shamir, U., and Marks, D. H. (1988a). "Water distribution reliability: Analytical methods." *J. Water Resour. Plann. Manage.*, 114(3), 253–275.

Wagner, J. M., Shamir, U., and Marks, D. H. (1988b). "Water distribution reliability: Simulation methods." *J. Water Resour. Plann. Manage.*, 114(3), 276–294.

Wasserman, S., and Faust, K. (1994). *Social network analysis: Methods and applications*, 1st Ed., Cambridge University Press, New York.

Watson, J., Greenberg, H. J., and Hart, W. E. (2004). "A multiple-objective analysis of sensor placement optimization in water networks." *Proc., 2004 World Water and Environmental Resources Congress: Critical Transitions in Water and Environmental Resources Management*, 4609–4618.