MULTIOBJECTIVE OPTIMIZATION OF CONTAMINANT SENSOR LOCATIONS IN DRINKING WATER DISTRIBUTION SYSTEMS USING NODAL IMPORTANCE CONCEPTS

A Thesis
Presented to
The Academic Faculty

By

Scott W. Rogers

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Environmental Engineering
School of Civil and Environmental Engineering

Georgia Institute of Technology

August 2009

MULTIOBJECTIVE OPTIMIZATION OF CONTAMINANT SENSOR LOCATIONS IN DRINKING WATER DISTRIBUTION SYSTEMS USING NODAL IMPORTANCE CONCEPTS

Approved by:

Dr. Mustafa M. Aral, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Jiabao Guan
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Wonyong Jang
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Seong Hee Kim
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Turgay Uzer
School of Physics
Georgia Institute of Technology

Date Approved: May 14, 2009

For my dear Denise

ACKNOWLEDGEMENTS

Compared with the journeys through graduate school of many of my colleagues, my graduate school experience has been exceptionally rigorous. Throughout graduate school, I have had to manage an array of challenges—both academic and personal. My life was even in question at one point during an ordeal with cancer. I could not have met these challenges and persevered to finish my doctoral work without the support of many people. Below, I will endeavor to give these people at least some of the credit they deserve.

First, I would like to express my gratitude for Dr. Mustafa Aral, my advisor. His guidance and patience throughout my graduate school experience were invaluable. I wish to thank my thesis committee as well for the time and input they contributed to help me conduct my research properly. Mr. Morris Maslia of the Agency for Toxic Substances and Disease Registry also eagerly provided guidance throughout my doctoral studies. The Multimedia Environmental Simulations Laboratory (MESL) has been my professional family for years; the research engineers and graduate students in MESL were always available to provide guidance and encouragement both professionally and personally even during the toughest of times. The Georgia Tech Environmental Engineering program as a whole also deserves credit for its support during both the coursework and research stages of my graduate school experience, and I greatly appreciate the Association of Environmental Engineers and Scientists for allowing me to

provide a variety of services for the program over the years so that I could reciprocate that support.

On a more personal level, I must express my infinite gratitude for my lifelong friends Mike and John; I called on them many times in the past few years for emotional support, and they always delivered. My family—my mother Martha, my father Wayne, my brother Ken, and my sister Renee—have also endured many hardships recently in addition to my own cancer, but they always expressed their love and eagerness to help whenever I needed them. While in graduate school, I met the love of my life, Denise. She has truly been a blessing and has had to be strong many times for both of us. I can safely say that I would not have been able to complete my doctorate or even stay grounded during my major personal ordeals without her love. For these reasons, I have dedicated my thesis to her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

BWSN	Battle of the Water Sensor Networks
EPA	Environmental Protection Agency
ft	feet
GA	genetic algorithms
gal	gallons
GB	gigabytes
GHz	gigahertz
g/min	grams per minute
gpm	gallons per minute
GRASP	Greedy Randomized Adaptive Search Procedure
h	hours
ISSM	Iterative Subset Search Method
L/h	liters per hour
MB	megabytes
mg/L	milligrams per liter
mi^2	square miles
min	minutes
MIP	mixed-integer program
RAM	random access memory
WDS	water distribution system
WDSA	Water Distribution System Analysis

SUMMARY

The monitoring of water distribution systems (WDSs) has been a very popular subject of study since the terrorist attacks of September 11, 2001, and the subsequent passing of laws motivating the study of WDS monitoring to provide system protection in the event of a terrorist attack. Inhibiting many WDS monitoring studies to date is the large amount of computational expense required to conduct meaningful studies, especially for larger WDSs that are of most interest.

In this study, methods were developed to determine the "importance" of WDS nodes in being considered as locations for sensors used to monitor a WDS in order to make sensor placement optimization more efficient. Single-objective protection goals considered individually in optimization were maximizing detection likelihood, minimizing expected detection time, and minimizing expected contaminated demand volume. A multiobjective protection goal accounting for all three single-objective goals concurrently was also considered; the formulation of the multiobjective optimization problem was intended to minimize tradeoffs among individual protection goals. Sensor placement optimization was carried out with the Iterative Subset Search Method (ISSM) employing genetic algorithms developed in this work; ISSM used nodal importance rankings to search a small subset of nodes for the optimal solution initially then broadened the search incrementally until convergence to a best solution occurred.

To demonstrate the effectiveness of the methods developed, sensor placement was performed according to each of the protection goals for three study systems—one small and two large—and a variety of attack conditions. Desirable sensor node solutions that provided for significant protection were found in all cases, and in many cases sensor placement results were comparable to or better than those of other works. Nodal importance narrowed the search for optimal sensor nodes to a relatively small proportion of WDS nodes in most cases.

CHAPTER 1

INTRODUCTION

In this introductory chapter, the motivations for the study at the center of this thesis are discussed, and the major objectives of the study in accordance with those motivations as well as the objectives of the thesis itself are outlined.

1.1 MOTIVATIONS FOR AND CHALLENGES OF WATER DISTRIBUTION SYSTEM MONITORING RESEARCH

The monitoring of water distribution systems (WDSs) has been a popular subject of study interest for many years. Initially, WDS monitoring research was motivated primarily by WDS operational needs and the potential for accidental WDS contamination (e.g., Kessler et al. 1998; Lee and Deininger 1992). However, the focus of study for WDS monitoring more recently has become the protection of a WDS in the event of intentional contamination of the system in the form of a terrorist attack (e.g., Berry et al. 2005; Ostfeld and Salomons 2004). In response to the terrorist attacks of September 11, 2001, the U.S. Congress amended the Safe Drinking Water Act (Title XIV of the Public Health Service Act) with the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 by mandating the development of methods to "prevent, detect, and respond to the intentional introduction of chemical, biological, or radiological contaminants into community water systems and source water for community systems" (U.S. Congress 2002). Consequently, the U.S. Environmental Protection Agency (U.S.

EPA) conducted a vulnerability assessment study in which individual vulnerability assessments of WDSs in communities with populations of 3,300 people or greater were examined in order to determine the aspects and degrees of vulnerability that exist for various aspects of water supply and distribution (U.S. EPA 2003). Factors that were considered in the individual assessments are the following:

- system characteristics,
- possible adverse consequences and their severity relative to each other,
- critical WDS assets subject to potential malevolent acts,
- likelihood of malevolent acts,
- existing system countermeasures, and
- measures being taken to develop a prioritized plan for risk reduction for the system.

The vulnerability assessment study indicated that the communities conducting individual assessments as a whole were not explicitly accounting for the potential of contamination—intentional or otherwise—as a threat to the WDS (ASCE 2004). Thus, there has been much room for the study of how contamination events with uncertain characteristics can affect the WDS and its customers. The first two vulnerability-related factors listed above have garnered the most attention from environmental engineers, health scientists, and the like who have been performing research to explicitly address the intentional contamination problem since the terrorist attacks of 2001. To quantifiably assess those two factors of vulnerability, several state and decision variables associated with specific systems must be taken into account:

- WDS physical characteristics. Every WDS has a unique set of physical characteristics that influence greatly the potential for adverse effects attributed to a terrorist attack and the ability to monitor the system adequately. WDS networks can range from hundreds to millions in numbers of nodes and links and cover areas of a few to hundreds of square kilometers. Monitoring the system becomes more difficult as system size increases for the number of possible attack locations and number of WDS components needing protection both increase. The set of nodes and links can be configured in a virtually infinite number of manners, allowing for very dynamic flow and transport behavior that can require complex modeling and a large amount of computational expense. Nodes mostly take the form of network junctions (i.e., connections of two or more pipe segments) but may take the form of hydraulic devices such a tank or a reservoir. Links typically represent pipe segments but may represent pumps and valves as well. Hydraulic devices can impact WDS flow and transport behavior to a large degree.
- WDS customer types. A WDS can serve a variety of types of customers such as residences, schools, hospitals, industries, and amusement parks. Each type of customer has particular water needs. Also, some types of customers may need more protection than others. For instance, a school may require more protection than an industry as contaminated water consumed by a school has a greater chance of causing adverse human effects and, in turn, psychological damage for the population-at-large.
- WDS hydraulic demand profile. Each node in the WDS is associated with a hydraulic demand time series over some specific period of time that is a function

of the number and type of customers served by the node. The system demand profile is the major contributor in the flow and transport behavior of the system.

- Sensors. Independent of a WDS in question, the type of contaminant sensor used to monitor the system and the number of sensors available to be employed are decision variables that affect WDS monitoring proficiency. The type of sensor used is tied to the ability of the sensor to detect a particular contaminant that is present at a particular concentration at the location of the sensor. Additionally, sensors represent the cost of monitoring the system. This cost is in terms of the sensors themselves and the implementation and continuous operation costs of the sensors.
- Hierarchy of WDS monitoring protection goals. The WDS protection provided by a set of sensors can be quantified in terms of performance measures like time after contaminant injection to detection of the contaminant, demand volume contaminated, the likelihood of detection, and others. As certain protection goals are in conflict with other protection goals (i.e., an improvement in performance measure value with regard to one protection goal may lead to a worsening in performance measure value with regard to another protection goal), it is up to decision-makers to prioritize protection goals. For instance, the protection of people may be the primary goal for a given WDS that serves a mostly residential area, so the demand volume contaminated may be considered a more important performance measure at the opportunity cost of decreased detection likelihood as well as possibly protection with respect to other performance measures.

However, time after injection to detection may be considered more important if the primary goal is the minimizing of remediation costs.

Unfortunately, human variables inherent to a terrorist attack of uncertain characteristics increase significantly the complexity of quantifying vulnerability. Some of these variables follow:

- Number of attack events. An indefinite number of contaminant injection events of potentially varied characteristics could occur in a particular attack.
- Location(s) of attack. Per attack, the entire WDS or some subset of WDS locations is susceptible to one or more contaminant injection events at one unknown location or at multiple unknown locations.
- Time(s) of attack. A contaminant injection event can occur at any time in a given period. Multiple events can occur simultaneously or at different times.
- Specifics of contaminant(s). A contaminant employed in a contaminant injection event can be biological or chemical in nature and be associated with a unique set of interesting physical and chemical properties. The contaminant likely would enter the system in a controlled manner at an either steady or unsteady rate and at either a constant or variable concentration. One or multiple contaminants can be employed in one attack.

Studies are documented that present methods for WDS monitoring through sensor placement considering the above variables. However, as acknowledged in various studies (e.g., Dorini et al. 2006; Huang et al. 2006), determining the WDS effects of a

realistic array of possible attack scenarios can require a great deal of computational expense in terms of both memory and runtime due to the large domains of values for variables corresponding to WDSs and terrorist attacks. In many studies that optimize sensor placement for a large-scale, real-world WDS, computational expense led to either computational infeasibility or oversimplifying assumptions that compromised accuracy for computational feasibility. These issues are especially concerning for urban systems serve more than 75% of the U.S. population and are at greater risks of being terrorist targets than are smaller systems (Copeland and Cody 2002). Thus, computational expense has hindered the development of definitive methods for WDS monitoring.

Given the negative correlation between WDS sensor placement proficiency and computational expense, a method is needed that assigns contaminant sensors to locations in a large WDS in order to acceptably maximize WDS protection while observing monitoring cost constraints and minimizing the computational expense needed.

1.2 OBJECTIVES OF THIS WORK

In this study, methods were developed that determine the "importance" of all eligible WDS locations (WDS nodes exclusively) in being considered for sensor placement in order to make the search for the optimal set of sensor locations more efficient. Nodal importance is directly tied to the on-average protection gained according to selected protection goals by placing a sensor at a given node. The protection goals of interest for this study are maximizing detection likelihood, minimizing the expected time after

contaminant injection to detection (i.e., expected detection time), and minimizing the expected demand volume contaminated. These protection goals are considered individually in single-objective optimization problems and collectively in a multiobjective optimization problem that minimizes protection tradeoffs among individual protection goals. The search for optimal sensor locations is conducted with the Iterative Subset Search Method employing genetic algorithms developed in this work. The Iterative Subset Search Method uses nodal importance rankings to begin the search for the optimal sensor nodes with a small subset of WDS nodes then broadens the search incrementally until convergence to a best solution occurs. Constraints of the optimization include a limit on sensor availability that represents in a lumped manner an upper bound on the cost of sensor placement. Methods of this study are applied to three study systems that provide a diversity of system characteristics appropriate for demonstrating the robustness of the developed sensor placement methods.

Chapter 2 of this thesis presents a discussion regarding the evolution of WDS monitoring over approximately the last two decades. More specifically, sensor placement methods submitted by various researchers will be explained and analyzed according to their ability to protect a WDS from the adverse effects of contamination and their computational expense drawbacks. In Chapter 3, the particular problem of this study is defined in detail. Mathematical representations of protection goal performance measures are given, and relationships among protection goals are discussed. Chapter 3 also gives the assumptions and other confining parameters of the study along with qualitative and quantitative evaluations of three study systems. Chapter 4 provides the solution approach that was

applied to the study problem. The single-objective and multiobjective optimization problems as well as nodal importance concepts are formulated mathematically, and the ways nodal importance concepts are used in conjunction with optimization are explained. Chapters 5, 6, and 7 present the results of application of the developed methods on the three study systems. In Chapter 8, the work presented in this thesis is summarized, and conclusions are drawn to justify the work as an advancement in WDS monitoring.

CHAPTER 2

LITERATURE SURVEY

An overview of many of the salient WDS monitoring studies documented to date is given in this chapter in order to provide context for the work detailed in this thesis.

2.1 WDS MONITORING STUDY PRIOR TO SEPTEMBER 11, 2001

As discussed in Chapter 1, sensor placement studies were mostly motivated by the possibility of accidental contamination before September 11, 2001. Several of these studies laid the foundation for future studies of WDS monitoring in the case of intentional contamination.

One such study was that of Lee and Deininger (1992). Lee and Deininger devised a method to place sensors at nodes in a WDS so that a sensor set maximized system "demand coverage", the percentage of system hydraulic demand monitored by the sensor set. This sensor placement optimization approach consisted of two major stages: constructing the "water fraction matrix" and solving an integer program. In the first stage, the authors used hydraulic pathway analysis to develop the water fraction matrix containing the particular fractions of demand for a WDS node flowing through specific upstream nodes. That information was employed in the second stage in the solving of the integer program with an objective function that maximized demand coverage subject to the constraint of sensor availability.

The method worked well for the small study system of the Lee and Deininger study, but drawbacks would exist if the method were applied to larger, real-world systems. A critical assumption of the method is that water quality lessens as time and distance from a given source of contaminant intrusion increases. Thus, the water quality at a node is the indication of the water quality of a node upstream from the node in question. The assumption may be valid if water quality deteriorates gradually, but contaminant intrusion generally causes rapid water quality deterioration. The assumption is a by-product of the more overarching assumption the method employs of time independence. System flow and contaminant transport are unrealistically considered steady-state. The time between contaminant intrusion and detection is not accounted-for. Such neglecting of time dependence can lead to the contamination of much hydraulic demand at many WDS nodes. In addition, the two-stage optimization process would be very computationally cumbersome for larger systems due primarily to the need to analyze all individual hydraulic pathways of the WDS.

Despite the aforementioned drawbacks, other researchers found the work of Lee and Deininger significant and continued the study of WDS monitoring by building on the work. Kumar et al. (1997) applied a greedy heuristic-based algorithm to solve the demand coverage problem and produced results similar to those presented by Lee and Deininger, though no proof of global optimality or near-optimality was given. Also, Kumar et al. claim extendibility of their work to larger systems, but they only tested their method on systems of 19 nodes or less. Harmant et al. (1999) modified the demand coverage-based sensor placement method by accounting for time dependence and

contaminant concentration in the integer program objective function. The means employed by Harmant et al. biased coverage consideration toward nodes with flows of higher magnitudes and more aged water. To correct this bias somewhat, Woo et al. (2001) modified the integer program objective function by normalizing concentration values according to contaminant source concentration values in order to emphasize coverage consideration for nodes of lower water quality. Al-Zahrani and Moied (2001) applied genetic algorithms to carry out the demand coverage-based optimization.

Another key study was conducted by Kessler et al. (1998). This study was based loosely on the Lee and Deininger (1992) study in terms of placing sensors to cover nodal demands. It was particularly innovative at the time for considering the random nature of accidental contamination through the development of the randomized pollution matrix and imposing a minimum protection-based level-of-service. The level-of-service in that study was a function of contaminated demand volume. Thus, to impose a minimum level-of-service was to impose a criterion for the maximum demand volume allowed to be contaminated prior to sensor detection of the contaminant. In short, the sensor placement optimization methodology consisted of establishing an auxiliary WDS network to represent all possible flow directions for a typical demand cycle, conducting a shortest-paths analysis to identify domains of pollution, and determining the "minimum covering set" for choosing optimal sensor locations. This study also has drawbacks that limit its applicability. As in the case of the Lee and Deininger (1992) study, the Kessler et al. (1998) study assumes time independence. Also, like Kumar et al. (1997), Kessler et

al. claim extendibility to larger systems, but the largest system employed in testing was the 18-node "Anytown, U.S.A.".

In summary, the WDS monitoring methods developed prior to September 11, 2001, described above have substantial drawbacks and are not applicable for real-world WDSs on the whole. However, the methodological contributions made by those studies would allow for significant advancements in WDS monitoring to be made immediately after September 11, 2001.

2.2 WDS MONITORING STUDY ADVANCES SINCE SEPTEMBER 11, 2001

Shifts in WDS monitoring priorities are evident in sensor placement work that has been documented since September 11, 2001. The uncertainties associated with a terrorist attack on a WDS explained in Chapter 1 have received increased attention. WDS protection goals have become more oriented toward safeguarding humans. Study systems have increased in size and complexity so that sensor placement methods can be tested under real-world conditions. Some of the more innovative studies that have been performed since September 11, 2001, are detailed and analyzed below.

Ostfeld and Salomons (2004) adapted the concepts of the randomized pollution matrix and the demand-based level-of-service developed by Kessler et al. (1998) for the set-covering sensor placement approach in order to address the problem of intentional contaminant intrusion. In their study, contaminant concentrations were taken into

account in determining if levels of service were met for water was considered contaminated only if its contaminant concentration equaled or exceeded some threshold value. The actual optimization to find the minimum covering set of sensors was carried out by genetic algorithms – a more straightforward means of optimization than that done by Kessler et al., though genetic algorithms cannot provide a guaranteed global optimal solution to an optimization problem. In addition, Ostfeld and Salomons tested their method under more realistic system conditions by considering unsteady-state flows and the possibility of multiple contaminant injection locations and times during one attack scenario. However, "Anytown U.S.A." was the largest WDS employed in their study, and therefore the extendibility of the method to larger systems of more interest was not tested. Optimization with genetic algorithms can be very computationally expensive in terms of runtime for larger-scale problems, as acknowledged by Ostfeld and Salomons. Thus, the extendibility of the method to larger systems cannot be assured at all.

Watson et al. (2004) expanded upon the Ostfeld and Salomons study by reformulating the objective function of the optimization problem in order to provide protection according to a variety of objectives, such as contaminant travel time as a function of contaminant propagation rates corresponding to individual flow patterns. Ostfeld and Salomons later extended their own work to account for uncertainties with regard to contaminant injection flow rate, consumer demands, sensor detection resolution, and response time after sensor detection (Ostfeld and Salomons 2005). In the 2005 study, Ostfeld and Salomons applied their method to "EPANET Example 3", a 97-node WDS. This system is larger than Anytown U.S.A., but it is small relative to urban-scale systems of most interest. Even in

testing on the relatively small EPANET Example 3, Ostfeld and Salomons acknowledge a great amount of computational expense. For instance, 4 to 7 hours were required to construct one randomized pollution matrix. As a consequence of the computational expense, they could only simulate short contaminant injection scenarios.

The study of Berry et al. (2005) was one of the first WDS monitoring studies to examine the effects of intentional contaminant injection on the population served by a WDS as well as the possibility of certain nodes in the WDS being at higher risks of attack than would be other nodes in the system. The population served at a given time by a WDS node was assumed correlated with the hydraulic demand experienced at the particular node and particular time. Without definitive information regarding the numbers of customers served by particular WDS nodes, Berry et al. accounted for population density uncertainty by somewhat arbitrarily fluctuating demands at nodes throughout the study period. Nodal attack risks were arbitrarily assigned to form a fixed probability distribution across the nodes of the WDS. Attack scenarios were assumed to consist of one contaminant injection event at one node. The sensor placement problem was solved as a mixed-integer program with an objective function that minimizes the expected fraction of the population served by a WDS that would be exposed to contaminant. As in the cases of earlier studies, a shortcoming of the Berry et al. (2005) work is that time independence is assumed for contaminant flow and transport conditions. Berry et al. tested their method on a 470-node system. The system was much larger than those employed by prior studies but still does not constitute an urban-scale system.

Shastri and Diwekar (2006) disagreed somewhat with the deterministic approach Berry et al. (2005) used to quantify uncertainty and developed a stochastic nonlinear optimization program with recourse to solve the sensor placement problem. The updated program formulation accounts for, in part, frequencies of flow patterns occurring and the costs of sensor placement. Shastri and Diwekar claimed to have objective function values superior to those of the Berry et al. (2005) study. However, the stochastic approach was acknowledged to be more computationally expensive on average than the deterministic approach, even for the 36-node "EPANET Example 2" WDS. As no large systems were used by Shastri and Diwekar to test their method, it is unclear if application of the method to urban-scale systems would be feasible. Carr et al. (2006) also extended the work of Berry et al. (2005) by employing the mixed-integer programming approach and quantifying uncertainty in three different manners: unweighted, linearly weighted, and bilinearly weighted. One observation Carr et al. make is that computational expense would increase by orders of magnitude with increases in the number of sensors used to monitor the WDS, system size, and robustness of the optimization model. Propato (2006) formulated the sensor placement problem as a binary polynomial program transformed into a mixed-integer program. Propato intended to improve upon the solution approaches presented in the work of Berry et al. (2005) as well as in similar works by accounting for the temporal aspects of contaminant transport and allowing for optimization to serve an array of protection objectives, though not in a multiobjective manner.

To address the lack of time dependence in their previous work, Berry et al. (2006b) revised the mixed-integer program (MIP) formulation to make it a temporal-MIP model.

Instead of solely minimizing the expected population fraction affected, the updated model was more generic so that sensors could be placed according to any protection objective performance level indicated by "nodal impact coefficients". The impact coefficient for a WDS node represents the expected adverse effects experienced at that node given contaminant detection by a sensor located at some particular node in the system. Contaminant concentrations experienced at nodes during the study period were determined for use in the optimization in this study, as opposed to the Berry et al. (2005) study that did not consider water quality variation. The temporal-MIP was solved as a *p*-median facility location problem using separately an MIP solver and the heuristic Resende-Werneck (RW) greedy randomized adaptive search procedure (GRASP). The large multidimensional nature of this approach can potentially be very computationally expensive, so Berry et al. compressed the impact coefficient matrix by combining in a sense information regarding nodes with equal coefficient values in order to reduce the memory required for computational feasibility. Though Berry et al. claim minimal runtimes, executing the GRASP solution approach on a 12,000-node system took over two days, factoring in the time required to run attack simulations needed for impact coefficient calculation. Furthermore, the authors had to confine the set of possible attack times to only four possible times in order to keep computational burden manageable.

In order to create a formal dialogue regarding WDS monitoring in light of terrorism concerns, the "Battle of the Water Sensor Networks" (BWSN) competition (Ostfeld et al. 2006) was held at the 8th Annual WDSA Symposium in August 2006. Researchers from around the world submitted various methods for sensor placement and the corresponding

sets of results based on a common study problem. The competition guidelines stipulated that participants optimize sensor placement according to four protection goals: minimization of expected contaminant detection time, minimization of expected number of people affected by contamination, minimization of expected contaminated demand volume, and maximization of detection likelihood. Participants were also instructed to examine the sensitivity of sensor placement to the attack uncertainties concerning the number of contaminant injection locations, duration of contaminant injection, and response time after contaminant detection. Two study systems, "BWSN Network 1" (129 nodes) and "BWSN Network 2" (12,527 nodes), were provided for participants to enable testing of methods on both a small system and a large system. These systems are discussed in greater detail in Chapter 3.

There were two major drawbacks of general consensus among participants in BWSN competition design. First, the competition guidelines requested that attack scenarios for which contaminant is not detected by a sensor should not be considered in calculating expected values of performance measures over all scenarios run. This condition led to significant underestimation of the realistic values for the expected number of people affected and the expected contaminated demand volume. Second, no scoring rubric for measuring the solution quality was initially presented to participants. Thus, participants conducted multiobjective optimization without understanding the desired weighting of protection goals and solved the problem according to different protection priorities. Despite these issues, BWSN provided calibrated data for two very different study systems, an array of submitted optimization methods with unique strengths and setbacks

explained, and an array of submitted solutions to the sensor placement problem. Therefore, the conditions of the work of this thesis mirror largely those of BWSN as explained in Chapter 3 in order to allow for comparisons of methods and results.

A variety of ways of formulating the optimization problem were presented by BWSN participants, including binary integer program (e.g., Guan et al. 2006), mixed-integer program (e.g., Krause et al. 2006; Propato and Piller 2006), temporal-MIP as a *p*-median problem (Berry et al. 2006c), and predator-prey model (Gueli 2006). Problem formulations were solved through a variety of means, but genetic algorithm-based means were particularly popular (e.g., Guan et al. 2006; Huang et al. 2006; Ostfeld and Salomons 2006; Wu and Walski 2006).

Most BWSN entries had "good" to near-optimal solutions. (Specific results of the competition are provided in Chapter 5 as needed for comparisons with results of this work.) However, the entries exhibited common drawbacks. One such drawback is the difficulty in handling the multiobjective aspect of the optimization. While some participants produced methods that are geared explicitly for the multiobjective nature of the problem (e.g., Guan et al. 2006; Ostfeld and Salomons 2006), other participants used a weighted-sum objective function that lumped all objectives into a single objective (e.g., Krause et al. 2006; Propato and Piller 2006; Wu and Walski 2006) or assumed that optimizing according to one objective would be sufficient for other objectives are correlated to the one objective (e.g., Berry et al. 2006c). In addition, great computational expense was acknowledged by most of the participants. The participants who provided

specific metrics of computational expense entailed in carrying out their methods are listed in Table 2.1 with the corresponding metrics. All studies in Table 2.1 except one had runtimes of multiple days. Eliades and Polycarpou (2006) had runtimes of less than one day, but their method did not search for the optimal sensor placement solution as extensively as did other methods. The work of Krause et al. (2006) could be considered a computational "worst-case scenario" due to the large memory and runtimes needed to exhaustively simulate all possible attack scenarios. Even though they simulated all possible scenarios, Krause et al. could use only a limited amount of the simulated data in optimization in order to maintain computational feasibility. Some participants did not attempt to solve the problem for the larger BWSN Network 2 (Gueli 2006; Propato and Piller 2006), perhaps due to the computational expense required. Others did not solve the problem for all of the contamination cases given by the competition guidelines for BWSN Network 2, such as Berry et al. (2006c) who cited computational expense as the reason for neglecting cases.

Table 2.1 Computational details provided by selected submissions of the Battle of the Water Sensor Networks competition.

Author(s)	Optimization Method(s)	Computational Platform	BWSN Network 1		BWSN Network 2		Relevant Computational Observations/ Decisions
			Reported Estimated Memory Required	Reported Estimated Total Runtime	Reported Estimated Memory Required	Reported Estimated Total Runtime	
Berry et al. (2006c)	*p*-median problem / GRASP heuristic	AMD, 2.2 GHz, 64-bit	*not given*	*total time not given*	*not given*	3-4.5 days	• too computationally expensive to place sensors in the case of 2 injection nodes for BWSN Network 2
Dorini et al. (2006)	Noisy Cross-Entropy Sensor Locator	Pentium M, 1.8 GHz, 1 GB RAM, 2 MB L2 cache	9-90 MB	30 min - 8 h	470-600 MB	2-3 days	• limited nodes considered for injection to make computationally feasible • "non-detected" scenarios excluded
Eliades and Polycarpou (2006)	Iterative Deepening of Pareto Solutions	Pentium IV, 3.4 GHz, 2 GB RAM	*not given*	20 min - 3 h	*not given*	1.5-14 h	• limited nodes considered for injection to make computationally feasible • precise enough to find a set of "good" solutions, but not necessarily near-optimal
Gueli (2006)	Predator-Prey Model	Pentium IV, 2.8 GHz, 1.5 GB RAM	225 MB	22 min maximum	*no tests run*		
Huang et al. (2006)	Genetic Algorithms with Data Mining	Pentium IV, 2.8 GHz, 704 MB RAM	25 MB	1.5 h - 7.5 days	50 MB	7.5-21 days	• limited nodes considered for injection to make computationally feasible
Krause et al. (2006)	Submodular Function Maximization	Intel Xeon 3 GHz, 20 GB RAM	14.6 MB - 1.6 GB	8 days per scenario	15-30 GB	8 days per scenario	• split task among 30 processors • had to use sparse representations of simulation results for computational feasibility

Most recently, studies explicitly addressing the need for increased efficiency in sensor placement have been performed. Such studies have aimed to relate effective sensor placement to the location and attributes of a WDS node or group of nodes. For instance, Isovitsch and VanBriesen (2008) analyzed sensor node solutions given by Krause et al. (2006) for BWSN Network 2 according to spatial trends of placement for various protection goals and attack parameters. They surmise that sensor placement is correlated with average "reachable" demand and "reachability" in manners dependent upon the protection goal or goals of interest. They also submit that sensor nodes for different arrays of attack parameters tend to be in close proximity to each other. Isovitsch and van Briesen believe knowledge of spatial trends in sensor placement could be useful in isolating desirable candidates for sensor placement from a large set of WDS nodes given further study.

Xu et al. (2008) more directly tried to use node locations and attributes to place sensors. Instead of using a flow and transport model to simulate WDS behavior, Xu et al. represented WDS behavior with three types of graphs (undirected, dynamic directed, and weighted dynamic directed). Graph theory allowed them to employ the concepts of "betweenness centrality" and "receivability" to concisely describe the desirability of a node as a sensor node candidate. While the idea of developing such indicators is sound, the particular concepts used by Xu et al. yield distinctly suboptimal solutions evident from their own comparison of results to those of Krause et al. (2006). The oversimplification of complex, unsteady-state WDS behavior through the use of graph theory may have contributed to the suboptimality of solutions. Xu et al. claim that

decreased protection performance is acceptable given gains in computational efficiency, but they do not provide any measurements of computational expense and only demonstrate their methods on the relatively small, 129-node BWSN Network 1. Also, it seems that computations to determine betweenness centrality and receivability could be rather computationally burdensome to carry out for larger systems. Therefore, the work of Xu et al. puts forth interesting ideas concerning how to prioritize nodes for sensor placement, but the work does not provide adequately viable means for prioritization.

2.3 NEEDED DIRECTIONS OF ONGOING WDS MONITORING WORK INFERRED FROM PREVIOUS STUDIES

Upon examining the nearly two decades of research pertinent to WDS monitoring for the purpose of minimizing the adverse effects of intentional contaminant injection summarized above, several observations are made that has guided the work of this thesis. The study of sensor placement appears to be in its infancy at present based on the lack of a definitive sensor placement method that would be adequate for large, real-world systems and a diverse set of protection goals. The amount of data available regarding the performance of methods on urban-scale WDSs is not abundant; urban-scale systems became a focus of earnest study only very recently. The continued exploration of the multiobjective aspect of sensor placement is needed as many researchers studying sensor placement differ from each other on how to handle the tradeoffs between conflicting protection goals. The paramount issue to address seems to be the computational expense required to optimize sensor placement for large systems; it has been consistently

acknowledged in documented works as a crippling limiting factor. Computational expense has been the cause for oversimplifying assumptions in optimizing sensor placement or even computational infeasibility in applying methods to large systems in prior studies. It is believed that the work detailed in this thesis will contribute to the reconciling of these issues.

CHAPTER 3

STUDY PROBLEM DEFINITION

The specific WDS monitoring problem studied in this work is defined extensively in this chapter.

3.1 PROTECTION GOAL PERFORMANCE MEASURES

As indicated in Chapter 1, the overarching goal of this study is the development of a method that provides maximum WDS protection from a contaminant injection attack with uncertain characteristics through the allocation of contaminant monitoring sensors to locations in the WDS network that observes cost constraints and minimizes computational expense required. WDS protection in this study is represented by three performance measures: detection likelihood, expected detection time, and expected contaminated demand volume. Preferred performance is associated with maximized detection likelihood, minimized detection time, and minimized contaminated demand volume. Performance measures are described in more detail below.

3.1.1 DETECTION LIKELIHOOD

The detection likelihood is the ratio of the number of attack scenarios for which contaminant is detected by any sensor in a given sensor set to the total number of attack

scenarios S_{opt} simulated to conduct optimization. Thus, detection likelihood is computed with the equation

$$Z_{lik} = \frac{1}{S_{opt}} \sum_{s=1}^{S_{opt}} d_s \tag{3.1}$$

where Z_{lik} is the detection likelihood performance measure, s designates a particular attack scenario, and d_s is a binary variable equaling 1 or 0 indicating detection or lack of detection, respectively, of contaminant by the given sensor set during scenario s. The total number of attack scenarios S_{opt} must be sufficiently large in order to minimize the error in the detection likelihood estimate introduced by the random selection of attack scenarios. Scenario selection is discussed further in sections below.

3.1.2 EXPECTED DETECTION TIME

For an individual sensor assigned to some WDS location (i.e., WDS node), detection time is defined as the time elapsed from the earliest injection of a contaminant into the WDS during a given attack scenario to the detection of the contaminant by the sensor. This detection time with respect to one sensor is represented by the variable t_{is}^{d}, where i denotes the WDS node and s denotes the particular attack scenario.

If contaminant is not detected by a sensor located at node i, or if a sensor is not located at node i, t_{is}^{d} is assumed to equal some large value in order to represent an "infinite" detection time; in this work, the large value is the study period length (discussed below). If a smaller time value were selected, a detection time for a detection instance may

actually be greater than that for a non-detection instance. Greater values would strengthen the correlation between expected detection time and detection likelihood for they would more negatively bias non-detection instances, making detection instances even more desirable and consequently reducing the number of Pareto-optimal solutions resulting from multiobjective optimization.

For an entire set of sensors, the detection time is the earliest time at which at least one of the allocated sensors detects contaminant during a given attack scenario and is represented by the equation

$$t_s^d = \min_s \left(t_{is}^d \right). \tag{3.2}$$

To determine the expected detection time for a set of sensors over many possible attack scenarios, sensor set detection times are found for each of S_{opt} attack scenarios, and the expected value of those detection times is calculated. Thus, the expected detection time performance measure is mathematically expressed as

$$Z_{time} = \frac{1}{S_{opt}} \sum_{s=1}^{S_{opt}} t_s^d \tag{3.3}$$

where Z_{time} represents expected detection time.

3.1.3 EXPECTED CONTAMINATED DEMAND VOLUME

Prior to sensor detection of contaminant associated with a given attack scenario and the subsequent response to the attack, water delivered to customers that exceeds a specific hazardous concentration threshold is considered contaminated. It is assumed that no

water is delivered to customers after response to the attack occurs. Under those conditions, each node *i* in the WDS has a corresponding total demand volume contaminated per attack scenario *s* represented by V_{is}^{d}. Therefore, the total demand volume contaminated for the entire WDS during scenario *s* can be found by the equation

$$V_s^d = \sum_i V_{is}^d \,. \tag{3.4}$$

The expected contaminated demand volume, in turn, is the expected value of the total demand volume contaminated for the entire WDS over the large set of randomly selected S_{opt} scenarios and is represented by the equation

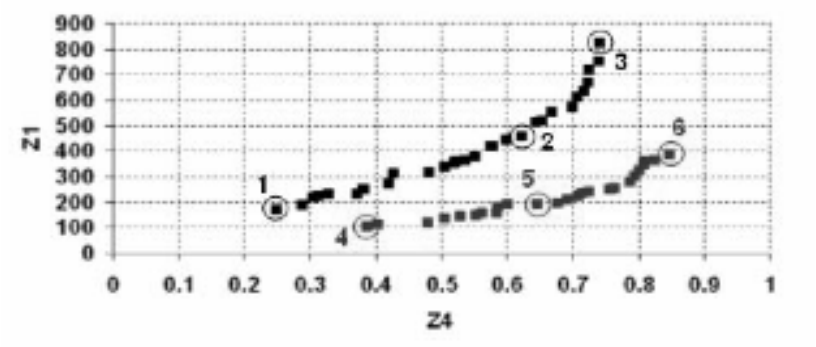
$$Z_{vol} = \frac{1}{S_{opt}} \sum_{s=1}^{S_{opt}} V_s^d \tag{3.5}$$

where Z_{vol} is the expected contaminated demand volume. This performance measure is typically very strongly correlated with the expected WDS population affected by an attack scenario, and the expected population affected is often considered a function of Z_{vol} (e.g., Ostfeld et al. 2006). Thus, in this study the expected population affected is assumed to be "lumped" with Z_{vol} and deemed unnecessary to be determined explicitly.

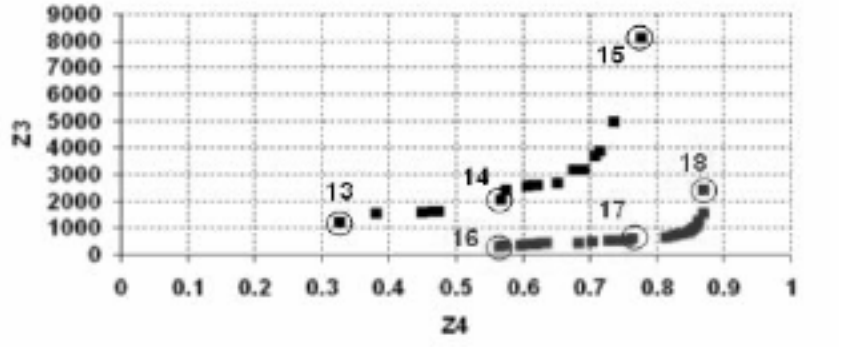
3.1.4 RELATIONSHIPS AMONG PERFORMANCE MEASURES

For this study, it is assumed that the three individual protection goals are of equal precedence relative to each other. A hierarchy among the protection goals can be established by WDS decision-makers, but preferences of certain goals over others are beyond the scope of this study.

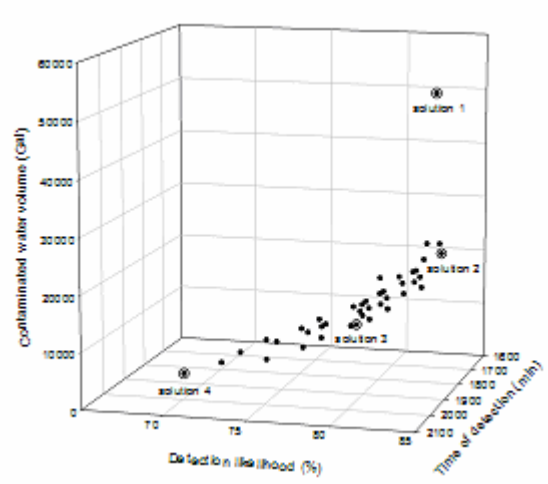
Although the protection goals are of equal relative preference, some goals are inversely related to other goals. WDS protection can improve with respect to one performance measure while protection decreases in terms of another performance measure, as stated in Chapter 1. In general, detection likelihood has inverse relationships with expected detection time and expected contaminated demand volume. These conflicting relationships are illustrated in curves of Pareto-optimal solutions found in several studies. Figure 3.1 provides some such Pareto-front curves from the works of Ostfeld and Salomons (2006) and Aral et al. (2008).



(a)



(b)



(c)

Figure 3.1 Examples of performance measure Pareto fronts. (a) expected detection time (Z1) & detection likelihood (Z4), Ostfeld and Salomons (2006); (b) expected contaminated demand volume (Z3) & detection likelihood, Ostfeld and Salomons (2006); (c) expected detection time, expected contaminated demand volume, & detection likelihood, Aral et al. (2008).

As the figure implies, tradeoffs in protection with regard to individual performance measures may need to occur in order to satisfy adequately all three equally-weighted performance measures of this study. For instance, it may be necessary to sacrifice some protected water demand volume in order to increase the detection likelihood significantly. With regard to actual sensor placement, this tradeoff could mean moving sensors away from WDS nodes or regions of higher demand in order for the sensors to monitor nodes or regions that are more often associated with contaminated flows.

3.2 COST MEASURES

Two types of costs are considered in this study: (i) those that are expected in actually placing sensors at locations in a WDS and (ii) those that are computational in nature resulting from carrying out the optimization of sensor placement. These costs are explained below.

3.2.1 COSTS ASSOCIATED WITH SENSOR PLACEMENT

The number of sensors employed in WDS monitoring (represented by the variable M) is the only cost variable considered in this study. Thus, the only limiting factor on sensor placement in this work is the number of sensors available for placement (M_{max}). WDS contaminant sensor technology is only emerging currently, so estimating sensor operation costs is not feasible at present. Without information regarding the physical characteristics of individual WDS nodes, the labor costs associated the geographic

location of the WDS, and other WDS-specific variables that would affect sensor implementation, it is very difficult to estimate with any accuracy the costs of sensor implementation. It seems reasonable that sensor implementation and operation costs would not vary substantially among WDS nodes and that the cost of a sensor itself would dominate the costs of implementation and operation over a reasonable period of time, so in this study the number of sensors employed is considered a lumped cost variable that implicitly accounts for sensor implementation and operation costs. The numbers of sensors available for placement in this study is either 5 or 20, as adopted from the Battle of the Water Sensor Networks competition (Ostfeld et al. 2006).

3.2.2 COMPUTATIONAL COSTS

The total computational memory and runtime requirements for sensor placement computational runs are reported in this thesis. It would be ideal to directly compare the memory and runtime requirements of this study with those of other documented studies. However, the reporting of those computational expense metrics in technical literature is sporadic and often inexact as implied by discussion in Chapter 2, so it is not entirely feasible to quantitatively compare computational expense between this study and another study. Also, computational expense is a function of the sensor placement methods employed as well as variables independent of those methods, such as computational platform and the design of programming code. In order to navigate these obstacles, the exact measurements of memory and runtime for this study are compared only qualitatively with memory and runtime requirements reported in other studies.

3.3 SIMULATED ATTACK CONDITIONS

Given the wide array of possible attack conditions implied by discussion in Chapter 1, it is necessary to confine the domain of attack possibilities while preserving the realistic nature of the study in order to provide for manageable study. As justified in Chapter 2, most attack conditions are patterned after the Battle of the Water Sensor Networks competition (Ostfeld et al. 2006) in order to provide some basis for comparison of sensor protection performance between this study and multiple other studies.

3.3.1 ELIGIBLE ATTACK LOCATIONS

Attacks are assumed to occur at WDS nodes only as nodes are on average more accessible for contaminant injection than are links in the WDS. In addition, all nodes in the WDS are equally eligible for attack. In reality, a WDS node may be physically situated so that it would be impossible to inject contaminant into the WDS at the node. This impossibility would probably only apply to a small minority of influential nodes, so it is considered negligible. Also, it is realistic that a node would be more desired by terrorists for attack due to better access to the node, high flow through the node, or the location of the node that may allow for greater psychological damage for WDS customers. Quantifiable risk data regarding node attack desirability is currently unavailable and therefore was not taken into account in this study.

3.3.2 ELIGIBLE ATTACK TIMES

In reality, an attack can occur at any time, creating an infinite domain of possible attack times. However, this domain must be constrained in order to make computational simulation feasible but allow enough time for contaminant to propagate throughout the system so that realistic values of performance measures can be found. Thus, attacks can only occur at five-minute multiples during the first quarter of a designated study period.

3.3.3 SCENARIO COMPOSITION AND SELECTION

A given attack scenario may take one of the following forms:

- one WDS node receiving a contaminant injection at one time,
- multiple nodes receiving injections at one time, or
- multiple nodes receiving injections at multiple times.

Therefore, the following types of attack scenarios are considered in this study for their effects on sensor placement:

- one node receiving an injection at one time,
- two nodes receiving injections at one time, and
- two nodes receiving injections at two different times.

The one-injection-node case is considered the "default" case for this study, while the two-injection-node cases are considered "variant" cases for the purposes of determining how sensor placement is affected by changing attack conditions. Above, the WDS nodes and times eligible for contaminant injection are described. To account for the uncertainty

regarding what eligible node(s) and time(s) would be selected for injection during a given attack scenario, the injection node(s) and injection time(s) are randomly sampled from discrete, uniform distributions for each scenario *s*.

3.3.4 CONTAMINANT PARAMETERS

A conservative tracer (namely fluoride) acts as the contaminant in this study. Beyond this study, accounting for the physical and chemical properties of real-world contaminants is simply a matter of adjusting constants for hydraulic simulation. Dilution and decay are also issues to consider in employing real-world contaminants, but these issues would not heavily influence WDS sensor placement results due to flow and transport behaviors and relatively short study periods common to WDSs.

Though some variance is possible, it is reasonable to assume that contaminant is injected into the WDS at a constant flow rate and at a constant concentration. The injection flow rate for this study is 125 L/h, and the contaminant concentration of the injection is 230,000 mg/L. Thus, in terms of mass flow, contaminant enters the system at approximately 479 g/min. This mass flow rate is extremely high, but the high rate allows for the effects of contamination on the WDS to be seen more clearly.

Another unknown of the contaminant injection that is examined in this study is the duration of the contaminant injection. It is reasonable to assume that contaminant is injected continuously for whatever duration chosen. Thus, the effects of a continuous 2-

hour injection (the default case) and a 10-hour injection (the variant case) are analyzed in this study.

Even though a contaminant may be present at a nonzero concentration at a WDS node, the contaminant concentration may be less than what is considered hazardous as implied above in the discussion of the Z_{vol} performance measure. To ascertain the sensitivity of sensor placement to the hazard concentration threshold, optimization was performed for two different threshold values: 0.3 mg/L (the default case) and 0.0 mg/L (the variant case).

3.3.5 SENSOR CAPABILITIES

A sensor at a WDS node may not be able to detect contaminant, although the contaminant may be present at the node in question. To elaborate, the contaminant concentration may be lower than the detection resolution of a sensor, and therefore the sensor would not register the contaminant as present at the node. To determine the possible effects on sensor placement of sensor resolution, optimization was conducted for detection resolution values of 0.0 mg/L (the default case) and 0.3 mg/L (the variant case). The tested hazard concentration thresholds given above and sensor resolution values are the same sets of values in this study; these sets were assumed to have identical values entirely for study convenience.

It is assumed that sensors register detection instantly upon experiencing contaminant concentration at or above the detection resolution of the sensor. There may be a small difference in time between contaminant presence and the registering of detection by the sensor, but this difference would be negligibly small for most cases.

3.3.6 RESPONSE AFTER CONTAMINANT DETECTION

Response to an attack is defined in this study as the actions taken to ensure that no contaminated water is delivered to WDS customers after detection during a given attack scenario. For instance, it may be the practice of a WDS to automatically or manually shut down the WDS entirely at once or in a piecewise fashion after contaminant detection. Response time can vary a great deal given the diversity in the possible methods used by individual WDSs for responding to an attack. In most of the studies described in Chapter 2, response is assumed to occur instantaneously after detection for computational ease and due to a lack of real-world response data. However, given current technological limits, it seems logical to account for response time in some manner. Thus, sensors are placed in this study in trials assuming instantaneous response (the default case) and trials assuming a 3-hour response delay (the variant case).

3.4 STUDY SYSTEMS

The sensor placement methods developed in this study were applied to three study systems of different characteristics in order to fully demonstrate the robustness of the methods.

3.4.1 BWSN NETWORK 1

"BWSN Network 1" is one of two hypothetical networks created by Ostfeld et al. (2006) for the BWSN competition discussed previously. This 129-node system is relatively small and manageable for flow and transport modeling and sensor set testing, but a lack of homogeneity contributes to flow and transport behavior complexity and, in turn, the potential for high sensitivity of sensor placement to WDS hydraulic and water quality conditions. This lack of homogeneity takes the form of spatial clustering of nodes, large variances in hydraulic demand and pipe length, a large number of valves, and localized flow behavior. The assumed study period of the system for this study is 96 hours which allows time in a given attack scenario for ample contaminant propagation throughout the WDS.

The physical layout of BWSN Network 1 is depicted in Figure 3.2 as plotted by EPANET (Rossman 2000). Table 3.1 provides the key information regarding the system and its components.

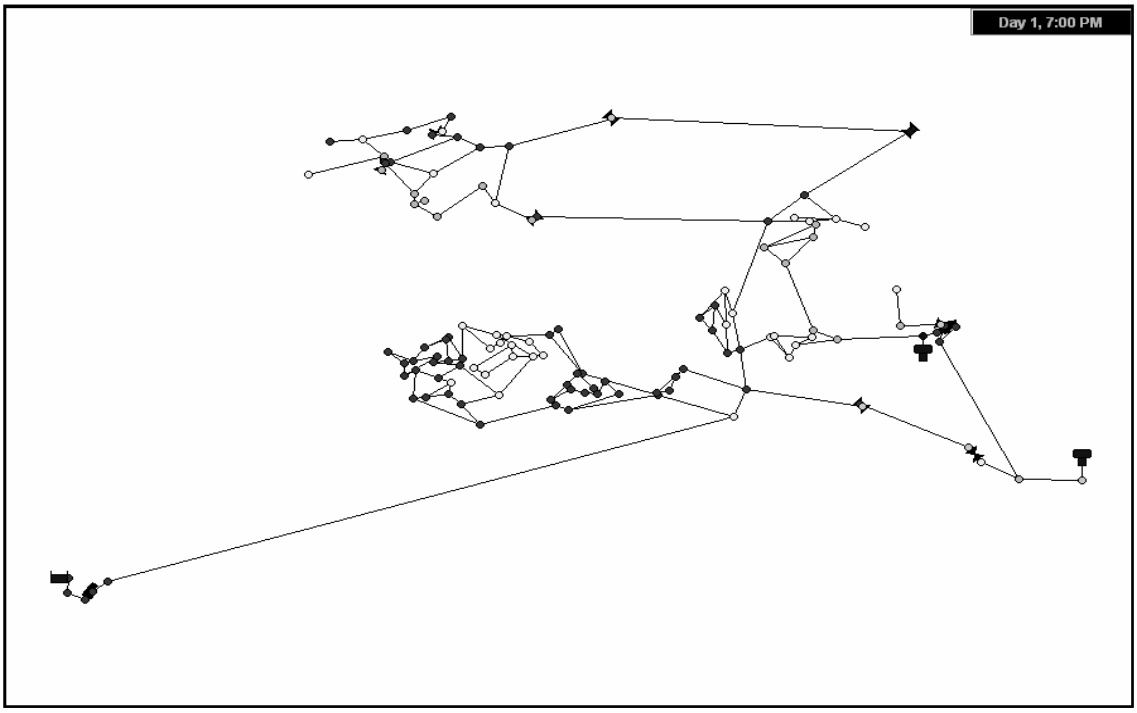


Figure 3.2 BWSN Network 1.

Table 3.1 Characteristics of BWSN Network 1.

Component	Attribute	Value
junctions	number	126
	number of (percentage of) junctions with demand	79 (63%)
base demand	maximum	198 gpm
	mean	7.5 gpm
	standard deviation	24.4 gpm
pipes	number	168
	length minimum	50 ft
	length maximum	10,260 ft
	length mean	733 ft
	length standard deviation	1102 ft
reservoirs	number	1
tanks	number	2
pumps	number	2
valves	number	8
overall system	approximate coverage area	7 mi^2
	study period	96 h

3.4.2 BWSN NETWORK 2

Though BWSN Network 1 provides a set of interesting challenges for sensor placement, it is not large enough for demonstrating the robustness of the developed sensor placement methods in application to systems of urban scale that are of primary interest. The other network provided in the BWSN competition, "BWSN Network 2", with 12,527 nodes is of the necessary scale. As inferred from Figure 3.3 and Table 3.2, BWSN Network 2 is much more homogeneous than BWSN Network 1 in terms of physical layout, demand profile, and the number of special hydraulic devices (e.g., valves) employed. BWSN Network 2 exhibits some localized flow behavior, but localized flow exists to proportionally a much lesser degree than it does for BWSN Network 1. As in the case of BWSN Network 1, a 96-hour study period is associated with BWSN Network 2.

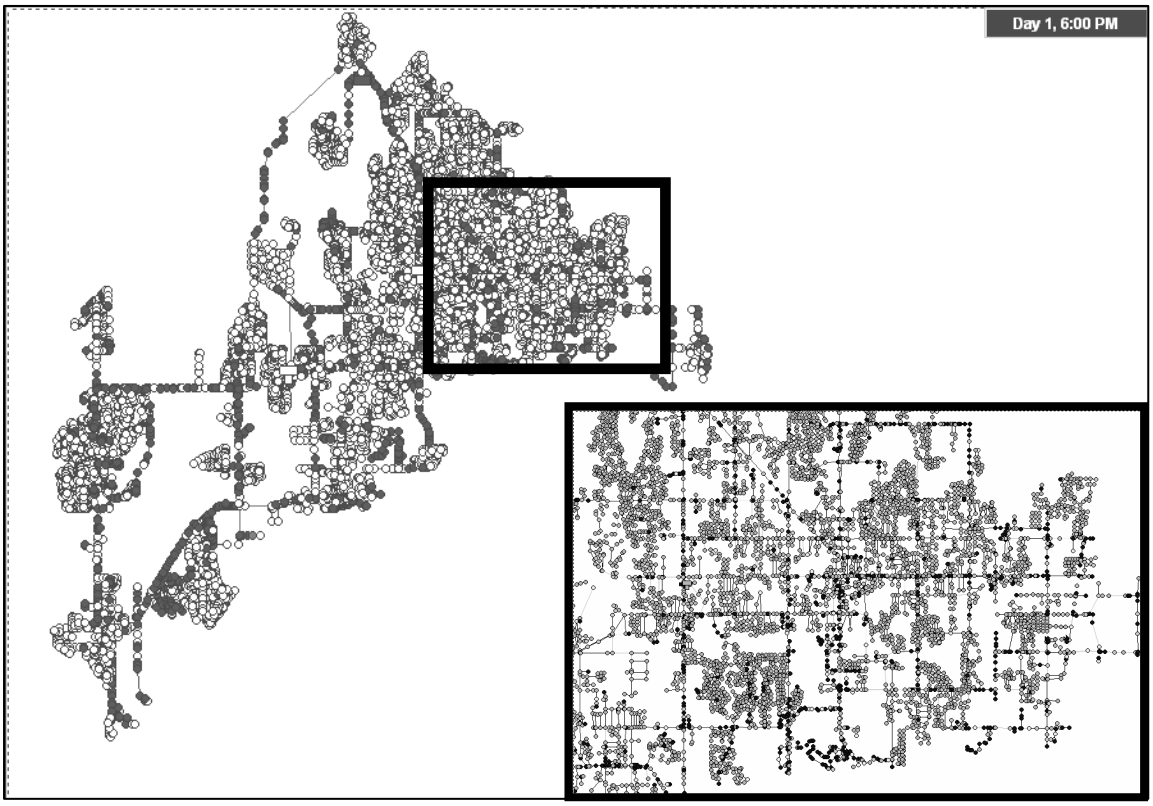


Figure 3.3 BWSN Network 2. The part of the system depicted in the inset box is the part encased in the box in the overall view of the system.

Table 3.2 Characteristics of BWSN Network 2.

Component	Attribute	Value
junctions	number	12,523
	number of (percentage of) junctions with demand	10,551 (84%)
base demand	maximum	247 gpm
	mean	1.6 gpm
	standard deviation	3.6 gpm
pipes	number	14,822
	length minimum	3.8 ft
	length maximum	13,185 ft
	length mean	408 ft
	length standard deviation	346 ft
reservoirs	number	2
tanks	number	2
pumps	number	4
valves	number	5
overall system	approximate coverage area	400 mi^2
	study period	96 h

3.4.3 TOMS RIVER WDS, DOVER TOWNSHIP, NEW JERSEY

The Toms River WDS is a real-world system located in Dover Township, New Jersey. Dover Township is depicted in the map of Figure 3.4 below. Toms River was selected for this study for it is of the desired urban scale and real-world data exists for it. Previous study involving this system provided all necessary data regarding system configuration, nodal demand profile, and flow and transport behavior (Maslia et al. 2000; Maslia et al. 2001). Data corresponding to July 1996 was used to simulate system conditions as (i) 1996 is the latest year for which data is available and (ii) July is a peak demand month for Toms River during which interesting flow behavior occurs in the system. The system is slightly larger than BWSN Network 2 in terms of nodes at 14,973 nodes, but it consumes much less land area than does BWSN Network 2. It serves approximately 44,000 people that reside in Dover Township, justifying its consideration as a larger system of interest by EPA based on criteria stated in Chapter 1. Like BWSN Network 2, Toms River exhibits homogeneous qualities as Figure 3.5 and Table 3.3 imply. It also exhibits some localized flow behavior and has a 96-hour study period.

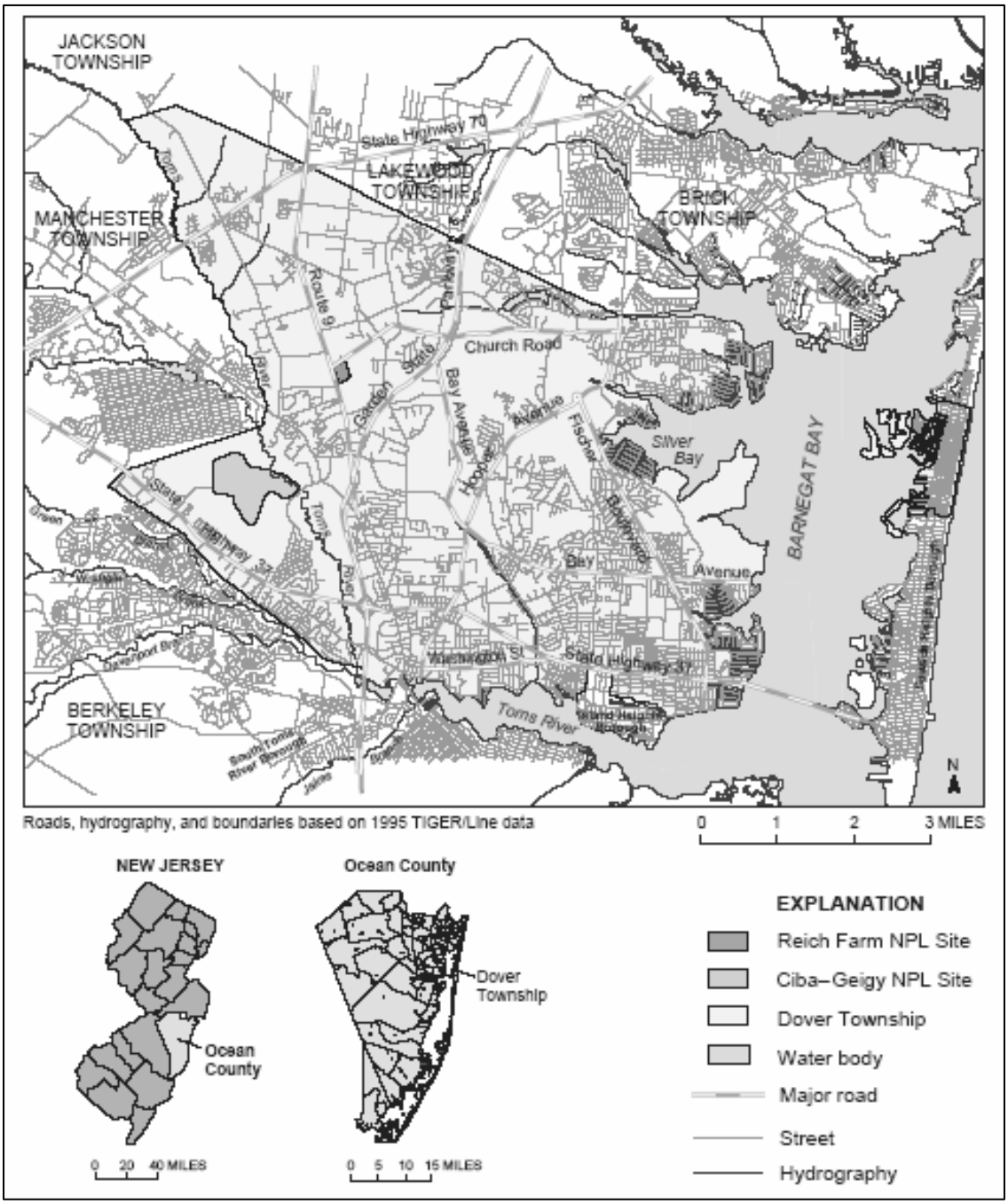


Figure 3.4 Dover Township, New Jersey. (Maslia et al. 2001)

Figure 3.5 The Toms River WDS. The part of the system depicted in the inset box is the part encased in the box in the overall view of the system.

Table 3.3 Characteristics of the Toms River WDS.

Component	Attribute	Value
junctions	number	14,965
	number of (percentage of) junctions with demand	13,287 (89%)
base demand	maximum	8.96 gpm
	mean	0.36 gpm
	standard deviation	0.38 gpm
pipes	number	16,048
	length minimum	1.0 ft
	length maximum	2813 ft
	length mean	158 ft
	length standard deviation	181 ft
reservoirs	number	0
tanks	number	8
pumps	number	12
valves	number	0
overall system	approximate coverage area	80 mi^2
	study period	96 h
	approximate population served	44,000

CHAPTER 4

SOLUTION APPROACH

To solve the study problem specified in Chapter 3 for the variety of systems and conditions described, the methodology explained in this chapter was employed.

4.1 GENERAL FORMULATION OF OPTIMIZATION PROBLEM

In Chapter 3, WDS protection goals for this thesis were mathematically described and the constraints on sensor placement in providing WDS protection were stated. Optimal locations for sensor placement can be determined by solving an optimization program that includes those objectives and criteria. The generic form of the multiobjective optimization program incorporating all three protection goals in a three-prong objective function using notation defined in Chapter 3 is given in (4.1).

$$\begin{aligned} &\max \quad Z_{lik} \\ &\min \quad Z_{time} \\ &\min \quad Z_{vol} \\ &\text{s.t.} \quad M = M_{max} \end{aligned} \tag{4.1}$$

In the given formulation, sensors are placed to minimize expected detection time, minimize expected contaminated demand volume, and maximize detection likelihood simultaneously while ensuring that the numbers of sensors employed does not exceed the number of sensors available. The optimization problem may be adjusted to exclude one prong or two prongs of the objective function if only two protection goals or one protection goal, respectively, are being considered. In this thesis, the optimization

problem is solved for the cases of considering only one protection goal and of all three protection goals at once.

4.2 OVERVIEW OF SOLUTION APPROACH

Solving the multiobjective optimization program to adequately satisfy all three prongs of the objective function concurrently is typically a non-trivial problem, even for smaller systems. In the case of a large WDS and relatively limited sensor availability, solving the problem could be especially difficult for the domain of possible sensor set solutions would be quite large. Therefore, it is advantageous to focus the solution search in a way that allows for both desirable values for the three performance measures and reduced computational expense required in solving the optimization program. Attaining this focus is the primary objective of the work of this thesis and is accomplished through the implementation of the "nodal importance concept."

The overall solution approach is explained in a broad sense in this section, and the specific components of the approach are explained in greater detail in sections below. In general, the solution of a sensor placement problem in this work consists of two main phases:

- determining the "importance" of WDS nodes (i.e., the "importance" phase), and
- the actual optimization to find the best-performing sensor nodes (i.e., the "optimization" phase).

Both of these phases are employed in solving any single-objective or multiobjective variant of the optimization problem in (4.1). The importance phase consists of computing importance values for all WDS nodes using functions each designed specifically to address a particular protection goal. The optimization phase consists of ranking nodes according to importance and using the Iterative Subset Search Method developed in this work in combination with genetic algorithms to solve the optimization problem. Before the two phases are carried out, the appropriate single-objective or multiobjective optimization problem is defined according to the protection goal or goals of interest and WDS flow and transport data are generated for the study system in question. Figure 4.1 illustrates the sequence of these major steps of the solution approach.

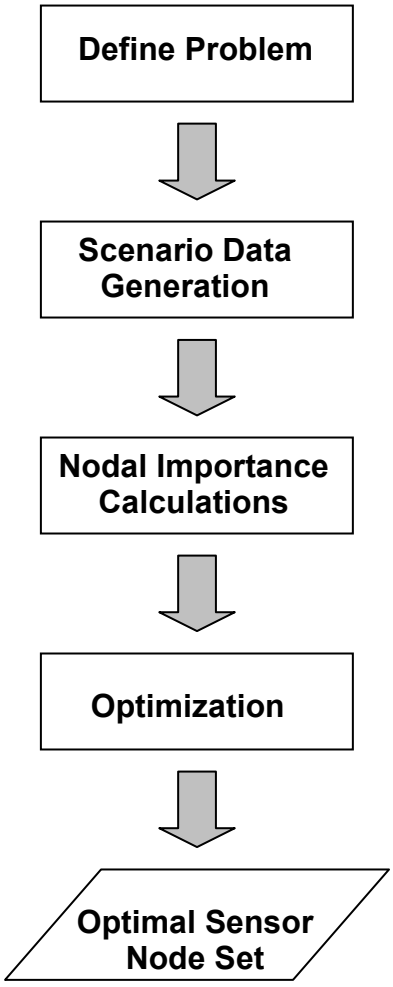


Figure 4.1 Sensor placement optimization problem solution approach overview.

4.3 SPECIFIC OPTIMIZATION PROBLEM DEFINITIONS

In this study, optimization of sensor placement according to both one particular projection goal and multiple protection goals concurrently is of interest. The problem formulations governing optimization for these objectives are provided below.

4.3.1 SINGLE-OBJECTIVE PROBLEMS

If only one protection goal is considered, the multi-prong objective function in (4.1) reduces to a simple single-prong function. For example, in the case of only maximizing detection likelihood, the optimization problem formulation takes the form seen in (4.2).

$$\begin{aligned} &\max \quad Z_{lik} \\ &\text{s.t.} \quad M = M_{max} \end{aligned} \tag{4.2}$$

Likewise, for only minimizing expected detection time, the formulation becomes the one in (4.3).

$$\begin{aligned} &\min \quad Z_{time} \\ &\text{s.t.} \quad M = M_{max} \end{aligned} \tag{4.3}$$

For only minimizing expected contaminated demand volume, the optimization problem becomes the one in (4.4).

$$\begin{aligned} &\min \quad Z_{vol} \\ &\text{s.t.} \quad M = M_{max} \end{aligned} \tag{4.4}$$

4.3.2 MULTIOBJECTIVE PROBLEM

If more than one protection goal is considered, then formulating the optimization problem in a manner that allows for practical solving of the problem is more complex than it is in single-objective cases. In the discussion that follows, the strengths and weaknesses of common approaches for handling multiobjective sensor placement problems are evaluated, and the specific definition of the multiobjective problem for this work that is tailored to take advantage of the strengths of the common multiobjective approaches is described in detail.

4.3.2.1 Analysis of Common Multiobjective Sensor Placement Approaches

Approaches for solving a multiobjective sensor placement problem documented to date generally take one of three forms:

- conversion of the multi-prong objective function into a weighted-sum function that acts a single-prong objective function with a "lumped" protection goal (e.g., Krause et al. 2006; Rogers et al. 2007),
- optimization according to a hierarchy of protection goals through "epsilon-constrained" optimization (e.g., Berry et al. 2006a), or
- generation of a Pareto front of equally optimal solutions relative to the multiple protection goals of interest (e.g., Aral et al. 2008; Ostfeld and Salomons 2006).

While these of approaches all have differing degrees of merit, they also have their respective significant drawbacks.

Weighted-sum optimization provides a straightforward way of proportioning weight to multiple protection goals and concisely yielding a single sensor node set solution through the use of the lumped protection goal. The process to solve the optimization problem with a weighted-sum objective function is typically no more complicated than it would be for a single-objective problem. However, proportioning weight to protection goals does not guarantee that the resulting solution is Pareto-optimal, even though the solutions may be very well-performing. When weighted-sum optimization is conducted in the conventional manner, much iteration of protection goal weights may be required to find a Pareto-optimal solution, and where that solution falls within a range of possible Pareto-optimal solutions is not readily apparent.

Epsilon-constrained optimization accounts for multiple objectives in a way that does concisely provide one Pareto-optimal solution. Protection goals in this optimization approach are addressed in an order of priority determined by human decision. To elaborate, the protection goal considered most important in the determined hierarchy of goals is treated initially as the only protection goal of interest, and the optimization problem is solved as a single-objective problem. For clarity, the solution for this first goal in the hierarchy is referred to as "Solution 1." After finding Solution 1, the next goal in the hierarchy is considered in another single-objective problem to find "Solution 2". The single-objective problem for finding Solution 2 has an additional constraint that Solution 2 loses no more than some specified margin of performance with respect to the more important goal considered to find Solution 1. This specified margin is "epsilon" (ε). For instance, if ε is 5%, Solution 2 should provide at least 95% of the protection with

respect to the more important goal that Solution 1 provides. Solution 2 is a solution on the Pareto front involving the two most important goals in a given hierarchy, and the position of Solution 2 on the Pareto front depends on the value of ε. If there is a third goal in the hierarchy, then the single-objective problem is solved according to the third goal and is epsilon-constrained by the two more important goals. Often, though, a Solution 3 and any subsequent solutions will not differ from Solution 2 due to the increasing number of epsilon constraints. In sum, despite providing a Pareto-optimal solution, epsilon-constrained optimization has trouble handling three or more protection goals simultaneously. Also, epsilon-constrained optimization hinges heavily on human input regarding both the hierarchy of protection goals and the value of ε, potentially allowing human decision to affect the resulting solution greatly.

Construction of a Pareto front has become popular in recent studies as it can provide a wide array of Pareto-optimal solutions for two and even three protection goals in one solution run, in turn quantifying the optimization tradeoffs among protection goals more completely than would be done through either weighted-sum or epsilon-constrained optimization. The Pareto front consists of all solutions that do not dominate each other but dominate all others that do not make up the Pareto front. (In order for one solution to dominate another, the solution in question must (i) perform at least as well as the other solution with respect to all protection goals and (ii) perform better than the other solution does with respect to at least one protection goal.) Regarding the drawbacks of this approach, the construction and subsequent illustration of the Pareto front can become unmanageable for multiobjective problems with more than three protection goals. A

Pareto front gives an extensive set of solutions, but it indicates no preference for certain Pareto-optimal solutions over others, even though some of the Pareto-optimal solutions may be more desirable in the practical sense than others. For instance, a boundary point of the Pareto front is a Pareto-optimal solution; this solution may provide a very desirable level of protection with regard to one protection goal but undesirable levels of protection with regard to other goals. However, a more-centered point on the Pareto front may provide desirable levels of protection with respect to all protection goals simultaneously.

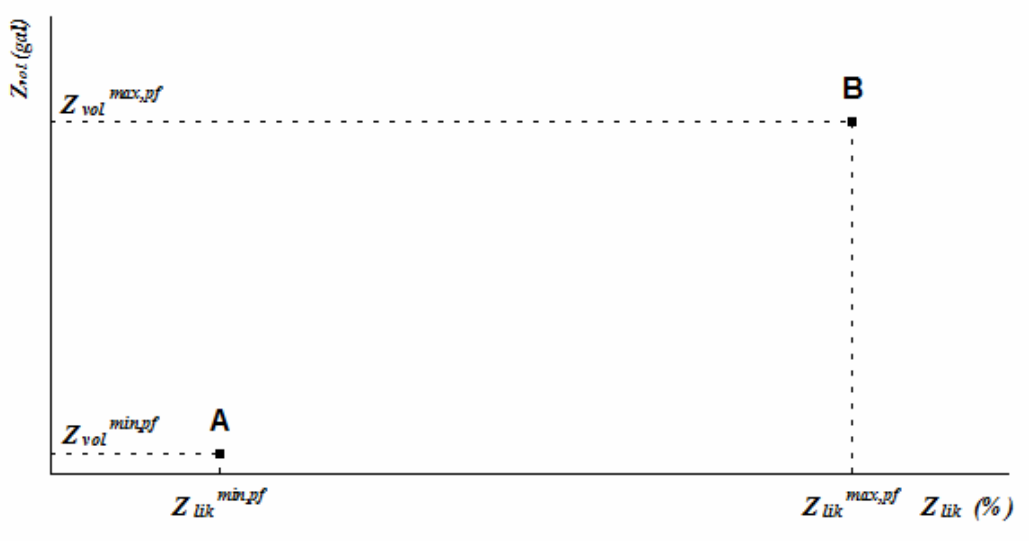
4.3.2.2 Formulation for This Work

In this work, a multiobjective problem formulation was developed to incorporate the strengths cited for the three approaches described above into one approach. More specifically, the overarching intention of this developed formulation is to concisely provide one sensor placement solution that best accounts for and minimizes the tradeoffs among any number of protection goals. Below, the formulation is derived with the aid of a two-prong multiobjective example—maximizing detection likelihood and minimizing expected contaminated demand volume—as an example with three or more prongs would pose problems in illustrating the derivation of the formulation effectively. After the formulation for the two-prong example is derived, the ways in which the formulation is extended to cases with three or more protection goals are discussed.

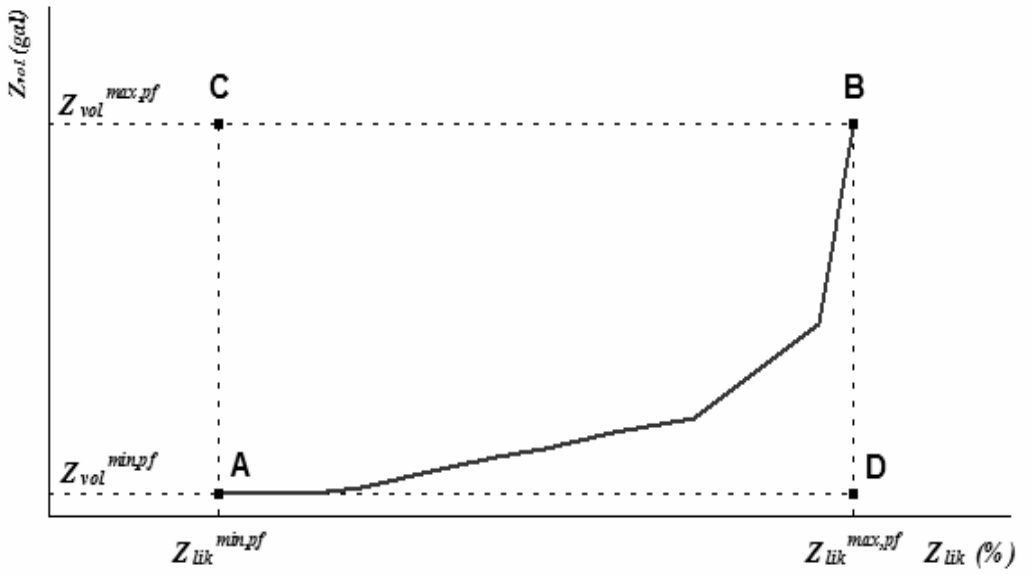
4.3.2.2.1 The Pareto Front Space

The formulation requires that the "Pareto front boundary points" are first identified. These points represent performance measure values of minimum and maximum possible protection with respect to particular protection goals found in solving single-objective optimization problems. Figure 4.2 (a) depicts generically where these points may lie with respect to each other in plotting a two-dimensional Pareto front. The boundary point A in Figure 4.2 represents the solution yielding both the minimum possible detection likelihood $\left(Z_{lik}^{min,pf}\right)$ and minimum possible expected contaminated demand volume $\left(Z_{vol}^{min,pf}\right)$ for a Pareto-optimal solution. Point B represents the solution yielding both the maximum possible detection likelihood $\left(Z_{lik}^{max,pf}\right)$ and maximum possible expected contaminated demand volume $\left(Z_{vol}^{max,pf}\right)$. Values for $Z_{lik}^{max,pf}$ and $Z_{vol}^{min,pf}$ are found by solving the single-objective problems for maximizing detection likelihood in (4.2) and minimizing expected contaminated demand volume in (4.4), but the values of $Z_{lik}^{min,pf}$ and $Z_{vol}^{max,pf}$ are the performance measure values for the protection goals not considered in the single-objective optimization problems. In other words, $Z_{lik}^{min,pf}$ is the value of Z_{lik} at $Z_{vol}^{min,pf}$, and $Z_{vol}^{max,pf}$ is the value of Z_{vol} at $Z_{lik}^{max,pf}$. Boundary points may actually be part of the Pareto front as shown in Figure 4.2 (b), but that is not necessarily the case. However, Pareto front points should not exist beyond the boundaries indicated by the boundary point coordinates.

The Pareto front is confined within a "space" defined using Pareto front boundary point information. In Figure 4.2 (b), the rectangle ACBD represents the Pareto front space (or, more specifically, the Pareto front "area" in the two-dimensional case) for the two-prong problem of interest in this discussion. This Pareto front space allows for better evaluation of the Pareto-optimality potential of a solution and, in turn, the potential for a solution to tradeoff among individual protection goals effectively.



(a)



(b)

Figure 4.2 Pareto front boundary points and the space containing the Pareto front for the generic two-prong multiobjective problem of maximizing detection likelihood and minimizing expected contaminated demand volume. The hypothetical Pareto front shown in (b) is an example of the possible shape the Pareto front may take.

4.3.2.2.2 The Lumped Protection Goal Performance Measure

With knowledge of the ranges of possible protection goal performance measure values for Pareto-optimal solutions, a means of quantifying the tradeoffs among protection goals associated with particular multiobjective solutions is beneficial to evaluate the degrees to which solutions sacrifice protection with respect to certain protection goals in order to accommodate multiple individual protection goals. With this means of quantifying tradeoffs, one solution can be identified that minimizes tradeoffs by maximizing the protection provided by a sensor node solution according to all protection goals collectively.

A "lumped" protection goal performance measure inspired by weighted-sum optimization is used for quantifying the collective protection provided by a solution. In this work, this lumped performance measure for any number of particular protection goals is referred to as Z_{all}. For the two-prong multiobjective problem defined above, the formula to find Z_{all} is

$$Z_{all} = Z_{lik}^{norm} + Z_{vol}^{norm} \tag{4.5}$$

where Z_{lik}^{norm} and Z_{vol}^{norm} are normalized terms for the detection likelihood and expected contaminated demand volume corresponding to a given feasible solution. .

Z_{lik}^{norm} and Z_{vol}^{norm} are Z_{lik} and Z_{vol} values normalized according to boundary point information. For solutions inside of the Pareto front space (rectangle ACBD in Figure 4.2 (b)), Z_{lik}^{norm} and Z_{vol}^{norm} each fall within the range [0, 1]. A value of 0 for one of the

terms translates to minimum protection with regard to a protection goal in question relative to all Pareto front solutions, while a value of 1 translates to maximum protection. For solutions falling outside of the Pareto front space, the terms are negative in value. Clearly, these outside solutions are not eligible to be optimal solutions, but assigning null fitness to all of these solutions may cause the genetic algorithms used in optimization to converge unnecessarily to one of the obviously suboptimal, null fitness solutions.

Given the above descriptions of the terms in (4.5), (4.5) can be rewritten as

$$Z_{all} = \frac{Z_{lik} - Z_{lik}^{min,pf}}{Z_{lik}^{max,pf} - Z_{lik}^{min,pf}} + \frac{Z_{vol}^{max,pf} - Z_{vol}}{Z_{vol}^{max,pf} - Z_{vol}^{min,pf}} \tag{4.6}$$

At either boundary point (A or B), Z_{all} is equal to 1 since the possible protection provided by a Pareto-optimal solution with respect to one of the protection goals is completely maximized while protection for the other goal is completely minimized. At point C, protection is completely minimized for both goals, so Z_{all} would equal 0. At point D, protection for both goals would be completely maximized, so Z_{all} would equal 2. Point D, though, would not be a realistic solution; it would be an ideal Pareto front consisting of only one solution that would maximize protection for both protection goals. (Feasible Pareto fronts consist of only one point when all protection goals in question are extremely correlated with each other. In such a case, multiobjective optimization is not necessary as the problem can be solved as a single-objective problem with the single-prong objective function addressing all protection goals. Also, (4.6) would be undefined for the particular case, and therefore the existence of this extreme correlation should be avoided in employing the multiobjective formulation of this work.)

Figure 4.3 gives a graphical indication of the ranges of possible Z_{all} values for solutions in the various areas of the Pareto front plot for the two-prong problem. It is noteworthy that on the line consisting of boundary points A and B, all solutions have a Z_{all} value of 1, as the trading-off of protection according to the normalized performance measures is occurring in a linear fashion such that the sum of Z_{lik}^{norm} and Z_{vol}^{norm} is always equal to 1. Pareto fronts for this particular two-prong problem are likely to exist within the triangle ABD in Figure 4.3, so the multiobjective tradeoff solutions found should have values no less than 1 and no greater than 2.

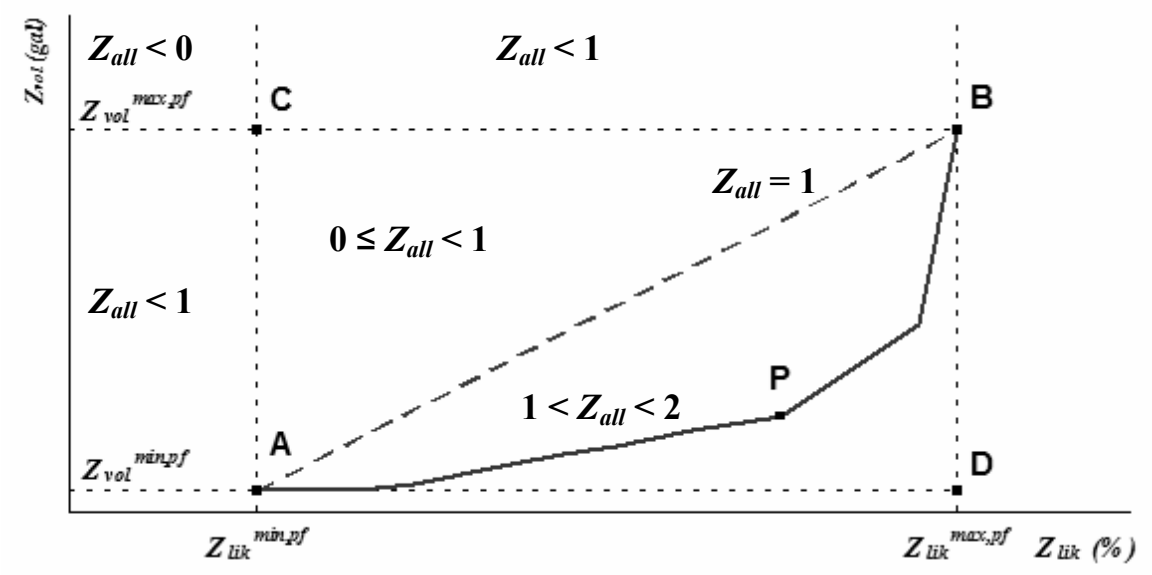


Figure 4.3 Possible ranges of Z_{all} values for solutions within particular areas of the Pareto front plot for the generic two-prong multiobjective problem of maximizing detection likelihood and minimizing expected contaminated demand volume. The point P on the Pareto front represents the hypothetical solution for which Z_{all} is maximized.

4.3.2.2.3 Converted Optimization Problem

To find the solution with the maximum possible value of Z_{all} (represented generically by point P in Figure 4.3 for the two-prong problem), the generic optimization problem in (4.1) is converted to the single-prong objective problem in (4.7) using Z_{all}.

$$\begin{aligned} \max \quad & Z_{all} \\ \text{s.t.} \quad & M = M_{max} \end{aligned} \tag{4.7}$$

In sum, (4.1) reduces to a relatively simple two-step problem with the developed approach. First, the single-objective problems corresponding to the protection goals of interest are solved to determine values of terms needed in the formula for Z_{all}, and then (4.7) is solved.

4.3.2.2.4 Extending Formulation for More than Two Protection Goals

A set of simple adjustments can be made to adapt the multiobjective problem formulation derived above for more than two protection goals. For a two-prong problem, the coordinates for boundary points are found by solving the single-objective problems for the two protection goals of interest and using performance measure information in the straightforward manner described above. For a multiobjective problem with three or more prongs, the identification of boundary point coordinates is not as straightforward. In these cases, performance value information from the relevant single-objective problems is still employed. As in the case of the two-prong problem, performance measure values for the protection goals of interest in the single-objective problems are boundary point coordinates. However, there are multiple candidates for all of the other

boundary point coordinates since three or more single-objective problems are being solved. Hence, the other coordinates are found by finding the possible Pareto front "worst-case" performance measures. For example, in the case of the three-prong multiobjective problem of this work, $Z_{lik}^{max,pf}$ is found as it would be found in a two-prong problem by solving the corresponding single-objective problem in (4.2). There are two candidates, though, for $Z_{lik}^{min,pf}$ —Z_{lik} at $Z_{time}^{min,pf}$ and Z_{lik} at $Z_{vol}^{min,pf}$. The lesser value of Z_{lik} of the two candidates (i.e., the worst-case Z_{lik}) would be $Z_{lik}^{min,pf}$. The formulas in (4.8) and (4.9) summarize the process of finding the Z_{lik} boundary point coordinates for the three-prong problem in (4.1).

$$Z_{lik}^{max,pf} = \max Z_{lik} \tag{4.8}$$

$$Z_{lik}^{min,pf} = \min\left(Z_{lik} \middle| Z_{time} = Z_{time}^{min,pf}, Z_{lik} \middle| Z_{vol} = Z_{vol}^{min,pf} \right) \tag{4.9}$$

Boundary point coordinates relating to the other two protection goals in the three-prong problem are found in similar fashion as shown in the formulas in (4.10) through (4.13).

$$Z_{time}^{min,pf} = \min Z_{time} \tag{4.10}$$

$$Z_{time}^{max,pf} = \max\left(Z_{time} \middle| Z_{lik} = Z_{lik}^{max,pf}, Z_{time} \middle| Z_{vol} = Z_{vol}^{min,pf} \right) \tag{4.11}$$

$$Z_{vol}^{min,pf} = \min Z_{vol} \tag{4.12}$$

$$Z_{vol}^{max,pf} = \max\left(Z_{vol} \middle| Z_{lik} = Z_{lik}^{max,pf}, Z_{vol} \middle| Z_{time} = Z_{time}^{min,pf} \right) \tag{4.13}$$

With these six values, the Pareto front space can be defined. In three-space, the Pareto front space would take the form of a rectangular prism as opposed to the rectangle associated with the two-prong problem illustrated in Figure 4.2 (b). As indicated

previously, the boundary points may not necessarily lie on the Pareto front itself, but they do at least provide for boundaries within which the Pareto front is contained.

The formula in (4.5) to determine Z_{all} can be extended with ease to accommodate more than two protection goals. The generic form of the formula for some specified number of protection goals is

$$Z_{all} = Z_1^{norm} + Z_2^{norm} + \cdots + Z_g^{norm} + \cdots + Z_G^{norm} \tag{4.14}$$

where g is the index for a particular protection goal, G is the number of protection goals for a given multiobjective problem.

Thus, for the three-prong multiobjective problem at the center of this work, the formula to determine Z_{all} for solutions within the Pareto front space becomes

$$Z_{all} = \frac{Z_{lik} - Z_{lik}^{min,pf}}{Z_{lik}^{max,pf} - Z_{lik}^{min,pf}} + \frac{Z_{time}^{max,pf} - Z_{time}}{Z_{time}^{max,pf} - Z_{time}^{min,pf}} + \frac{Z_{vol}^{max,pf} - Z_{vol}}{Z_{vol}^{max,pf} - Z_{vol}^{min,pf}}. \tag{4.15}$$

The value of Z_{all} calculated with (4.15) falls in the range [0, 3) for solutions within the Pareto front space.

4.4 ATTACK SCENARIO DATA GENERATION

EPANET Version 2.0 (Rossman 2000) was used to generate hydraulic and contaminant transport data for each study system. EPANET requires calibrated data on WDS components, connectivity, water sources, nodal demands, flow patterns, and other WDS-related state variables to carry out modeling; these data existed for all three study systems

prior to this work. For an individual system, independent sets of data were generated for the importance phase and the optimization phase; data were generated for each of S_{imp} attack scenarios for the importance phase and each of S_{opt} scenarios for the optimization phase. These generated data allows for the degrees to which contaminant is present at particular WDS nodes at given times during the study period to be determined for every attack scenario. As demonstrated in sections below, this information is employed to calculate nodal importance values and determine the effectiveness of sensor node set candidates during optimization.

4.5 NODAL IMPORTANCE

The theoretical fundamentals of the nodal preprocessing approach developed in this study are detailed in this section.

4.5.1 BASIC CONCEPT AND GENERAL APPROACH

The two most computationally-expensive aspects of solving the WDS sensor placement problem are (i) simulating the hydraulic and contaminant transport behavior of the WDS and (ii) the testing of sensor node combinations in order to find the node set providing maximum protection on-average. At present, little can be done to reduce the burden associated with WDS flow and transport simulation; EPANET is the definitive WDS flow and transport model used by most researchers conducting sensor placement optimization studies. The simulation burden is best managed by minimizing the need to

run simulations. The computational expense associated with the testing of sensor node candidates, though, can be addressed much more directly by performing a preliminary examination of each WDS node to gauge its potential fitness as a sensor node candidate relative to the other nodes in the system. Thus, the nodes that are of higher fitness can be identified and can receive more focus in the search for the optimal set of sensor nodes. The nodal importance concept developed in this work allows for such "preprocessing" of WDS nodes by assigning weights to nodes in the importance phase before the actual sensor placement optimization is carried out in the optimization phase.

In this work, the potential fitness of node as a sensor node candidate is considered synonymous with the "importance" of the node. With knowledge of (i) how contaminant moves throughout the system in the S_{imp} scenarios modeled in the scenario data generation described above and (ii) the potential for individual WDS nodes to experience adverse effects resulting from the S_{imp} scenarios, the importance of a particular node in sensor placement consideration is concisely estimated according to a function of "importance variables." A particular importance function is designed to yield one numerical value per node for a given protection goal or set of goals. The specific functions designed for particular protection goals and sets of goals are derived in sections below. In order to ensure that the determined value of importance for each WDS node is sufficiently representative of the true expected importance of the node, a relatively large number of scenarios should be employed in the importance phase. This work employed 3,000 scenarios in the importance phase (i.e., $S_{imp} = 3{,}000$) for each of the three study systems.

The idea of preprocessing WDS nodes for sensor placement consideration has precedent. For instance, the nodal importance concept of this work is not entirely unlike the nodal impact coefficient concept developed by Berry et al. (2006b). However, determining impact coefficients for all WDS nodes as done by Berry et al. requires placing a hypothetical sensor at each node—one node at a time—for every attack scenario simulated and quantifying the protection impacts of placing sensors at individual nodes. The nodal importance concept developed in this work has the advantage of not requiring separate simulations for every WDS node per attack scenario and, therefore, is more computationally efficient. Also, the nodal importance concept better accounts for protection provided by other possible sensor nodes in addition to a particular node in question as will be made apparent in sections below. Another example of the use of preprocessing is the data mining procedure conducted by Huang et al. (2006) to determine detection probability information (specifically the frequency at which contaminant would be first detected by a sensor at a node and the frequency at which contaminant would be detected at a node during a particular period of time after injection) for each WDS node in order to provide optimization algorithms with reasonable initial candidate solutions. As discussed in Chapter 2, the methods of Huang et al. seem very computationally inefficient. Most recently, Xu et al. (2008) use "betweenness centrality" and "receivability" to identify key nodes for sensor placement consideration as described in Chapter 2, but their methods yield suboptimal solutions and have not been applied to larger systems.

4.5.2 NODAL IMPORTANCE FUNCTIONS

This section provides the derivation of the mathematical functions used to calculate nodal importance for the single-objective protection goals and the lumped multiobjective protection goal.

4.5.2.1 Maximizing Detection Likelihood

If the protection goal of interest is maximizing detection likelihood, the importance of a WDS node *i* for an attack scenario *s* is a function of

- the presence or absence of contaminant at or above the sensor resolution at node *i* during scenario *s*, and
- the number of nodes other than node *i* that experience contamination during scenario *s*.

From these two ideas stem the concepts of "frequency," "uniqueness," "end-of-line" nodes, and "slow-outflow" nodes that were used to formulate nodal importance functions.

4.5.2.1.1 Frequency

"Frequency" refers to the expected proportion of scenarios for which contaminant is present at node *i* over S_{imp} scenarios. Mathematically, this proportion is expressed as

$$FREQ_i^{lik} = \frac{1}{S_{imp}} \sum_{s=1}^{S_{imp}} D_{is} \tag{4.16}$$

where D_{is} is equal to 1 or 0, indicating the presence or absence, respectively, of contaminant at node *i* at any time within the study period during scenario *s*.

The logic behind the relevance of this frequency measurement to nodal importance is intuitive. If a sensor placed at node *i* is expected to detect contaminant for a significant number of attack scenarios, then contaminant should be present at node *i* for a significant number of scenarios.

4.5.2.1.2 Uniqueness

Frequency alone cannot indicate the importance of a node with respect to maximizing detection likelihood. To elaborate, there is no need to place a sensor at a particular node to detect contaminant for some given number of attack scenarios if the contaminant would be detected anyway for those scenarios by some other sensor located elsewhere in the system, even if the computed frequency of the node in question is relatively high. Instead of placing the sensor at the node in question, it would be more beneficial to place the sensor at another location to detect contaminant for other scenarios not detected by any other sensors in the sensor set.

"Uniqueness" is used to estimate how unique a node is in experiencing contamination relative to the other WDS nodes. The effectiveness of uniqueness as a variable of nodal importance can be made clear with some hypothetical extreme cases. For example, a node that is the only node in the WDS to experience contamination during a given

scenario would be as unique as possible and extremely important with respect to that scenario for it would be the only node in the system at which a sensor could detect contaminant. In contrast, if a node were one of very many that experience contamination during a scenario, the node would be much less unique and much less important as so many other nodes would also be suitable locations for a sensor to detect contaminant.

Uniqueness per node per scenario is quantified by the equation

$$UNIQ_{is}^{lik} = \begin{cases} \dfrac{1}{\sum_j D_{js}} & D_{is} = 1 \\ 0 & D_{is} = 0 \end{cases} \tag{4.17}$$

where j is a specific node in the set of all WDS nodes that includes node i, and D_{js} is equal to 1 or 0 indicating the presence or absence, respectively, of contaminant at node j during scenario s. Put simply, the value of uniqueness in (4.17) is the ratio of one to the total number of WDS nodes experiencing contamination during scenario s if node i experiences contamination during the scenario (i.e., $D_{is} = 1$). If node i does not experience contamination at all during scenario s, then the uniqueness of node i for scenario s is equal to zero.

As implied by (4.17), uniqueness formulated in this manner is inherently a function of frequency for a greater frequency yields a greater number of non-zero nodal uniqueness values over S_{imp} scenarios. Thus, the correlation between uniqueness and frequency is positive, but the strength of the correlation is directly related to the sizes of nodal uniqueness values for scenarios with D_{is} equal to 1.

To obtain an expected value of uniqueness over S_{imp} scenarios, the mean of values for nodal uniqueness per scenario determined by (4.17) is computed using (4.18).

$$UNIQ_i^{lik,avg} = \frac{1}{S_{imp}} \sum_{s=1}^{S_{imp}} UNIQ_{is}^{lik} \tag{4.18}$$

It is also helpful on occasion to know the maximum value of nodal uniqueness per scenario as made evident in discussion below. The mathematical expression for this maximum value is found with (4.19).

$$UNIQ_i^{lik,max} = \max_s \left(UNIQ_{is}^{lik} \right) \tag{4.19}$$

The uniqueness concept is very similar in principle to the "receivability" concept developed by Xu et al. (2008) for which the ultimate goal is to maximize detection "coverage" of the WDS with a fixed number of sensors, consequently minimizing unnecessary sensor detection redundancy. Uniqueness is more advantageous, though, for its value can be found with the simple equations above, whereas receivability requires a multiple-step, iterative procedure that could be computationally expensive for larger systems. Using uniqueness over receivability also appears to yield higher detection likelihood values as will be shown in Chapter 5.

4.5.2.1.3 End-of-Line Nodes

An "end-of-line" node is a node that has positive pipe inflow but has zero pipe outflow. Instead, outflow from the node takes the form of hydraulic demand at a junction or of tank inflow such that hydraulic continuity is preserved. This particular flow behavior in

the case of a junction with demand is illustrated in Figure 4.4 (a). If this flow behavior occurs at a node at any time during the study period, the node is considered an end-of-line node. End-of-lines nodes are relatively straightforward to determine as only a simple hydraulic analysis is required; no contaminant transport analysis is needed. End-of-line junctions always have a maximum uniqueness as formulated in (4.19) equal to 1 for contaminant injection could occur at a particular junction at a time when the junction is experiencing end-of-line conditions and would only affect that junction due to the lack of pipe outflow to other WDS nodes. End-of-line node status for a node *i* can be expressed by the variable EOL_i; EOL_i is equal to 1 if node *i* is an end-of-line node or 0 if otherwise.

End-of-line nodes are the endpoints of possible contaminant transport paths in the WDS. A sensor placed at one of these nodes would eventually detect contamination if contaminant is present anywhere on the node's respective contaminant transport path, assuming that the contaminant when present is present at or above the sensor resolution. Thus, identifying end-of-line nodes can allow for the narrowing of the sensor node set solution search domain while still accounting for all contaminant transport paths in the WDS. In other words, end-of-line nodes would be the only important nodes if the only protection goal of concern is maximizing detection likelihood.

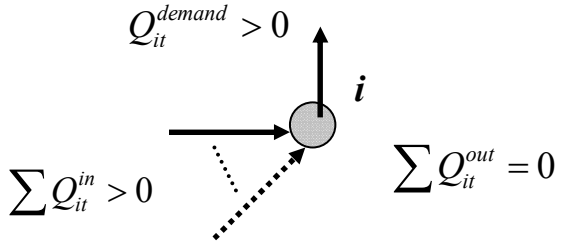
4.5.2.1.4 Slow-Outflow Nodes

A slow-outflow node has positive pipe outflow at the latest time during the study period when contaminant arrives at the node (i.e., the "latest contaminant arrival time" of the

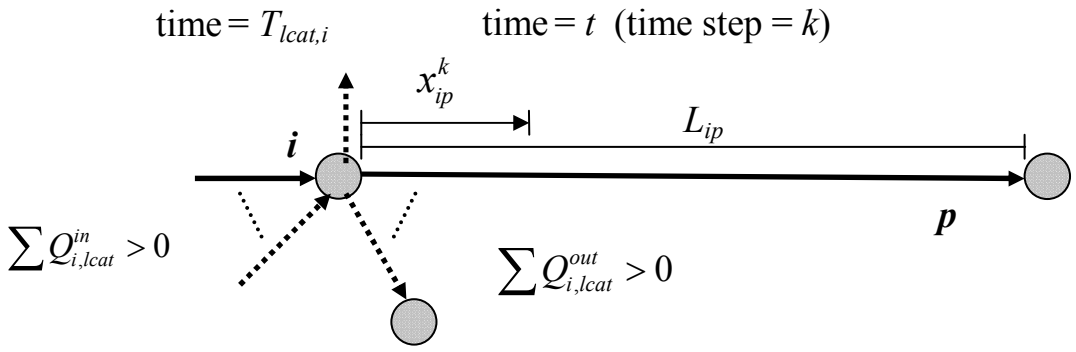
node) for a given attack scenario, but contaminant arriving at the latest contaminant arrival time and leaving the node through pipe outflow cannot travel from the node in question to any downstream node by the end of the study period. In essence, a slow-outflow node is a *de facto* end-of-line node for contaminant can travel to no other WDS nodes before the end of the study period upon reaching the node. Slow-outflow node status for a node i is represented by the variable SOF_i that has a value of 1 if node i is a slow-outflow node or 0 if otherwise. Figure 4.4 (b) illustrates slow-outflow node flow conditions for a junction.

Ascertaining whether or not a WDS node i is a slow-outflow node entails the following:

- determining the latest contaminant arrival time for the node ($T_{lcat,i}$) by simulating an injection at every WDS node at the latest eligible injection time (i.e., at 24 hours into the 96-hour study period for all three systems in this work) and recording the latest contaminant arrival time for the node,
- calculating the total distance water contaminated at the node's latest contaminant arrival time travels from the node i through a pipe p with positive outflow at the node's latest contaminant arrival time toward a downstream node for every time step k after the latest contaminant arrival time (x_{ip}^k in Figure 4.4 (b)) until the end of the study period for every pipe connected to the node with positive outflow at the latest arrival time, and
- comparing the total distance calculated for each pipe p with the length of each respective pipe (L_{ip}).



(a)



(b)

Figure 4.4 Flow behaviors at a WDS junction, node *i*, under (a) end-of-line conditions and (b) slow-outflow conditions. Solid arrows in the diagrams represent pipe or demand flows that are required to be present, and dashed arrows represent pipe or demand flows that are not required but can be present. For (a), *t* represents any time within the study period. As shown in (a), the junction should have a number of inflow pipes greater than zero, a total pipe inflow $\left(\sum Q_{it}^{in}\right)$ greater than zero, and a hydraulic demand Q_{it}^{demand} greater than zero, while the total pipe outflow $\left(\sum Q_{it}^{out}\right)$ should be equal to zero. With regard to (b), contaminant is assumed to reach node *i* at the node's latest contaminant arrival time, $T_{lcat,i}$. The junction should have at least one pipe with positive inflow and at least one pipe with positive outflow as indicated in (b). Thus, the total pipe inflow $\left(\sum Q_{i,lcat}^{in}\right)$ and total pipe outflow $\left(\sum Q_{i,lcat}^{out}\right)$ are greater than zero. For a given pipe *p* connected to node *i* with positive outflow, L_{ip} is the length of the pipe, and x_{ip}^{k} is the distance contaminated water travels in the time from $T_{lcat,i}$ to time *t* (i.e., time step *k*). For node *i* under slow-outflow conditions, x_{ip}^{k} at the end of the study period should be less than L_{ip}.

If for no outflow pipe p (i) the total distance traveled by the end of the study period (i.e., x_{ip}^{k} when k corresponds to the end of the study period) is greater than or equal to the length of the pipe or (ii) a negative value is found for x_{ip}^{k} for any time step k indicating that the contaminated water reversed flow and exited the pipe in question, the node in question is considered a slow-outflow node. Mathematically, the total contaminated water pipe distance at a time step k is determined with the equation

$$x_{ip}^{k} = x_{ip}^{k-1} + u_{ip}^{k}\Delta t \tag{4.20}$$

where k is the time step index corresponding to a particular five-minute multiple of time that falls after the latest contaminated arrival time for the node i and before the end of the study period, Δt is the five-minute time step, and u_{ip}^{k} is the EPANET-computed water velocity for pipe p connected to node i for time step k. The distance x_{ip}^{k} for the time step k associated with the latest contaminated arrival time is equal to zero.

The maximum nodal uniqueness value of a slow-outflow junction is not necessarily equal to 1 as it is in the case of an end-of-line node. For instance, if an injection actually occurs at a node identified as a slow-outflow node, contaminant may be able to reach some node downstream of the slow-outflow node by the end of the study period, whereas an injection occurring at a node other than that node may cause that node to be the last node on a contaminant transport path reached by contaminated water before the end of the study period, in turn causing the node to exhibit slow-outflow conditions. Under these circumstances, the slow-outflow node would have a maximum uniqueness value according to (4.19) of at least 0.5 (1/2), even though it is possible that the node can act as

a *de facto* end-of-line node. In most cases, it is expected that the maximum uniqueness value for a slow-outflow node would be relatively high.

Identifying slow-outflow nodes can be a computationally burdensome process compared to that for identifying end-of-line nodes, especially for larger systems. Thus, the slow-outflow node concept should be employed only when its benefits vastly outweigh its computational costs—usually in smaller system cases. The implementation of the slow-outflow node concept for determining nodal importance in smaller systems is discussed in sections that follow.

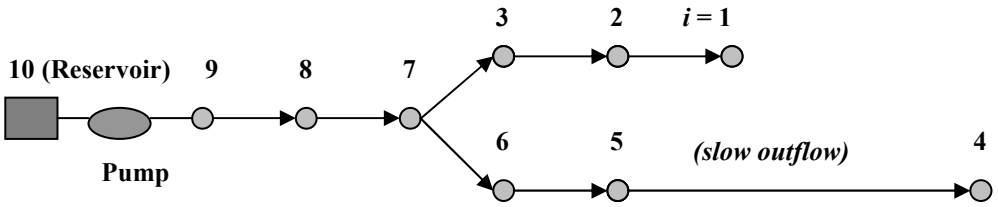
4.5.2.1.5 Application of Defined Concepts

The example problem in Figure 4.5 demonstrates how frequency, uniqueness, and end-of-line or slow-outflow node status can influence a node's importance with respect to maximizing detection likelihood. In the case of one sensor employed in solving the example problem, frequency is most important as a higher frequency directly translates to a higher detection likelihood. Thus, the one sensor is placed at Node 1 in the example WDS because it experiences contamination during more scenarios than do the other nodes. However, when more than one sensor is employed, minimizing detection redundancy becomes a crucial factor in optimal placement. In the case of two sensors, the second sensor is placed at Node 5 for a sensor there detects contaminant for an additional 20% of scenarios beyond the 70% of scenarios for which contaminant is detected by the sensor at Node 1; a second sensor at Node 4 or Node 6 would only detect

contaminant for an additional 10% of scenarios, and the sensor at any other node would detect contaminant for the scenarios already accounted for by the first sensor at Node 1. In the case of three sensors, the remaining 10% of non-detection scenarios are accounted for with a sensor at Node 4 to fully minimize detection redundancy and yield a detection likelihood of 100%.

Expected and maximum nodal uniqueness values found in the importance phase using (4.18) and (4.19), respectively, for Nodes 4 and 5 as well as for Node 1 give indications of sensor placement desirability with respect to minimizing detection redundancy in the optimization phase. Nodes 1 and 5 have expected uniqueness values greater than the values for the other eight nodes, and Node 4 has a value that is relatively high, albeit the fifth-largest of the ten values. These values demonstrate that expected uniqueness values can provide good estimates of nodal importance for a small system such as the one in Figure 4.5 but are not necessarily exact measurements of importance. Regarding maximum uniqueness, Nodes 1, 4, and 5 are the only nodes with a maximum nodal uniqueness value of 1; all other nodes have lower values. (Under certain flow and transport conditions, Node 5 may have a lesser maximum uniqueness value, though it would probably be relatively high compared to those of most other nodes. For the example problem, a realistic assumption is made that the maximum uniqueness value of Node 5 is 1.) Though maximum uniqueness does not indicate the on-average uniqueness of a node over many scenarios, it does give an upper bound of possible uniqueness for a given attack scenario.

It is no surprise that Nodes 1, 4, and 5 comprise the subset of end-of-line and slow-outflow nodes for the example WDS. This subset effectively contains the endpoints of all possible hydraulic paths on which contaminant can travel, as long as the assumption that the contaminant concentration is at or above the sensor resolution at nodes experiencing contamination holds. The same optimization results presented in Figure 4.5 would be found by only searching the three-node subset of end-of-line and slow-outflow nodes instead of searching the set of all 10 nodes, increasing optimization efficiency.



i	$FREQ_i^{lik}$	$UNIQ_i^{lik,avg}$	$UNIQ_i^{lik,max}$	Node Type
1	70%	0.238	1/1 = 1.00	end-of-line
2	60%	0.138	1/2 = 0.50	
3	50%	0.088	1/3 = 0.33	
4	10%	0.100	1/1 = 1.00	end-of-line
5	60%	0.205	1/1 = 1.00	slow-outflow
6	50%	0.105	1/2 = 0.50	
7	40%	0.055	1/6 = 0.17	
8	30%	0.038	1/7 = 0.14	
9	20%	0.024	1/8 = 0.13	
10	10%	0.011	1/9 = 0.11	

M	Sensor i's	Z_{lik}
1	1	70%
2	1, 5	90%
3	1, 5, 4	100%

Figure 4.5 Hypothetical example problem for optimizing sensor placement to maximize detection likelihood. The system depicted is assumed to be under steady-state conditions. At nodes experiencing contamination, the contaminant concentration is assumed to be at or above the sensor resolution. The numbers of scenarios for the importance phase (S_{imp}) and the optimization phase (S_{opt}) are both equal to 10 (i.e., one contaminant injection for every node). The problem is solved for one sensor, two sensors, and three sensors employed (i.e., $M = 1$, $M = 2$, and $M = 3$). The left table provides nodal frequency, expected uniqueness, maximum uniqueness values calculated using (4.16), (4.18), and (4.19), respectively, and indicators of end-of-line or slow-outflow node status for the nodes as appropriate. It is assumed that the procedure described in Section 4.5.2.1.4 was used to identify Node 5 as a slow-outflow node. The right table gives optimal sensor nodes and resulting detection likelihood values.

The relationships between the concepts defined above and the optimization of sensor placement with respect to maximizing detection likelihood are summarized by Figure 4.6. Using these concepts, functions for calculating the nodal importance value for a node *i* with respect to maximizing detection likelihood (IMP_i^{lik}) were formulated. The first function presented below is designed for "large" systems (i.e., systems with a number of nodes at least in the high hundreds). This function is considered the default function for maximizing detection likelihood and is used unless another function is more advantageous for a special case. Special cases include cases of maximizing detection likelihood for "small" systems (i.e., systems with a number of nodes in the low hundreds or less). The small system functions for determining nodal importance developed in this work are also provided below.

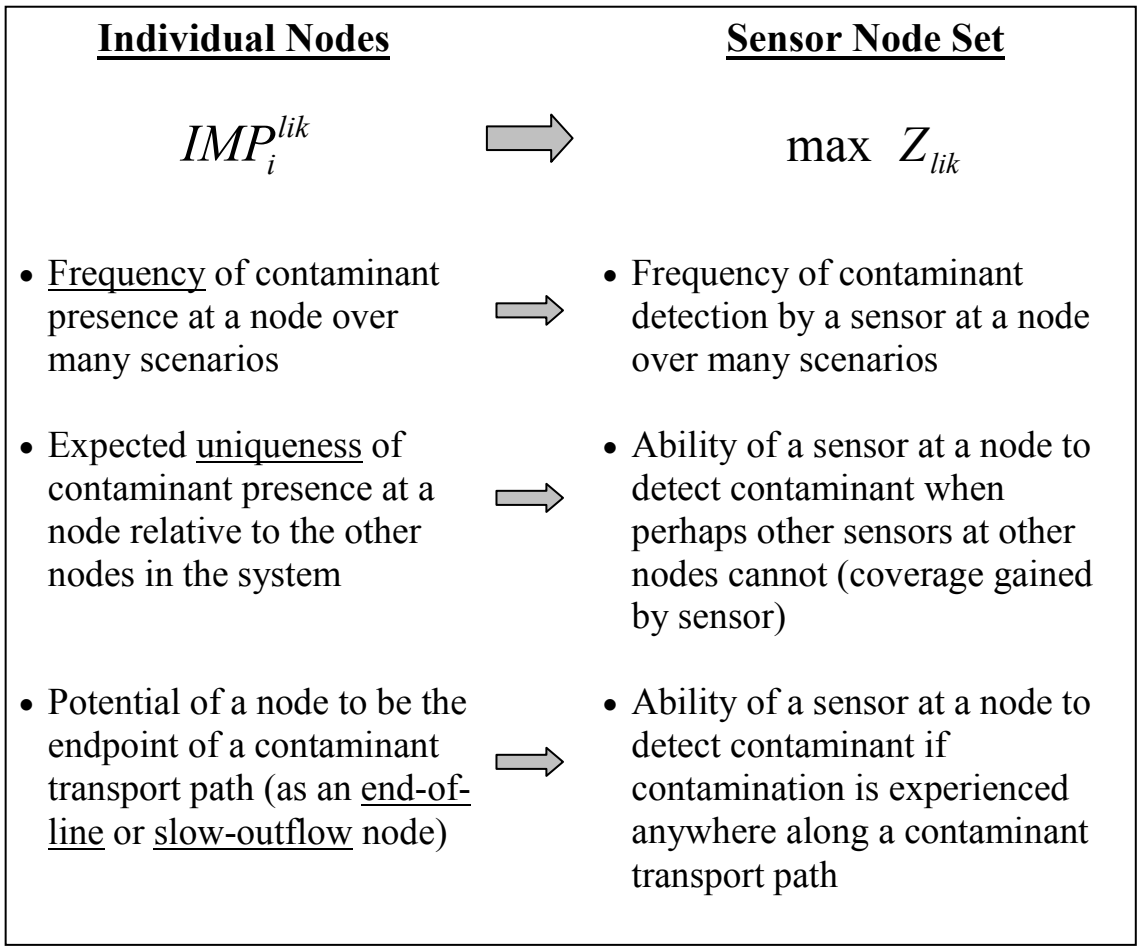


Figure 4.6 Relationships between nodal importance concepts and the optimization of sensor placement to maximize detection likelihood.

4.5.2.1.6 Function for Large Systems

It was established above that uniqueness as formulated in (4.17) is inherently a function of frequency. For large systems, though, the correlation between uniqueness and frequency is weak on average relative to the correlation for small systems due to fewer opportunities for nodes to experience contamination over S_{imp} scenarios. Therefore, in determining the importance of a certain node in a large system with respect to maximizing detection likelihood, frequency and uniqueness can be considered nearly independent of each other and should be accounted for explicitly in the nodal importance function as indicated in (4.21).

$$IMP_i^{lik} = UNIQ_i^{lik,avg} \cdot FREQ_i^{lik} \tag{4.21}$$

The nodal importance function in (4.21) makes nodes with relatively high values for both expected uniqueness and frequency stand out from other WDS nodes. Sensors located at more important nodes according to values calculated with (4.21) are expected to detect contamination for greater numbers of scenarios with minimal detection redundancy.

Large systems usually have very large subsets of end-of-line and slow-outflow nodes. For example, BWSN Network 2 has 2,140 end-of-line nodes (more than 17% of the nodes in the system), and the number of slow-outflow nodes in the system makes the size of the end-of-line and slow-outflow node subset greater than 2,140. Identifying slow-outflow nodes for a large system can be extremely computationally expensive as explained previously. Thus, for large systems, the subset of only end-of-line nodes is determined, and slow-outflow nodes are neglected. With a number of end-of-line nodes

that should far exceed the number of sensors available, the subset of only end-of-line nodes should provide an ample number of candidates for placing sensors at the ends of contaminant transport paths in a typical large system.

If contaminant is assumed to be at or above the sensor resolution where contamination exists, end-of-line nodes allow for the sorting of nodes in order to more thoroughly quantify nodal importance. Nodes can be sorted into two subsets according to whether or not they are end-of-line nodes with all end-of-line nodes considered more important than those that are not end-of-line nodes. Importance values are assigned to nodes according to (4.21), but a node's importance value is only relevant in comparing the node to other nodes in a particular subset within which it belongs. In other words, nodal importance is determined first by end-of-line status and then by calculated importance value. If two nodes with different importance values in the end-of-line node subset are compared, clearly the node with the higher importance value is considered more important. However, if one node in the end-of-line node subset and another node in the other subset are compared, the node in the end-of-line node subset is considered more important no matter what the importance values of the two nodes are. In cases for which the above contaminant concentration assumption is not valid, nodes are not sorted, and nodal importance is determined only by calculating importance values.

4.5.2.1.7 Function for Small Systems

The formula in (4.21) can be adjusted to assign importance with respect to maximizing detection likelihood to nodes in small systems. For small systems, uniqueness is more strongly correlated with frequency than it is for large systems due to the greater numbers of opportunities nodes have to experience contamination. It is somewhat redundant to include a frequency term in the nodal importance function when expected uniqueness is a variable in the function. Thus, nodal importance is considered synonymous with expected uniqueness, as shown in (4.22).

$$IMP_i^{lik} = UNIQ_i^{lik,avg} \tag{4.22}$$

The identification of slow-outflow nodes in a small system should typically be a computationally manageable process, and therefore these nodes are identified. Assuming that the contaminant concentration is at or above the sensor resolution, the subset of end-of-line and slow-outflow nodes can aid the determination of nodal importance for a small system by allowing for the sorting of nodes to further distinguish nodes as important in the same way the subset of only end-of-line nodes does for large systems. It is possible in the small-system case that the number of sensors available may be larger than the number of nodes in the end-of-line and slow-outflow node subset. If so, all nodes in that subset would be assigned sensors in the optimization phase, and the remaining sensors would be placed at "unimportant" nodes outside of the subset and would have no impact on detection likelihood values.

4.5.2.2 Minimizing Expected Detection Time

As discussed in Chapter 3, the calculation of expected detection time takes into account scenarios for which contaminant is not detected by assigning a detection time equal to the study period length to those scenarios in order to quantify "infinite" detection time. A by-product of this decision is the relatively strong positive correlation between expected detection time and detection likelihood. To reflect this correlation, the nodal importance function with respect to minimizing expected detection time is a function of the nodal importance value for maximizing detection likelihood as shown in the equation

$$IMP_i^{time} = \frac{1}{S_{imp}} \sum_{s=1}^{S_{imp}} \tau_{is} \cdot IMP_i^{lik} \tag{4.23}$$

where τ_{is} is a time-based variable for a node i and scenario s defined below. The value for IMP_i^{lik} can be found with (4.21) if the WDS in question is large or (4.22) if the WDS in question is small.

The function in (4.23) essentially scales IMP_i^{lik} by a time-based importance variable averaged over S_{imp} scenarios. That time-based variable gives more importance weight to a node with an earlier mean contaminant arrival time for an earlier mean arrival time implies that a sensor placed at the node in the optimization phase would detect contamination sooner on average than it would at a node with a later mean arrival time. Coupling the variable with IMP_i^{lik} as done in (4.23) has the effect of further distinguishing nodes that are relatively more important in terms of maximizing detection

likelihood according to their contaminant arrival times. The specific formula for the time-based variable τ_{is} is

$$\tau_{is} = \begin{cases} 1 - \dfrac{t_{is}^{d}}{T_{sp}} & D_{is} = 1 \\ 0 & D_{is} = 0 \end{cases} \tag{4.24}$$

where t_{is}^{d} is the time after the first (or only) contaminant injection when contaminant is first present at node i at a concentration above the sensor resolution during scenario s, and T_{sp} is the study period length. To express (4.24) in simple terms, the contaminant arrival time during scenario s for node i is normalized to fall on the [0, 1] scale. For example, if contaminant were to arrive at node i instantaneously after contaminant injection, τ_{is} would equal 1, or, if the time of one study period were to elapse before contaminant arrived, τ_{is} would equal 0.

The relationships between variables affecting nodal importance and the optimization of sensor placement with respect to minimizing expected detection time are summarized by Figure 4.7.

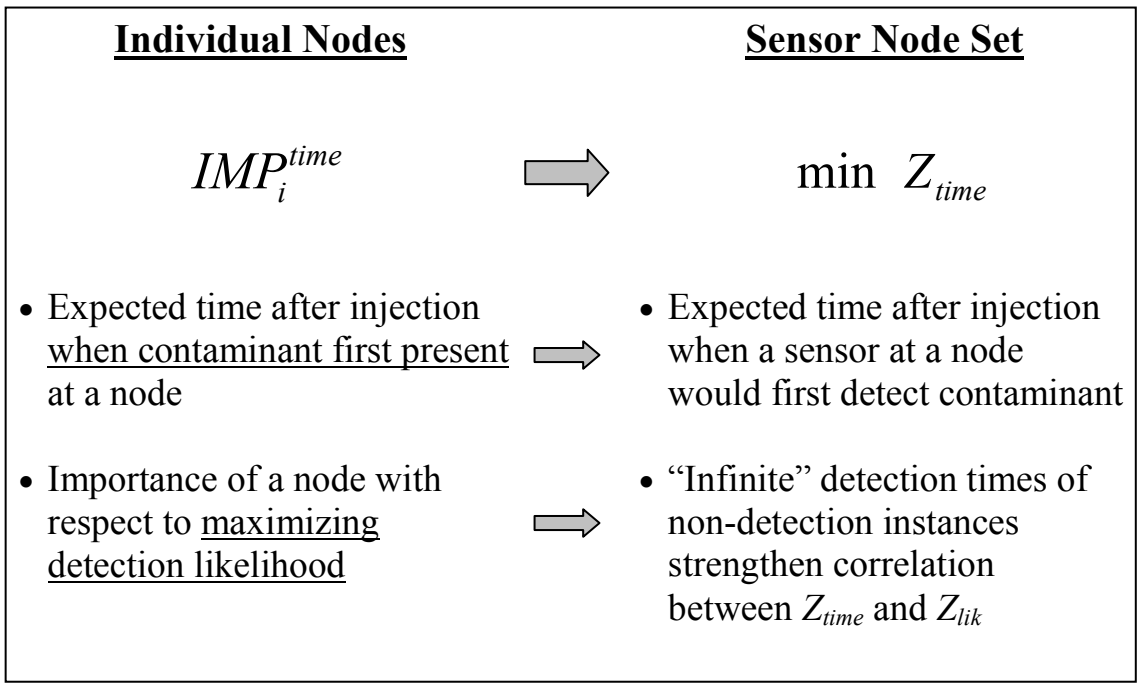


Figure 4.7 Relationships between variables affecting nodal importance and the optimization of sensor placement to minimize expected detection time.

4.5.2.3 Minimizing Expected Contaminated Demand Volume

Regarding the minimizing of expected contaminated demand volume, the more important nodes for sensor placement consideration are those that would allow a sensor to

- maximize the volume of water prone to contamination monitored, and
- detect contaminant more quickly after injection so that contaminant propagation would be limited.

Thus, the nodal importance function to address minimizing expected contaminated demand volume is a function of two variables corresponding to those attributes:

- the sum of the volume at or above the hazard threshold concentration discharged from a node as either hydraulic demand or pipe outflow, and

- the time after injection when the contaminated volume is discharged from the node.

Sensors placed in the optimization phase at nodes passing larger contaminated volumes at earlier times in the importance phase are expected to have greater probabilities of minimizing contaminated demand volume. The relationships between variables affecting nodal importance and the optimization of sensor placement to minimize expected contaminated volume are summarized in Figure 4.8. As in the case of maximizing detection likelihood, it was found that different nodal importance functions incorporating those importance variables are needed to appropriate quantify importance for different system sizes.

The total contaminated volume discharge V_{isk}^{cont} for node *i*, scenario *s*, and time step *k* in the importance phase is employed for both large and small systems and is found with the equation

$$V_{isk}^{cont} = \left(\sum_{p} Q_{ikp} + Q_{ik}^{demand} \right) \cdot \delta_{isk} \cdot \Delta t \tag{4.25}$$

where *p* is the index of a pipe connected to node *i* with positive outflow at time step *k*, Q_{ikp} is the hydraulic flow leaving node *i* at time step *k* through outflow pipe *p*, Q_{ik}^{demand} is the hydraulic demand at node *i* at time step *k*, δ_{isk} is a binary variable equal to 1 or 0 if contaminant is present or not present, respectively, at or above the hazard threshold concentration at node *i* at time step *k*, and Δt is the length of the time step (a constant 5 minutes in this work).

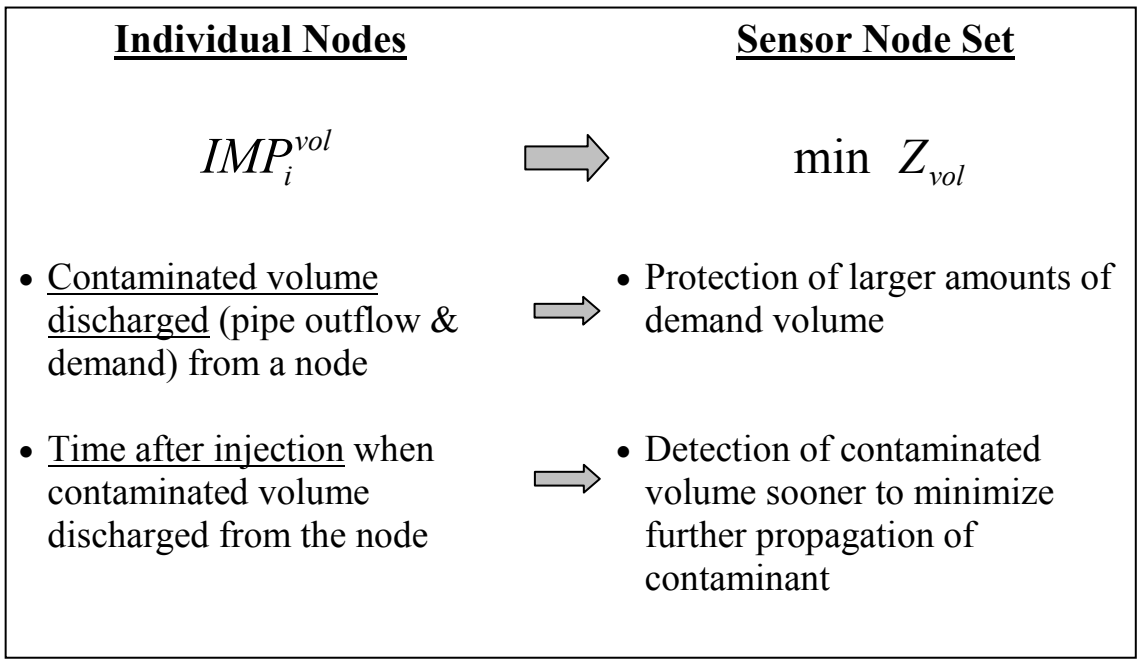


Figure 4.8 Relationships between variables affecting nodal importance and the optimization of sensor placement to minimize expected contaminated demand volume.

4.5.2.3.1 Function for Large Systems

For large systems, the volume calculated with (4.25) is coupled with the time-based term θ_{isk} in the relevant nodal importance function. Consequently, this coupling gives more importance weight to a contaminated volume experienced at the node at an earlier time. The time-based term is normalized on [0, 1]; it equals 1 at the time of contaminant injection and decreases linearly as time increases toward the end of the study period. The formula for the term is

$$\theta_{isk} = 1 - \frac{k - k_s^{inj}}{K - k_s^{inj}} \tag{4.26}$$

where k_s^{inj} represents the time step at which contaminant injection occurs during scenario *s*, and *K* represents the time step at the end of the study period.

The equation in (4.27) is the nodal importance function for minimizing contaminated demand volume employing the two variables defined above.

$$IMP_i^{vol} = \frac{1}{S_{imp}} \sum_{s=1}^{S_{imp}} \left(\frac{1}{K - k_s^{inj} + 1} \sum_{k=k_s^{inj}}^{K} \theta_{isk} V_{isk}^{cont} \right) \tag{4.27}$$

In (4.27), the θ_{isk}-V_{isk}^{cont} couple is averaged over all time steps associated with a particular scenario and then averaged over S_{imp} scenarios. The number of time steps varies among scenarios due to the random selection of injection time for each scenario; averaging the variable couple over the number of time steps for a given scenario normalizes the value of the couple so that it is compatible for comparison with couple values corresponding to other scenarios. The denominator "$K - k_s^{inj} + 1$" in (4.27) may seem odd upon first inspection for calculating the number of time steps from injection to study period end, but "$K - k_s^{inj}$" is the number of time steps in the study period beyond k_s^{inj}, so "+1" accounts for k_s^{inj} itself.

4.5.2.3.2 Function for Small Systems

For small systems, reducing contaminated volume is more dependent on the time of detection of contamination as detection occurs much more rapidly on average for small systems that it does for large systems. Through trial-and-error, it was found that the best nodal importance function for small systems accounting for that point is

$$IMP_i^{vol} = \frac{1}{S_{imp}} \sum_{s=1}^{S_{imp}} \left(\frac{1}{K - k_s^{inj} + 1} \sum_{k=k_s^{inj}}^{K} V_{isk}^{cont} \right) \cdot IMP_i^{time} . \quad (4.28)$$

This function is very similar to that in (4.27); the key difference is that the dependency on time of importance is represented by the nodal importance value for minimizing expected detection time which acts as an independent variable that is multiplied with an averaged volume. This formulation is in contrast to the incorporation of time dependency in the volume-averaging process done by (4.27).

While (4.27) gives a good estimate of nodal importance in most cases, this importance function is not at all a perfect measurement. For instance, a large influx of flow may occur downstream of a given node that experiences contamination; the equation in (4.27) may indicate that a downstream node is more important than the node in question because of a much greater contaminated volume, but in actuality the node in question may be more important as a sensor placed at the node may detect contaminant sooner and prevent the contamination of even greater volumes. Such instances, though, should not jeopardize the overall validity of (4.27) as an importance function.

4.5.2.4 Summary of Importance-Optimization Relationships

Table 4.1 summarizes how the variables and concepts influencing nodal importance defined above relate to the three protection goals of interest in the work of this thesis.

Table 4.1 Summary of relationships between importance variables and protection goals. Variable notations are those used in (4.16) through (4.28).

Protection Goal	Nodal Importance Function	Variable/Concept Affecting Nodal Importance	Corresponding Importance Variable(s)	Value(s) Associated with Higher Nodal Importance
Maximizing Detection Likelihood (Z_{lik})	IMP_i^{lik}	Frequency of contaminant presence at a node over many scenarios	$FREQ_i^{lik}$	higher values
		Expected uniqueness of contaminant presence at a node relative to the other nodes in the system	$UNIQ_i^{lik,avg}$	higher values
		Potential of a node to be the endpoint of a contaminant transport path (as an end-of-line or slow-outflow node)	EOL_i SOF_i	"1" for either variable
Minimizing Expected Detection Time (Z_{time})	IMP_i^{time}	Expected time after injection when contaminant first present at a node	τ_{is}	earlier times
		Importance of a node with respect to maximizing detection likelihood	IMP_i^{lik}	higher values
Minimizing Expected Contaminated Demand Volume (Z_{vol})	IMP_i^{vol}	Contaminated volume discharged (pipe outflow & demand) from a node	V_{isk}^{cont}	higher values
		Time after injection when contaminated volume discharged from the node	θ_{isk} IMP_i^{time}	earlier times

4.5.2.5 Multiple Objectives

In order for the nodal importance concept to be effective in multiobjective optimization, the concept must be tailored to accommodate the particular multiobjective optimization approach of interest, which in this work pertains to finding a Pareto-optimal solution that maximizes protection with respect to all protection goals collectively as explained above. Thus, multiobjective nodal importance in this work relates to the potential a node has in providing protection with respect to all three protection goals collectively as a sensor node. Figure 4.9 summarizes the relationship of multiobjective nodal importance with optimal sensor placement in a multiobjective problem.

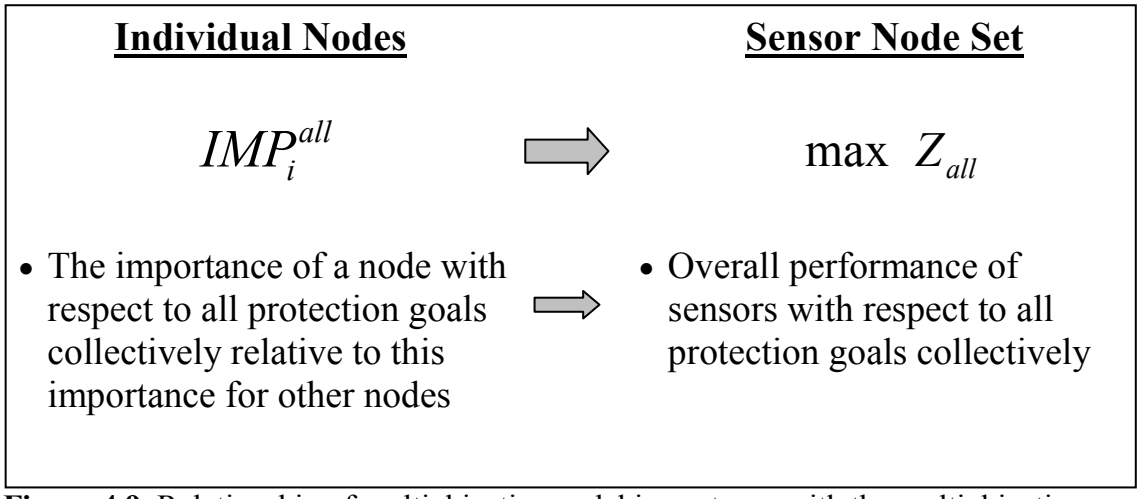


Figure 4.9 Relationship of multiobjective nodal importance with the multiobjective optimization of sensor placement.

Determining nodal importance values with respect to all three protection goals individually provides the information needed to determine multiobjective nodal importance values. It stands to reason that a node that has the potential to provide more

protection according to all three protection goals would be considered more important with respect to all individual goals. To quantify concisely with one value the on-average degree a node is more important with respect to all protection goals than other WDS nodes, multiobjective nodal importance employs the Pareto dominance concept. As stated above, Pareto dominance of a sensor placement solution over some other solution implies that

- the solution in question performs at least as well as the other solution with respect to all protection goals, and
- the solution performs better than the other solution does with respect to at least one protection goal.

Applying this idea to nodal importance instead of sensor node performance, one node is dominant over another node in terms of importance if

- the importance values for all protection goals for the node in question are greater than or equal to the counterpart importance values for all protection goals of the other node, and
- at least one of the importance values for a particular protection goal for the node in question is strictly greater than its counterpart for the other node.

Conversely, the node in question is dominated by the other node if

- the importance values for all protection goals for the node in question are less than or equal to the counterpart importance values for all protection goals of the other node, and
- at least one of the importance values for a particular protection goal for the node in question is strictly less than its counterpart for the other node.

Using this concept, multiobjective importance for a particular node in this work is a function of how many other nodes that the node in question dominates and how many other nodes dominate the node in question.

The actual multiobjective nodal importance function for this work is an extension of prior work done to model groundwater flow and transport. Wang (2008) developed a "criticalness index" that measured how well a set of uncertain parameters (referred to as a "realization" by Wang) used in modeling groundwater flow and transport provides for the simulation of true extreme contaminant arrival times relative to other realizations. Criticalness index values allowed Wang to conduct simulations with a limited number of "more critical" realizations that acceptably modeled extreme arrival times, reducing the computational expense necessary for adequate modeling. This criticalness index for a particular realization is a function of the number of other realizations the realization in question dominates in terms of providing for the simulation of true extreme arrival times and the number of other realizations that dominate the realization in question. The specific equation to determine criticalness index used by Wang is

$$CI_i = \frac{J_i}{n-1} + \frac{n-1-J_i-K_i}{(n-1)(n-J_i)} \tag{4.29}$$

where i in that work is a particular realization, CI_i is the criticalness index for realization i, n is the total number of realizations considered, J_i is the number of realizations that realization i dominates, and K_i is the number of realizations that dominate realization i.

Even though the types of problem of this work and the work of Wang (2008) are different, the criticalness index concept and nodal importance concept both serve to

improve computational efficiency in the search for better-performing candidates by providing a concise numerical index that allows for the narrowing of the search. It is reasonable, then, that the criticalness index approach can be applied to determining multiobjective nodal importance.

In this work, the WDS node takes the place of the realization in the criticalness index function in (4.29). With this change, the criticalness index function becomes the multiobjective nodal importance function

$$IMP_i^{all} = \frac{\alpha_i}{N-1} + \frac{N-1-\alpha_i-\beta_i}{(N-1)(N-\alpha_i)} \tag{4.30}$$

where i is a particular WDS node, IMP_i^{lik} is the multiobjective nodal importance value for node i, N is the total number of WDS nodes, α_i is the number of nodes that node i dominates in terms f nodal importance, and β_i is the number of nodes that dominate node i.

Following the conditions of Pareto dominance, a node (e.g., Node "1") dominates another node (e.g., Node "2") in terms of importance if the following logical relationships are satisfied:

- $IMP_1^{lik} \geq IMP_2^{lik} \wedge IMP_1^{time} \geq IMP_2^{time} \wedge IMP_1^{vol} \geq IMP_2^{vol}$, and (4.31)
- $IMP_1^{lik} > IMP_2^{lik} \vee IMP_1^{time} > IMP_2^{time} \vee IMP_1^{vol} > IMP_2^{vol}$. (4.32)

If these conditions are indeed met, the value of α_1 increases by 1. This analysis is continued for the remainder of the WDS nodes and eventually yields the α_1 value used in (4.30). In contrast, Node 1 is dominated by Node 2 if the following relationships are satisfied:

- $IMP_1^{lik} \leq IMP_2^{lik} \wedge IMP_1^{time} \leq IMP_2^{time} \wedge IMP_1^{vol} \leq IMP_2^{vol}$, and (4.33)
- $IMP_1^{lik} < IMP_2^{lik} \vee IMP_1^{time} < IMP_2^{time} \vee IMP_1^{vol} < IMP_2^{vol}$. (4.34)

The value of β_1 increases by 1 if these conditions are met.

It is possible that other functions may serve as well as or even better than (4.30) serves to assign importance according to the multiobjective optimization approach of this work, but (4.30) is based on established work and appears to be effective in assigning importance, as will be demonstrated in Chapters 5, 6, and 7. Therefore, the function is considered appropriate for this work.

4.6 OPTIMIZATION METHODS

The methods used to actually carry out the optimization of sensor placement in this study are detailed here.

4.6.1 THE ITERATIVE SUBSET SEARCH METHOD

Nodal importance values are reasonable estimates of true nodal importance as indicated above, but they are imperfect measurements. In other words, there is no lockstep correspondence between the importance value of a node and the protection performance of a sensor placed at the node. Modeling the error associated with calculated nodal importance with precision is virtually impossible. Given these conditions, a method used to search for the optimal sensor node set that takes advantage of nodal importance

estimates in order to narrow the search and subsequently promotes computational efficiency while keeping the search broad enough to account for nodal importance calculation error as best as possible is warranted.

The "Iterative Subset Search Method" (ISSM) developed in this work brings such a balance to the optimization. In short, ISSM begins the search with a small subset primarily consisting of "more important" nodes according to calculated nodal importance values and incrementally increases the size of the subset to broaden the search until convergence to a solution of best protection performance occurs. This approach makes it possible for the optimal (or at least very near-optimal) sensor node set to be found while minimizing the number of suboptimal candidates that are evaluated during the search.

In addition to the higher-importance nodes, a particular ISSM node subset also contains a relatively small number of nodes randomly chosen from the WDS nodes that are not among the higher-importance nodes. The inclusion of these randomly chosen nodes is a measure used to verify as best as possible that ISSM does indeed find the optimal solution. If all sensor nodes in the best solution found are among the higher-importance nodes, there is evidence that the best solution is the optimal solution and increasing the subset size would not yield a better solution. However, if any sensor node in the best solution is among the nodes randomly chosen, then the subset size is increased in order to potentially find a better solution.

ISSM employs several user-defined decision variables that affect the ability of ISSM to find the optimal sensor node set in an effective and computationally efficient manner.

- Proportions of higher-importance nodes (i.e., "ranked" nodes) and randomly chosen nodes (i.e., "unranked" nodes) in the subset. The majority of nodes in the subset should be ranked as these nodes have much higher probabilities of being better sensor node candidates on the whole than unranked nodes have. However, there should be a reasonable number of unranked nodes in the subset to provide a sensible verification metric. Therefore, the proper balance between the proportions of ranked and unranked nodes is crucial. For this work, the ranked and unranked node proportions vary with study system; the proportions used for specific study systems are reported in chapters that follow. If the proportions chosen lead to non-integer numbers of ranked and unranked nodes, the number of unranked nodes is rounded up to the nearest integer, and the number of ranked nodes is rounded down.
- Initial subset size. The initial subset should be large enough to include a best sensor node set solution of respectable performance and enough unranked nodes such that the best solution has realistic opportunity to include unranked nodes. The initial subset size varied with study system in this work as will be shown in Chapters 5, 6, and 7.
- Rate of subset size increase. It is important to increase the subset size in broadening the search at a rate that will provide for an ample chance of finding better sensor node solutions. A low rate of increase can lead to the erroneous conclusion of attaining solution convergence. However, too large of a rate of

increase may make the search for the optimal solution unnecessarily computationally inefficient. The rates selected for particular study systems are reported in chapters that follow.

- Iteration stopping criteria. In general, the subset size should incrementally increase until both (i) the best solution no longer improves with increasing subset size and (ii) all sensor nodes in the best solution are ranked nodes. Thus, for every increase in subset size, the best solution found for the current subset is compared to overall best solution found from searching prior subsets. If the subset best solution is better than the overall best solution, then the subset size is again increased for solution convergence has not occurred. If the subset best solution is not better than the overall best solution, then the overall best solution is evaluated to determine if any of the nodes in the solution are unranked nodes. If there are unranked nodes in the solution, then subset size is increased for it is demonstrated that the ranked nodes in the subset do not completely include the optimal solution. If there are no unranked nodes in the overall best solution, then there is evidence suggesting that convergence has occurred. However, additional iterations of subset size increase are needed to ensure that convergence has indeed occurred. For each of the additional iterations, there should be no improvement in the subset best solution. The number of additional iterations is somewhat arbitrarily chosen. For this work, three additional iterations are conducted to ensure convergence.

Figure 4.10 illustrates the entire ISSM procedure incorporating the decision variables discussed above.

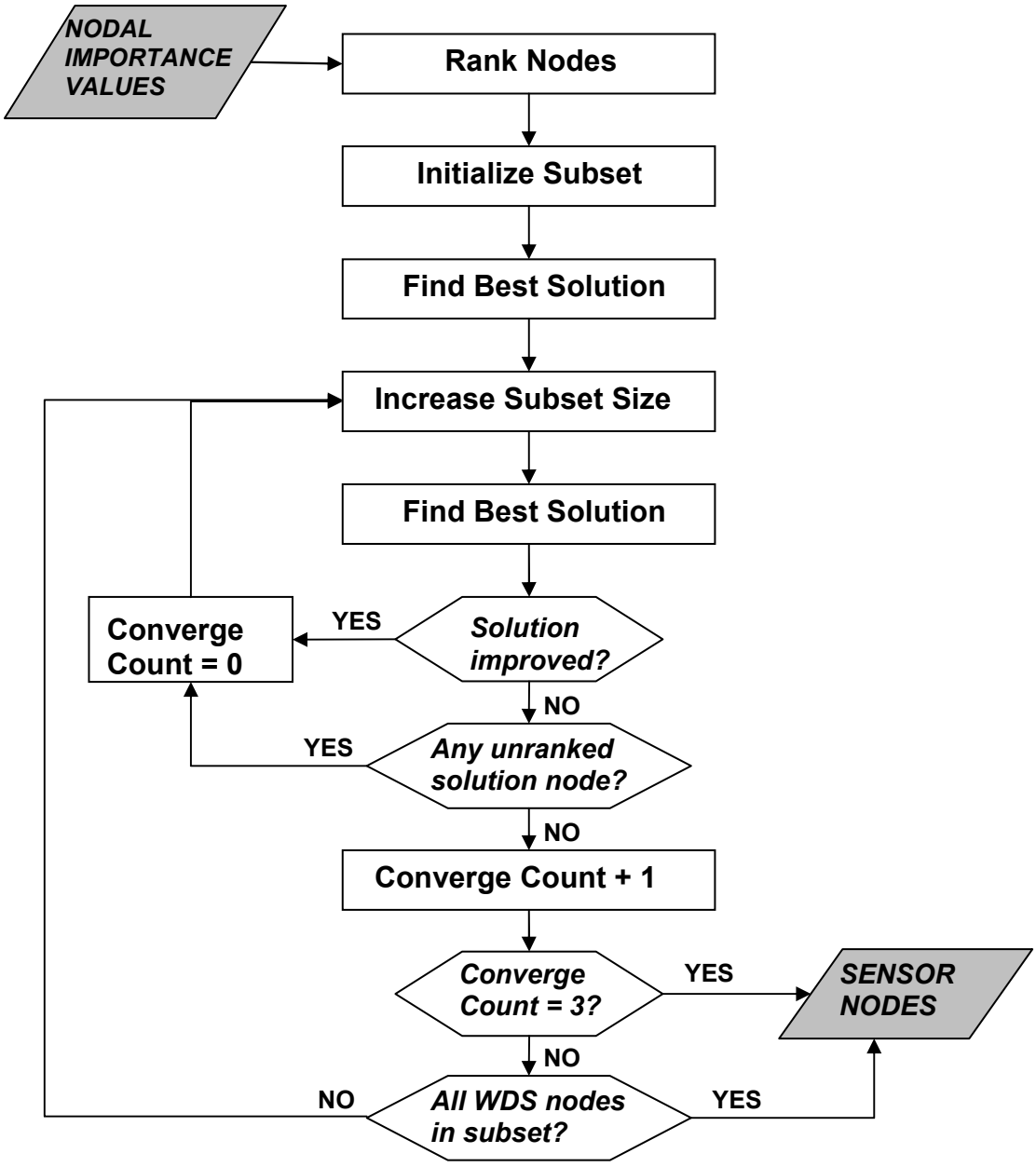


Figure 4.10 The Iterative Subset Search Method.

4.6.2 THE GENETIC ALGORITHM SOLVER

The search of an ISSM subset for the best sensor node solution is conducted by a genetic algorithm (GA) solver. The use of GAs is common and has been shown to be effective (e.g., Aral et al. 2008; Guan et al. 2006; Ostfeld and Salomons 2006; Rogers et al. 2007) in solving the optimal sensor placement problem. More specifically, the "simple" GA (Goldberg 1989) as a component of the GA library (Wall 2005) was used in this work. The flowchart in Figure 4.11 illustrates the evolutionary search procedure implemented by the simple GA in this work.

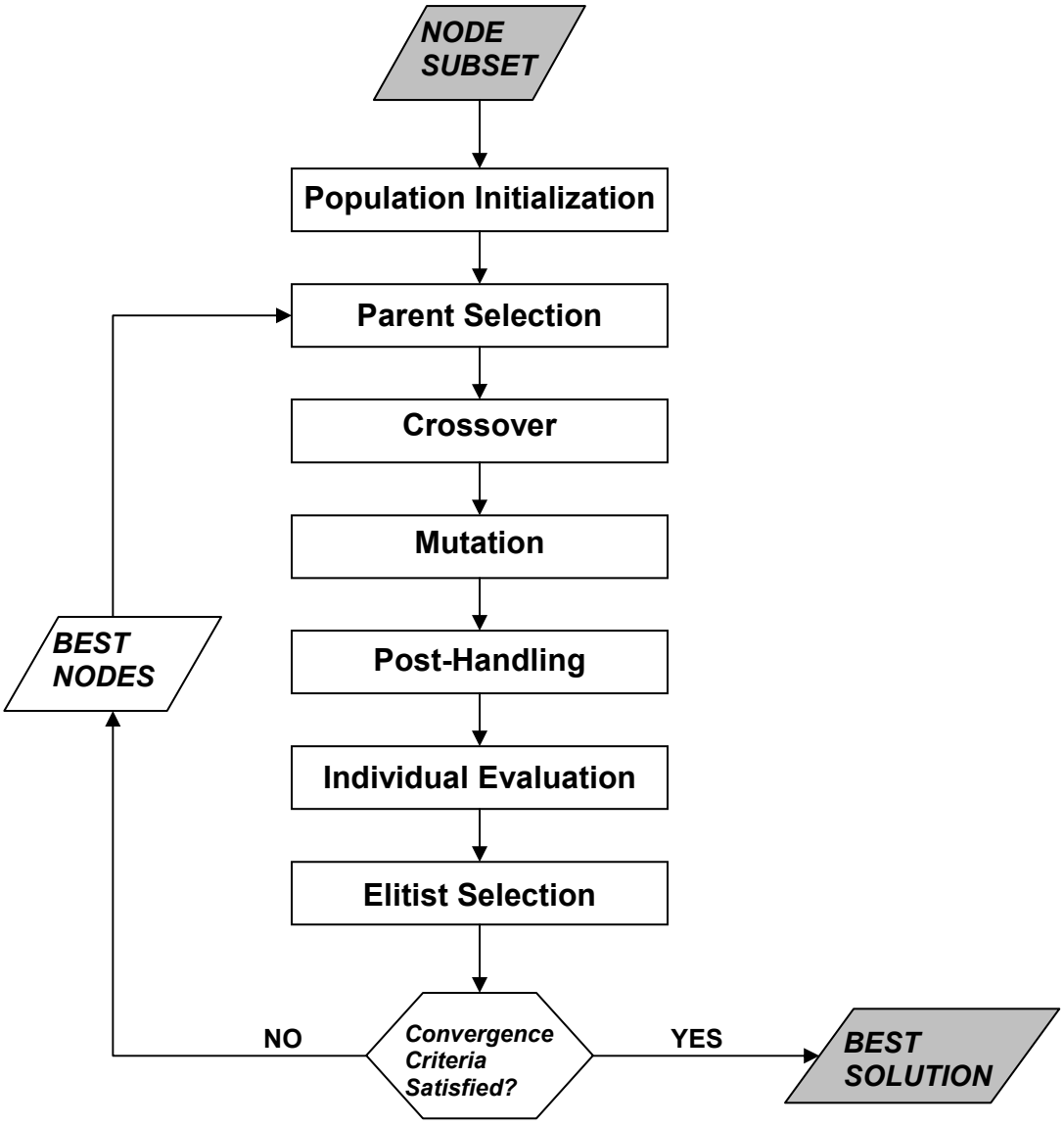


Figure 4.11 The genetic algorithm evolutionary procedure for one generation.

The GA solver for this work employs a binary string chromosome structure. The chromosome represents a particular subset of WDS nodes in question, and each bit in the chromosome represents a node either with a sensor (value of the bit equals "1") or without a sensor (value of the bit equals "0"). Thus, a chromosome with a particular set of bit values represents a sensor node solution candidate (i.e., an individual). The population of chromosomes for each generation of GA evolution is initialized in a uniform, random manner.

The actual evolutionary operations for the GA model are crossover, mutation, post-handling, population individual evaluation, and elitist selection.

Crossover is performed in the common two-point fashion under a crossover probability of 0.95. "Parents" for the crossover operation are determined through roulette-wheel selection of population individuals with selection probabilities derived using objective function values for the particular optimization problem of interest. Crossover results in the generation of new individuals (i.e., "children"). The mutation operation is performed on these children through "bit-flipping," changing the bit value from one binary value to the other, in a uniform, random manner, under a mutation probability of 0.05.

Post-handling entails uniform, random bit-flipping such that the end result for each individual satisfies the sensor availability constraint. After post-handling, the sum of the bit values for each evolved individual should be less than or equal to the number of sensors available. Performed in this manner, post-handling essentially forces mutation of

randomly chosen bits, resulting in a relatively large mutation of an individual in some cases. Such a large mutation increases the evolution time for the GA to converge to a best solution, but it also helps to prevent convergence to a local optimal solution, which is prone to occur as ISSM isolates many well-performing sensor node set candidates.

Evolved individuals that have undergone the crossover, mutation, and post-handling operations are evaluated with respect to their protection performance determined by the objective function of the optimization problem. The individual providing the most protection is preserved in elitist selection and is carried to the population of the next GA generation for continued evolution.

CHAPTER 5

RESULTS AND ANALYSIS: BWSN NETWORK 1

In this chapter, the sensor placement optimization results from employing the solution approach outlined in Chapter 4 for the study system BWSN Network 1 and for the various attack cases described in Chapter 3 are presented. The results for the other two study systems, BWSN Network 2 and the Toms River WDS, are given in Chapters 6 and 7, respectively.

5.1 OVERVIEW OF RESULTS AND ANALYSES PRESENTED

Specifically, sensor node solutions for each of the study systems for the single-objective problems represented by (4.2) through (4.4) and for the multiobjective problem represented by (4.7) are given in sections below. These solutions are analyzed to

- compare the protection provided by using the approaches of this work to protection provided by approaches documented in other works,
- ascertain any spatial trends in the optimal placement of sensors,
- gauge how sensitive solutions are to attack case differences,
- quantify the performance tradeoffs with respect to individual protection goals in employing multiobjective sensor node solutions, and
- evaluate the effectiveness of using nodal importance as a part of ISSM detailed in Chapter 4 in narrowing the search for the optimal sensor nodes.

5.2 SINGLE-OBJECTIVE PROBLEM: MAXIMIZING Z_{lik}

Results for sensor placement optimization with respect to maximizing detection likelihood are presented and discussed in this section.

5.2.1 SENSOR NODE SOLUTIONS

Sensor node solutions for maximizing detection likelihood in the default attack case for BWSN Network 1 are given in Table 5.1 for the cases of employing 5 and 20 sensors. Upon examining the results in the table independently, it seems that the sets of sensor nodes selected in optimization are associated with high performance in maximizing detection likelihood, especially considering that only 5 sensors in a 129-node system can detect contaminant for nearly 90% of attack scenarios. Figures 5.1 and 5.2 illustrate the locations of the optimal sensor nodes for $M = 5$ and $M = 20$, respectively. Figure 5.1 indicates well how many sensors for maximizing detection likelihood tend to be placed in the peripheral areas of the system where many contaminant pathways end. This tendency is not surprising given the preference for contaminant pathway endpoints as sensor locations to maximize detection likelihood. The clustering of sensor nodes on the western side of the WDS with 20 sensors employed seen in Figure 5.2 is a consequence of the maximum possible detection likelihood being reached with less than 20 sensors. As discussed in greater detail below, 15 nodes represent all of the contaminant pathway endpoints in the WDS. Once sensors are placed at these 15 nodes, the maximum possible detection likelihood is reached, and it does not matter where the remaining 5 sensors are

located. The cluster in Figure 5.2 represents most of these "extra" sensor nodes that were chosen due to the way nodes are ranked according to nodal importance. To verify maximum protection is reached with those 15 sensor nodes, a sensor was placed at every node in system for validation purposes only; the resulting Z_{lik} value was 94.7%--the same value associated with the 20-sensor solution in Table 5.1. The 5.3% of scenarios for which contamination was not detected corresponds to scenarios with injection nodes selected that are hydraulically inactive such that it would be impossible for contamination to be detected. This 20-sensor case is explored further in the evaluation of nodal importance below.

Table 5.2 compares the sensor nodes selected in this work for the default attack case to those selected in other works. The table shows that the protection in terms of maximizing detection likelihood provided by the sensor locations of this work is equal or better to that of sensor locations for other studies. The deficiencies of approaches used in other works to characterize the desirability of WDS nodes as sensor locations are also indicated by the table; the "nontime-constrained receivability" concept of Xu et al. (2008) led to clearly inferior results for both the $M = 5$ and $M = 20$ cases, and the use of "nodal impact coefficients" by Berry et al. (2006c) yielded slightly inferior results for the 20-sensor case.

Table 5.1 Optimal sensor locations and performance measures for maximizing detection likelihood in the default attack case for BWSN Network 1.

M	Sensor Nodes*	Z_{lik}** (%)	Z_{time} (min)	Z_{vol} (gal)
5	10, 45, 83, 100, 126	89.7	1,685.2	59,701
20	10, 11, 19, 34, 35, 39, 41, 42, 45, 79, 81, 82, 83, 84, 100, 114, 118, 123, 124, 126	94.7	806.3	4,734

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

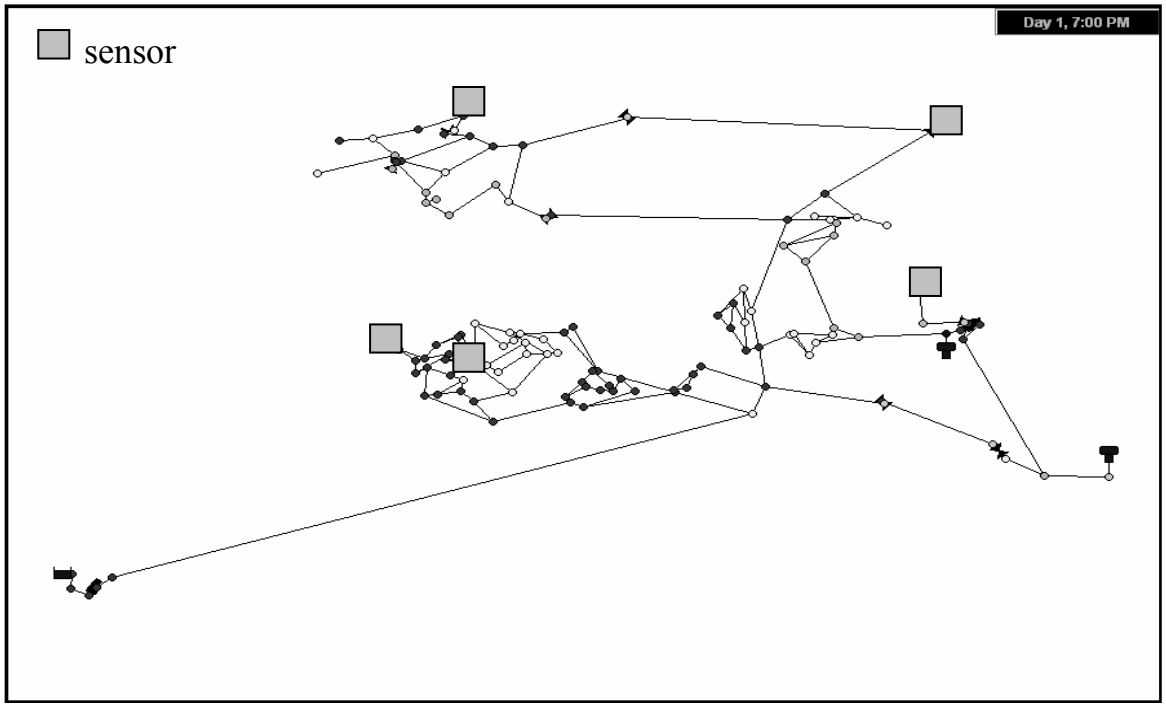


Figure 5.1 Optimal sensor locations with $M = 5$ for maximizing detection likelihood in the default attack case for BWSN Network 1.

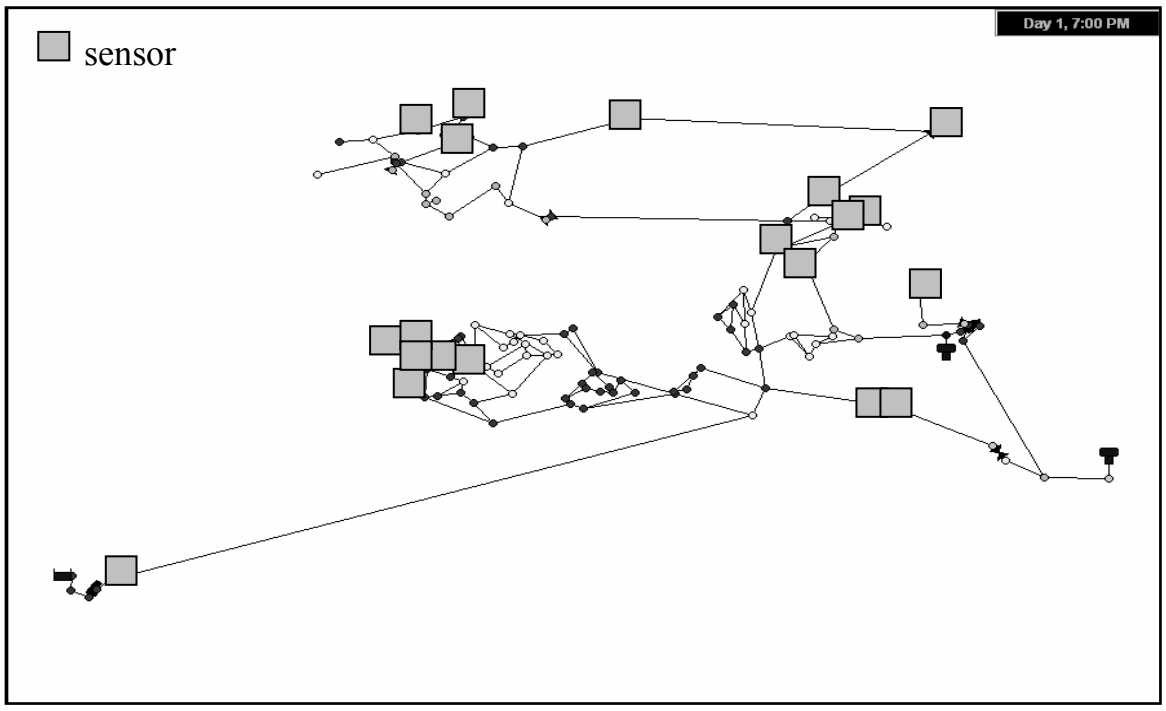


Figure 5.2 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the default attack case for BWSN Network 1.

Table 5.2 Comparison of optimal sensor locations for maximizing detection likelihood in the default attack case for BWSN Network 1 determined in this study with locations found in other works.

M	**Work**	**Sensor Nodes***	Z_{lik}** **(%)**
5	This Study	10, 45, 83, 100, 126	89.7
	Berry et al. (2006c)	10, 45, 83, 100, 126	89.7
	Eliades and Polycarpou (2006)	10, 45, 83, 100, 126	89.7
	Krause et al. (2006)	10, 45, 83, 100, 126	89.7
	Xu et al. (2008)***	45, 83, 100, 114, 126	86.0
20	This Study	10, 11, 19, 34, 35, 39, 41, 42, 45, 79, 81, 82, 83, 84, 100, 114, 118, 123, 124, 126	94.7
	Berry et al. (2006c)	1, 7, 10, 13, 16, 19, 35, 36, 38, 45, 83, 100, 106, 113, 114, 122, 123, 124, 125, 126	94.5
	Eliades and Polycarpou (2006)	10, 11, 14, 17, 19, 21, 31, 34, 35, 45, 68, 74, 83, 90, 100, 102, 114, 123, 124, 126	94.7
	Krause et al. (2006)	10, 11, 19, 34, 35, 45, 83, 100, 106, 110, 114, 123, 124, 126, 128, 129 †	94.7
	Xu et al. (2008)***	7, 13, 36, 38, 45, 83, 100, 113, 114, 123, 124, 125, 126, TANK-131 †	89.3

* All sensor node numbers preceded by "JUNCTION-", unless otherwise indicated.
** Values computed using the software developed in this work.
*** Solution for sensor placement according to "nontime-constrained receivability".
† Work provided less than *M* sensors, citing protection limit reached.

Table 5.3 gives the sensor nodes found in this study with 5 sensors employed for the default attack case already given in Table 5.1 as well as the variant attack cases outlined in Chapter 3. It is noted that for all study systems when maximizing detection likelihood there are no results for the attack cases of a contaminant hazard threshold = 0.0 mg/L and a response delay after detection of 3 hours because these variant conditions do not affect sensor placement according to the maximization of detection likelihood. Sensor nodes would be the same in those cases as in the default attack case, and Z_{lik} and Z_{time} values would be the same as in the default case, but Z_{vol} values would be greater than or equal to (very likely greater than) those found in the default attack case.

As shown in Table 5.3, the sensor nodes for all variant attack cases were the same as those for the default attack case. Even though these sensor nodes are the same for all attack cases, Z_{lik} values differ in expected manners. For increased sensor detection resolution, detection likelihood is slightly decreased due to an inability for a sensor to detect contamination for some scenarios because the contaminant concentration had dropped to less than 0.3 mg/L by the time the contaminant reached the sensor node. A slight increase in detection likelihood for the case of a longer injection duration is not unexpected because a larger window of time for injection during which WDS hydraulic behavior can change dramatically and allow contaminated water opportunities to travel to different and/or more places in system and subsequently sensors in these other places in the system to detect contamination. There is a substantial increase in detection likelihood for the two injection node attack cases due to contaminant access to additional

contaminant pathways. Even if one injection node is hydraulically inactive, the other injection node may be active, in turn, making contamination detectible.

It is seen in Table 5.3 that for all attack cases sensor nodes selected are the same as those for the default attack case. Isovitsch and VanBriesen (2008) observed that it is typical for most if not all locations of sensor nodes not to change greatly as attack conditions change, so the resulting sensor nodes in Table 5.3 are not surprising. This observation is also made for sensor nodes selected for different protection goals and study systems in sections and chapters that follow.

Table 5.4 shows that other works agree with this work in sensor placement for selected variant attack cases. The attack cases of a 10-hour contaminant injection duration and 2 nodes injected at 1 time were studied in BWSN, so data is available for those cases for comparison purposes. No comparisons can be made for the cases of a sensor detection resolution of 0.3 mg/L and 2 nodes injected at 2 times.

Table 5.3 Optimal sensor locations and performance measures for maximizing detection likelihood with $M = 5$ in the variant attack cases for BWSN Network 1.

Attack Case	Sensor Nodes*	Z_{lik}** (%)	Z_{time} (min)	Z_{vol} (gal)
default	10, 45, 83, 100, 126	89.7	1,685.2	59,701
sensor detection resolution = 0.3 mg/L	10, 45, 83, 100, 126	88.4	1,753.4	58,356
contaminant injection duration = 10 h	10, 45, 83, 100, 126	89.9	1,642.3	75,724
2 injection nodes at 1 time	10, 45, 83, 100, 126	98.9	919.5	80,483
2 injection nodes at 2 times	10, 45, 83, 100, 126	98.3	1,191.5	69,062

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

Table 5.4 Comparison of optimal sensor locations for maximizing detection likelihood with $M = 5$ in the variant attack cases for BWSN Network 1 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{lik}** (%)
contaminant injection duration = 10 h	This Study	10, 45, 83, 100, 126	89.9
	Berry et al. (2006c)	10, 45, 83, 100, 126	89.9
	Krause et al. (2006)	10, 45, 83, 100, 126	89.9
2 injection nodes at 1 time	This Study	10, 45, 83, 100, 126	98.9
	Krause et al. (2006)	10, 45, 83, 100, 126	98.9

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

5.2.2 EVALUATION OF NODAL IMPORTANCE

Since BWSN Network 1 is considered a small system, both (i) the small-system nodal importance function in (4.22) and (ii) the procedure of stratifying nodes that are either end-of-line or slow-outflow according to the definitions in Section 4.5.2.1 from other WDS nodes were used to rank nodes in terms of importance for maximizing detection likelihood in the default attack case. The end-of-line and slow-outflow nodes identified

for BWSN Network 1 are listed in Table 5.5 and depicted in Figure 5.3. With these nodes identified, the system nodes were ranked according to end-of-line/slow-outflow status first then by nodal importance value calculated with (4.22). The Iterative Subset Search Method was then carried out to find the optimal solution as illustrated in Figure 5.4 for the default attack case with 5 sensors employed. The ISSM parameters selected for this particular study system for all protection goals and attack cases are given in Table 5.6. It was felt that these parameters would provide ample opportunity for solution improvement if such potential exists with each iteration of ISSM and allow for the realistic opportunity for "unranked" nodes to be included in best solution of particular iteration such that the inclusion of unranked nodes in ISSM subsets would be a functional check for convergence.

Table 5.5 End-of-line nodes and slow-outflow nodes identified for BWSN Network 1.

Node Category	Identified Nodes*
End-of-Line	19, 34, 35, 39, 41, 42, 45, 83, 100, 114, 123, 124, 126
Slow-Outflow	10, 11

* All node numbers preceded by "JUNCTION-".

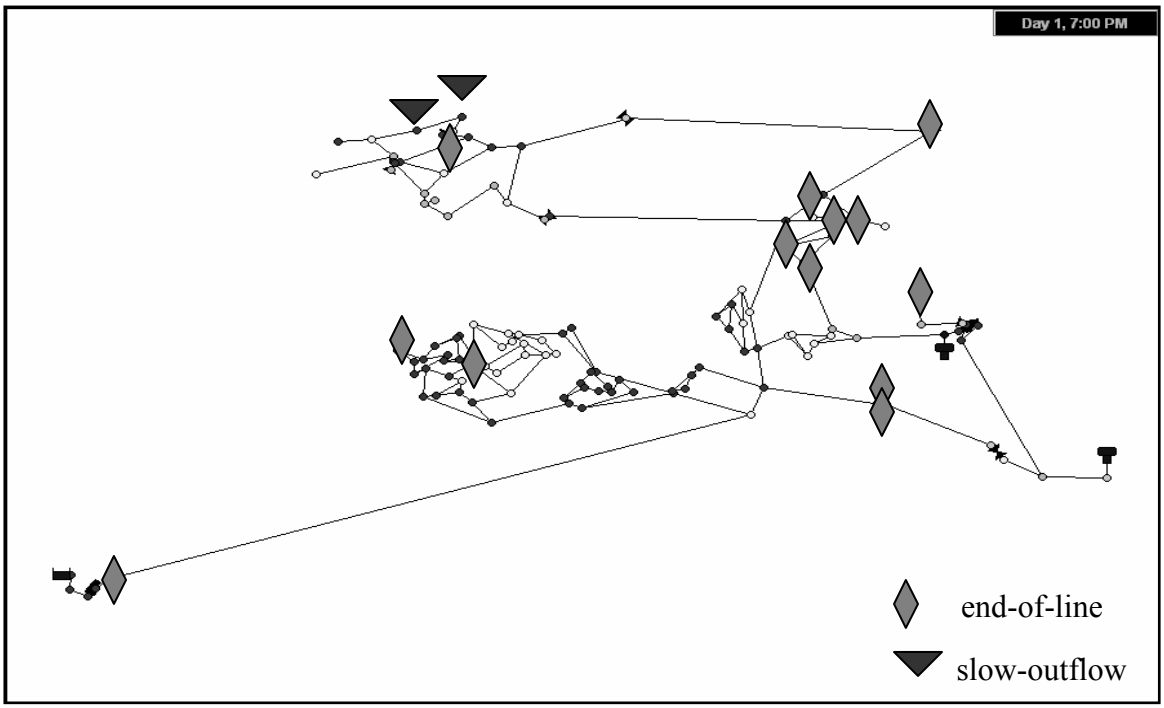


Figure 5.3 BWSN Network 1 end-of-line and slow-outflow node locations.

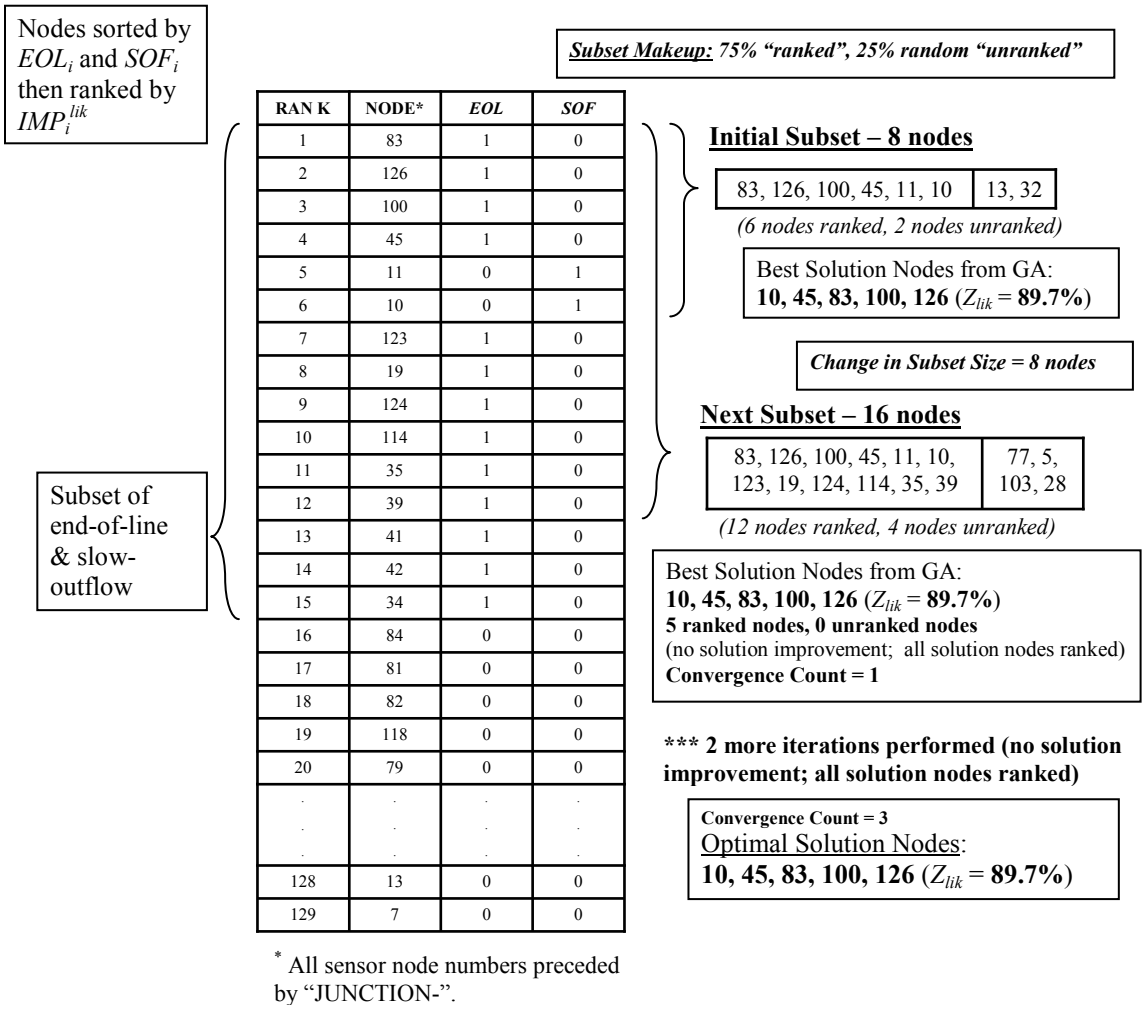


Figure 5.4 An illustration of ISSM applied to the single-objective problem of maximizing detection likelihood with $M = 5$ in the default attack case for BWSN Network 1.

Table 5.6 Parameters selected for use in ISSM for BWSN Network 1.

M	Initial Subset Size	Change in Subset Size	Proportion of Ranked Nodes in Subset	Proportion of Unranked Nodes in Subset
5	8	8	75%	25%
20	20	10	75%	25%

The one case for which all of the ISSM parameters for BWSN Network 1 in Table 5.6 do not apply is in the maximization of detection likelihood in the default attack case with 20 sensors employed. As noted previously, 15 nodes represent the endpoints of all contaminant pathways in BWSN Network 1; sensors at these nodes allow for the maximum possible detection likelihood to be achieved. These 15 nodes are the 13 end-of-line and 2 slow-outflow nodes constituting the top 15 ranked nodes in Figure 5.4. Thus, an initial subset size of 20 (with 15 ranked nodes and 5 unranked nodes under the ISSM parameters in Table 5.6) when 20 sensors are employed allows for the maximum detection likelihood possible to be found. ISSM as programmed would record the 20 nodes in the subset as the best solution and would conduct iteration until all of the unranked nodes in that solution are ranked, which may require a relatively large subset. Therefore, for this case, all ISSM subset nodes are ranked so that it is demonstrated that the optimal solution can be found with a subset of the minimum size possible.

The effectiveness of the nodal importance function and the sorting procedure is clear upon considering Tables 5.7 and 5.8. The tables indicate the rank of the solution sensor node ranked lowest in terms of importance for each case listed and thereby allow for the proportion of WDS nodes containing the optimal solution to be computed. For instance, it can be seen in Figure 5.4 that the 5 sensor nodes selected for the default attack case

with 5 sensors employed are ranked 1, 2, 3, 4, and 6, so lowest rank of 6 is reported in Table 5.7. Subsequently, the proportion of WDS nodes containing the solution is 6/129, or 4.7%. In some cases, the solution is isolated to smallest number of nodes possible – the number of sensors employed. Given the data in Tables 5.7 and 5.8, it is reasonable to claim that the nodal importance concepts developed for maximizing detection likelihood make it possible for the optimal sensor node solution for all cases to be isolated to relatively very small proportions of WDS nodes.

The benefit of including slow-outflow nodes in the sorting procedure when maximizing detection likelihood for small systems is demonstrated by the data in Tables 5.7 and 5.8. The solutions for employing 5 sensors include 4 end-of-line nodes and 1 slow-outflow node (JUNCTION-10). JUNCTION-10 has the rank of either 5 or 6 for all attack cases where the underlying assumption regarding a sensor detection resolution of zero (i.e., all cases but the one with a 0.3 mg/L detection resolution). If slow-outflow nodes were not included in the sorting procedure, then JUNCTION-10 would have been ranked much lower in these cases and would require a longer running of ISSM in order to determine the optimal sensor node solution.

Even without implementing the sorting procedure, ranking according to nodal importance value only should still result in the isolating of solution sensor nodes to small proportions of WDS nodes, as demonstrated by sensor placement carried out for the case of a 0.3 mg/L detection resolution. With a non-zero detection resolution, the end-of-line and slow-outflow conditions as defined in Section 4.5.2.1 do not apply, so nodes can only be

ranked according to nodal importance value. As indicated in Table 5.8, the lowest ranked sensor node in this case is ranked 11, which still makes for a reasonably small proportion of nodes containing the optimal solution.

Table 5.7 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for maximizing detection likelihood in the default attack case and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	6	4.7%
20	20	15.5%

Table 5.8 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for maximizing detection likelihood with $M = 5$ in the variant attack cases and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	6	4.7%
sensor detection resolution = 0.3 mg/L	11	8.5%
contaminant injection duration = 10 h	5	3.9%
2 injection nodes at 1 time	6	4.7%
2 injection nodes at 2 times	5	3.9%

In addition to noting the lowest ranked optimal sensor node, the effectiveness of nodal importance is further assessed in this study by analyzing solution convergence as ISSM is carried out for a particular protection goal, attack case, number of sensors employed, and study system. For this solution convergence examination, the ISSM stopping criteria of

the three consecutive ISSM iterations without solution improvement and all sensor nodes being "ranked" nodes is relaxed in order to test the validity of the stopping criteria. For maximizing detection likelihood in BWSN Network 1, the solution convergence plots in Figures 5.5 and 5.6 reinforce the effectiveness of nodal importance as the plots show rapid convergence to the optimal solution Z_{lik} value in each case. The optimal solution Z_{lik} value is found using the initial ISSM subset in all attack cases except in the case of a detection resolution of 0.3 mg/L. Even in that case, the optimal solution Z_{lik} value is found relatively quickly (i.e., after the first increase in subset size), and the Z_{lik} value found in the initial subset is very close to that of the optimal solution. Thus, solution convergence behavior confirms the extreme effectiveness of nodal importance in narrowing the search for the optimal sensor nodes to maximize detection likelihood in BWSN Network 1. Obviously, ISSM with parameters and stopping criteria chosen are adequate for carrying out optimization properly.

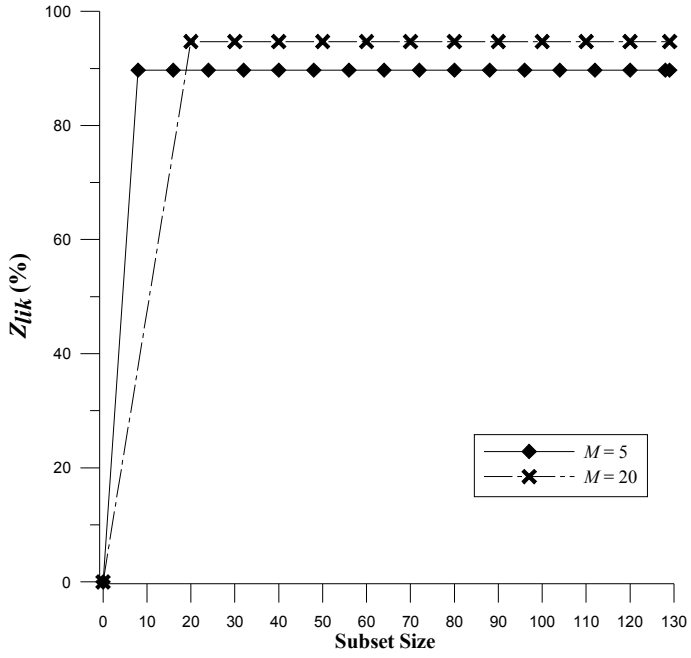


Figure 5.5 Solution convergence as ISSM is carried out for maximizing detection likelihood in the default attack case for BWSN Network 1.

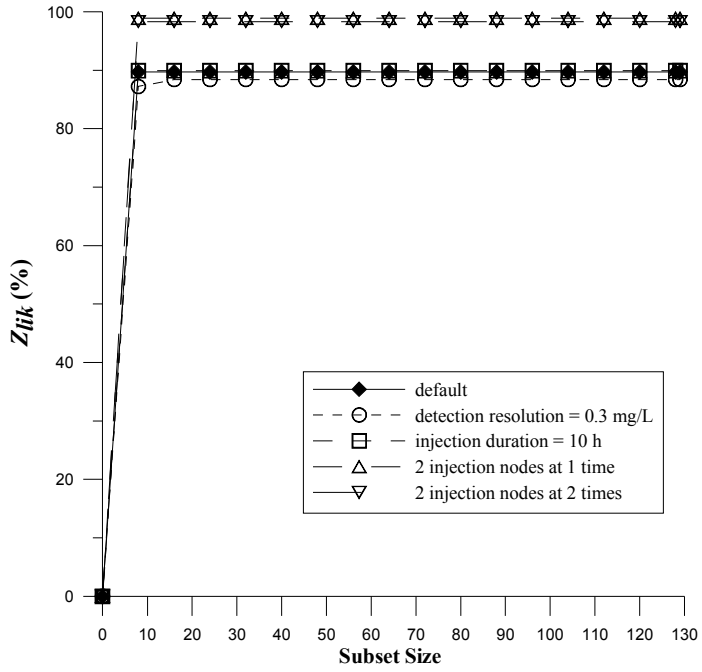


Figure 5.6 Solution convergence as ISSM is carried out for maximizing detection likelihood with $M = 5$ in the default and variant attack cases for BWSN Network 1.

5.3 SINGLE-OBJECTIVE PROBLEM: MINIMIZING Z_{time}

Sensor placement optimization results with respect to minimizing expected detection time are presented and discussed here.

5.3.1 SENSOR NODE SOLUTIONS

Table 5.9 provides the sensor placement optimization results for minimizing expected detection time in the default attack case, and Figures 5.7 and 5.8 illustrate optimal sensor node locations for employing 5 and 20 sensors, respectively. As stated previously, expected detection time is strongly correlated with detection likelihood; hence, it is not surprising that many of the sensor nodes in Table 5.9 are nodes close to those that are optimal for maximizing detection likelihood or are actually nodes selected for maximizing detection likelihood. For instance, in the case of 5 sensors employed, JUNCTION-11 is just upstream from JUNCTION-10, and JUNCTION-118 is just upstream from JUNCTION-126; JUNCTION-45, JUNCTION-83, and JUNCTION-100 are nodes optimal for both protection goals. It makes sense that JUNCTION-11 and JUNCTION-118 represent a shift upstream from the nodes selected for maximizing detection likelihood so that contamination can be detected sooner while mostly preserving a high detection likelihood. In other words, the results of optimization are sensor nodes that provide almost as much if not as much protection in terms of maximizing detection likelihood as nodes listed in Table 5.1 while significantly reducing Z_{time} values relative to those in Table 5.1.

Table 5.10 compares sensor nodes found in this work for minimizing expected detection time in the default attack case with those submitted by other works. The nodes selected in this work match those selected by Krause et al. (2006) in performance and are significantly better in performance than those selected by Eliades and Polycarpou (2006).

Table 5.9 Optimal sensor locations and performance measures for minimizing expected detection time in the default attack case for BWSN Network 1.

M	**Sensor Nodes***	Z_{lik} **(%)**	Z_{time}** **(min)**	Z_{vol} **(gal)**
5	11, 45, 83, 100, 118	88.3	1,436.8	16,145
20	10, 11, 14, 19, 21, 34, 35, 45, 68, 72, 75, 83, 90, 100, 101, 114, 118, 123, 124, 126	94.7	617.4	2,620

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

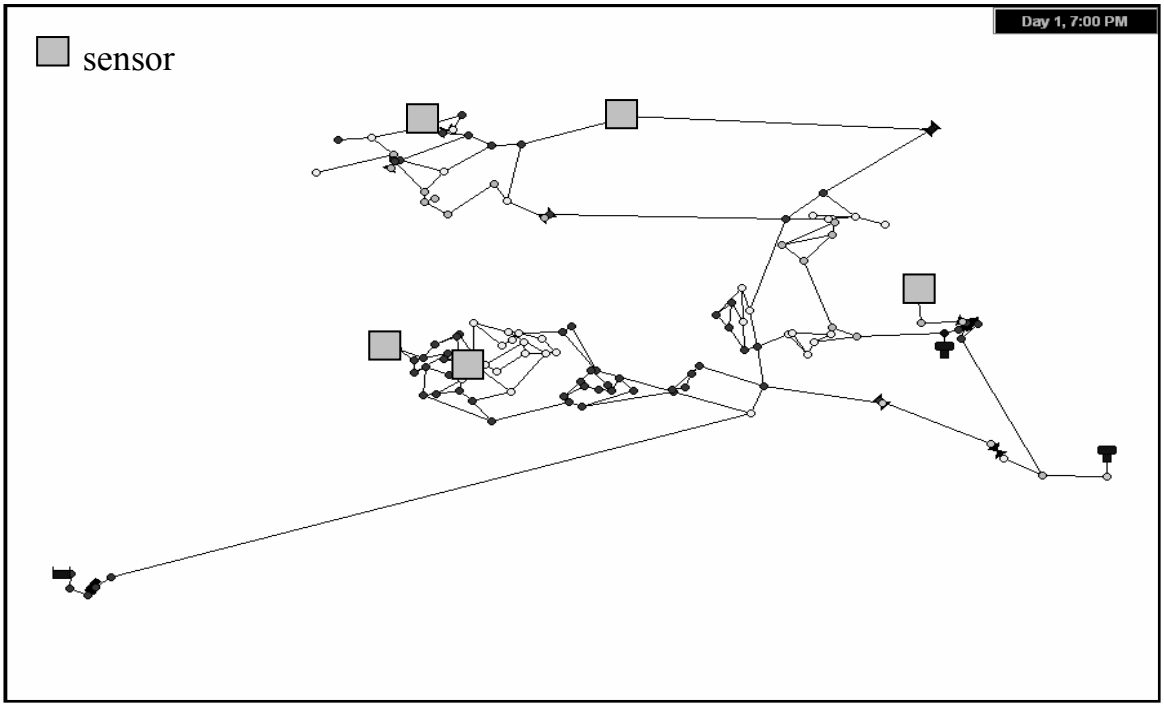


Figure 5.7 Optimal sensor locations with $M = 5$ for minimizing expected detection time in the default attack case for BWSN Network 1.

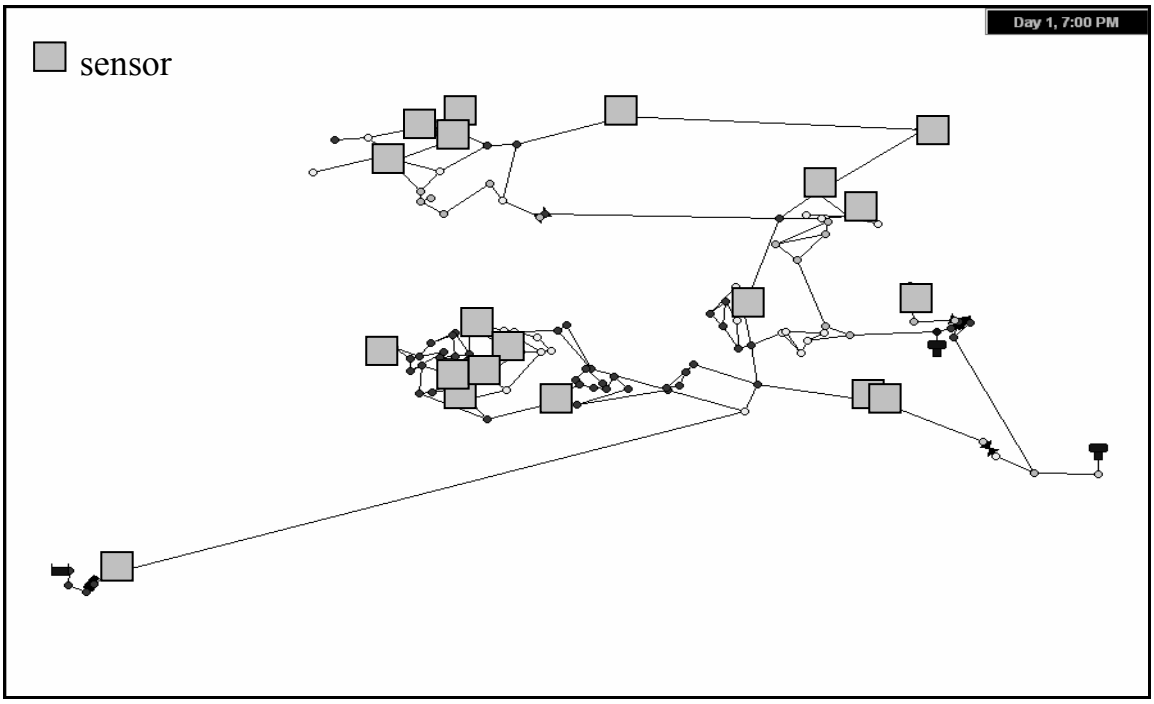


Figure 5.8 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the default attack case for BWSN Network 1.

Table 5.10 Comparison of optimal sensor locations for minimizing expected detection time in the default attack case for BWSN Network 1 determined in this study with locations found in other works.

M	**Work**	**Sensor Nodes***	Z_{time}** **(min)**
5	This Study	11, 45, 83, 100, 118	1,436.8
	Eliades and Polycarpou (2006)	17, 83, 101, 123, 126	1,581.3
	Krause et al. (2006)	11, 45, 83, 100, 118	1,436.8
20	This Study	10, 11, 14, 19, 21, 34, 35, 45, 68, 72, 75, 83, 90, 100, 101, 114, 118, 123, 124, 126	617.4
	Eliades and Polycarpou (2006)	10, 11, 14, 17, 19, 21, 31, 35, 45, 68, 74, 76, 83, 90, 100, 101, 114, 123, 124, 126	641.0
	Krause et al. (2006)	10, 11, 14, 19, 21, 34, 35, 45, 68, 72, 75, 83, 90, 100, 101, 114, 118, 123, 124, 126	617.4

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

Table 5.11 gives sensor node solutions for minimizing expected detection time with 5 sensors employed for the variant attack cases. As with maximizing detection likelihood, the attack cases of a hazard threshold of 0.0 mg/L and a response delay of 3 hours do not affect sensor placement for minimizing expected detection time, so sensor nodes in these cases would be the same as those for the default attack case with Z_{vol} values being greater than or equal to those associated with the default case.

Sensor nodes sets for all one injection node cases are the same with expected differences in Z_{time} values. Expected detection time is greater for an increased sensor detection resolution since less instances of detection leads to more scenarios associated with an "infinite" detection time. The opposite is true for the case of an increased injection duration; more instances of detection leads to less scenarios associated with an "infinite" detection time. The sensor node sets for the two injection node cases are interestingly different from those for the one injection node case. As shown in Figures 5.9 and 5.10, sensor nodes are more congregated toward the center of the system in these cases. With two injection nodes, sensors do not need to provide as much coverage by being located on the periphery of the system in order to reduce instances of "infinite" detection time, so sensors can afford to move farther upstream to detect contamination sooner. Though there is not a substantial difference in sensor node locations between the two different two injection node cases, there is a substantial difference in Z_{time} value between the two cases. This sensitivity to time differences between injections occurring at two nodes is due to hydraulic behavior in the system that does not allow a "first" injection to be detectible, though the time-after-injection value for a given scenario is increasing and

will contribute to the increasing of Z_{time} value even before a detectible "second" injection occurs.

As indicated in Table 5.12, the optimal sensor nodes found in this work for the case of a 10-hour injection duration are the same as those found by Krause et al. (2006) and are very similar to those of Krause et al. for the case of 2 nodes injected at 1 time. The only difference in the two injection nodes case is the selection of JUNCTION-46 in this work, which is upstream of JUNCTION-45 (selected by Krause et al.), allowing for slightly quicker contaminant detection on average and in turn a better Z_{time} value for this work.

Table 5.11 Optimal sensor locations and performance measures for minimizing expected detection time with $M = 5$ in the variant attack cases for BWSN Network 1.

Attack Case	Sensor Nodes*	Z_{lik} (%)	Z_{time}** (min)	Z_{vol} (gal)
default	11, 45, 83, 100, 118	88.3	1,436.8	16,145
sensor detection resolution = 0.3 mg/L	11, 45, 83, 100, 118	87.2	1,485.8	16,623
contaminant injection duration = 10 h	11, 45, 83, 100, 118	88.4	1,408.9	22,848
2 injection nodes at 1 time	46, 68, 83, 101, 118	96.6	576.3	10,091
2 injection nodes at 2 times	46, 70, 83, 101, 118	95.7	831.7	12,497

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

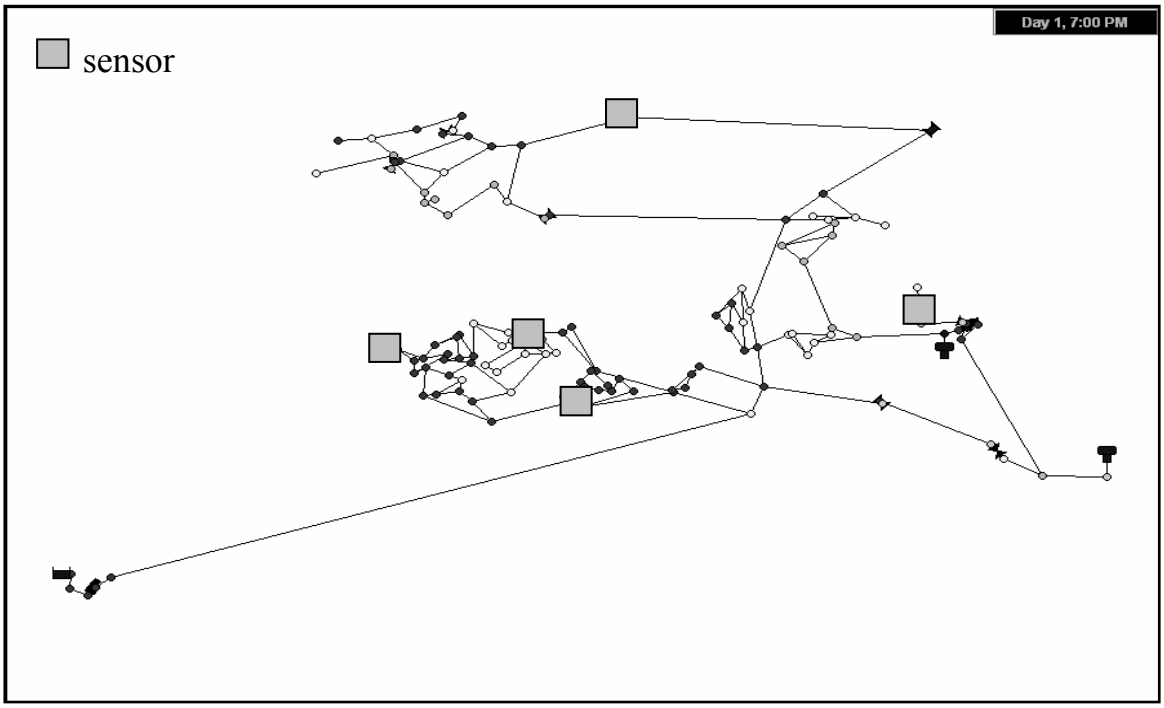


Figure 5.9 Optimal sensor locations with $M = 5$ for minimizing expected detection time in the case of 2 nodes injected at 1 time for BWSN Network 1.

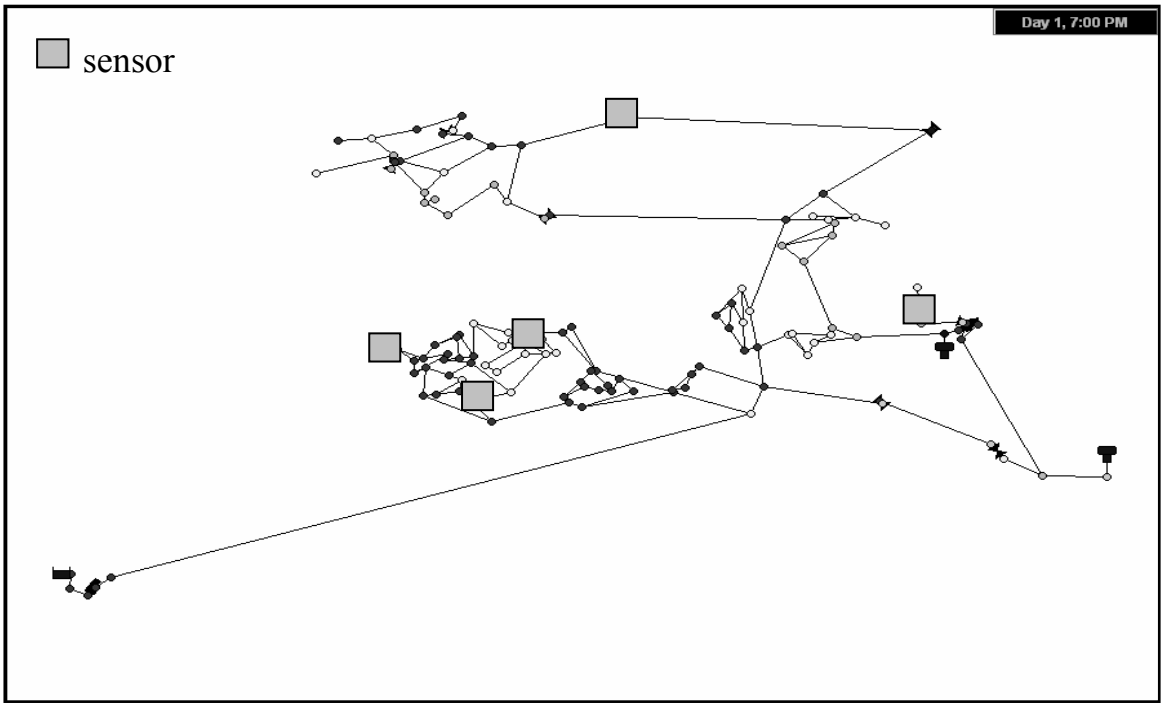


Figure 5.10 Optimal sensor locations with $M = 5$ for minimizing expected detection time in the case of 2 nodes injected at 2 times for BWSN Network 1.

Table 5.12 Comparison of optimal sensor locations for minimizing expected detection time with $M = 5$ in the variant attack cases for BWSN Network 1 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{time}** (min)
contaminant injection duration = 10 h	This Study	11, 45, 83, 100, 118	1,408.9
	Krause et al. (2006)	11, 45, 83, 100, 118	1,408.9
2 injection nodes at 1 time	This Study	46, 68, 83, 101, 118	576.3
	Krause et al. (2006)	45, 68, 83, 101, 118	592.2

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

5.3.2 EVALUATION OF NODAL IMPORTANCE

ISSM was carried out using the nodal importance function in (4.23) to find the optimal sensor nodes with respect to minimizing expected detection time presented above. An examination of the lowest-ranked sensor nodes of optimal solutions in Tables 5.13 and 5.14 and the solution convergence plots in Figures 5.11 and 5.12 reveals interesting relationships of the ability of nodal importance to narrow the search for the optimal solution with both the number of sensors employed and the particular attack case. In Table 5.13, it is reported that to find the optimal sensor node solution with 5 sensors employed, the solution can be isolated to the top 14% of the WDS nodes when ranked according to values calculated with (4.23). However, with 20 sensors employed, the ranking does not isolate the entire set of optimal sensor nodes to a small proportion of WDS nodes. This instance brings to light one drawback of nodal importance as formulated in this study: cases of a high number of sensors employed to number of WDS nodes ratio (M/N ratio). When the number of sensors is few (e.g., 5) and subsequently M/N ratio is small, nodal importance is very effective in identifying the nodes that would

provide the "major" protection with regard to a particular protection goal as sensor locations. In some cases when the number of sensors becomes relatively high (e.g., 20), the nodes providing the major protection as sensor locations are accounted for, and increased monitoring performance for the remaining sensor locations is much more a function of increasing detection likelihood. For minimizing expected detection time, increasing detection likelihood leads to a reduction of instances of "infinite" detection time. Unfortunately, the nodal importance concepts in this study are not designed to adjust to the "changed" priority of increasing detection likelihood after the major protection is accounted for. Despite the lack of complete narrowing of the search domain when 20 sensors are employed, Figure 5.11 shows that nodal importance provides for near-optimal solutions when ISSM subset size is relatively small.

When 5 sensors are employed, nodal importance provides for a very large degree of narrowing of the search domain for all attack cases studied with one injection node. However, for two injection node cases, achieving convergence to the optimal solution requires a larger subset, as implied by data in Table 5.14. As mentioned above, having two injection nodes allows for sensor nodes to shift upstream, but consequently the optimal sensor nodes are not necessarily important in terms of maximizing detection likelihood which would make the nodes in question not as important in terms of minimizing expected detection time. Thus, the ranks of the lowest ranked nodes for the two injection node cases are relatively low. As in the case of 20 sensors employed in the default attack case, nodal importance does provide ISSM the means for finding near-optimal solutions with smaller subsets, evidenced by the solution convergence plot in Figure 5.12.

Table 5.13 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for minimizing expected detection time for the default attack case and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	18	14.0%
20	117	90.7%

Table 5.14 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for minimizing expected detection time with $M = 5$ in the variant attack cases and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	18	14.0%
sensor detection resolution = 0.3 mg/L	18	14.0%
contaminant injection duration = 10 h	18	14.0%
2 injection nodes at 1 time	45	34.9%
2 injection nodes at 2 times	42	32.6%

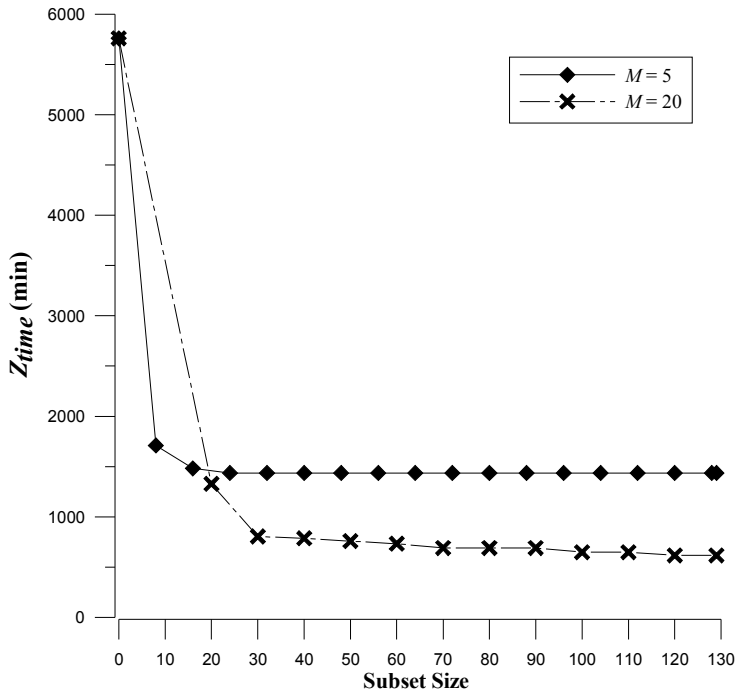


Figure 5.11 Solution convergence as ISSM is carried out for minimizing expected detection time in the default attack case for BWSN Network 1.

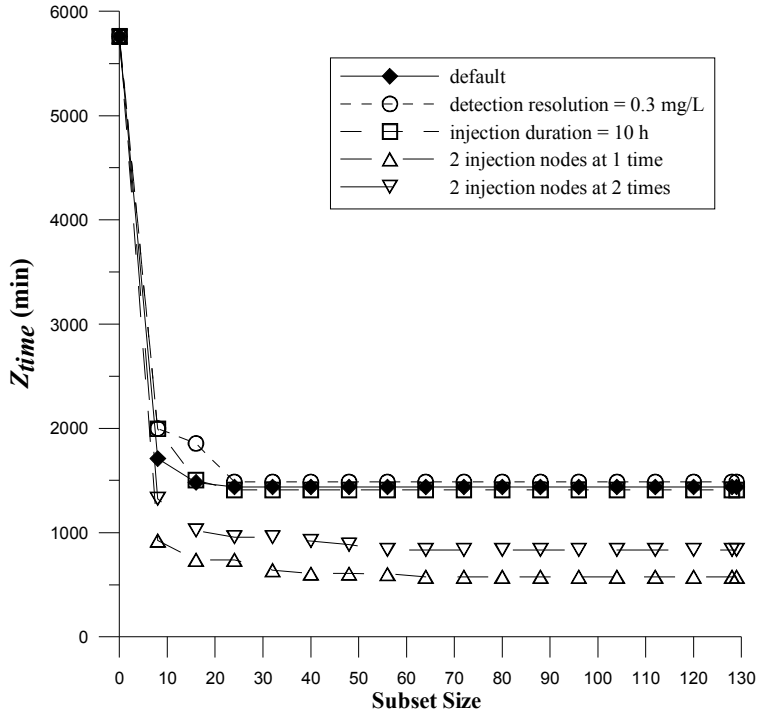


Figure 5.12 Solution convergence as ISSM is carried out for minimizing expected detection time with $M = 5$ in the default and variant attack cases for BWSN Network 1.

5.4 SINGLE-OBJECTIVE PROBLEM: MINIMIZING Z_{vol}

The optimization findings regarding the minimizing of expected contaminated demand volume are presented and discussed in this section.

5.4.1 SENSOR NODE SOLUTIONS

Table 5.15 and Figures 5.13 and 5.14 indicate the sensor nodes selected in this work to minimize expected contaminated demand volume in the default attack case for BWSN Network 1. Based on the data given in the table and how they relate to data given in Tables 5.1 and 5.9, it is obvious that sensors at the nodes listed would provide a significant reduction of volume contaminated on average at the cost of some performance in terms of Z_{lik} and Z_{time}. In Figure 5.13, sensor nodes are more upstream relative to those for maximizing detection likelihood and minimizing expected detection time in Figures 5.1 and 5.7, respectively. This overall shift in sensor node location is reasonable as it is most important to detect contaminant in a timely manner in order to minimize the volume that could potentially be contaminated farther downstream. The sensor nodes selected for the default attack case are at least as good as or better than those submitted by other works in minimizing expected contaminated demand volume, as indicated in Table 5.16.

Table 5.15 Optimal sensor locations and performance measures for minimizing expected contaminated demand volume in the default attack case for BWSN Network 1.

M	Sensor Nodes*	Z_{lik} (%)	Z_{time} (min)	Z_{vol}** (gal)
5	17, 22, 68, 79, 102	66.9	2,256.9	3,864
20	4, 17, 21, 28, 30, 31, 34, 37, 46, 49, 68, 74, 79, 83, 90, 98, 102, 118, 122, 126	84.2	1,115.2	486

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

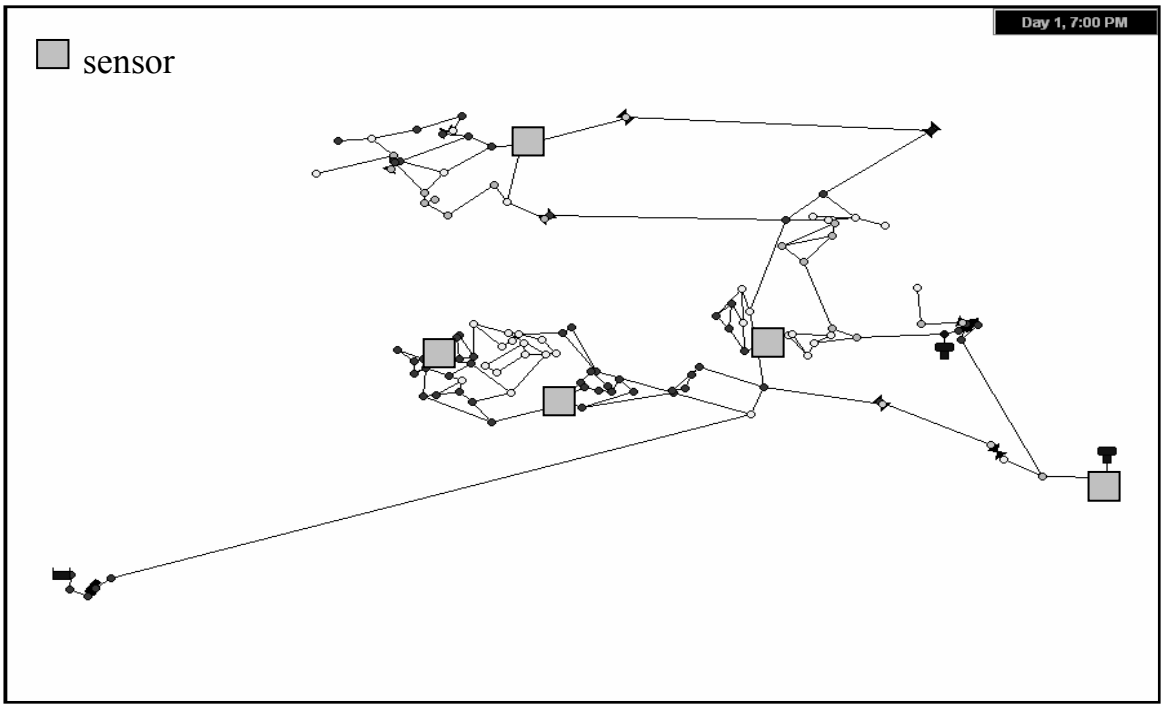


Figure 5.13 Optimal sensor locations with $M = 5$ for minimizing expected contaminated demand volume in the default attack case for BWSN Network 1.

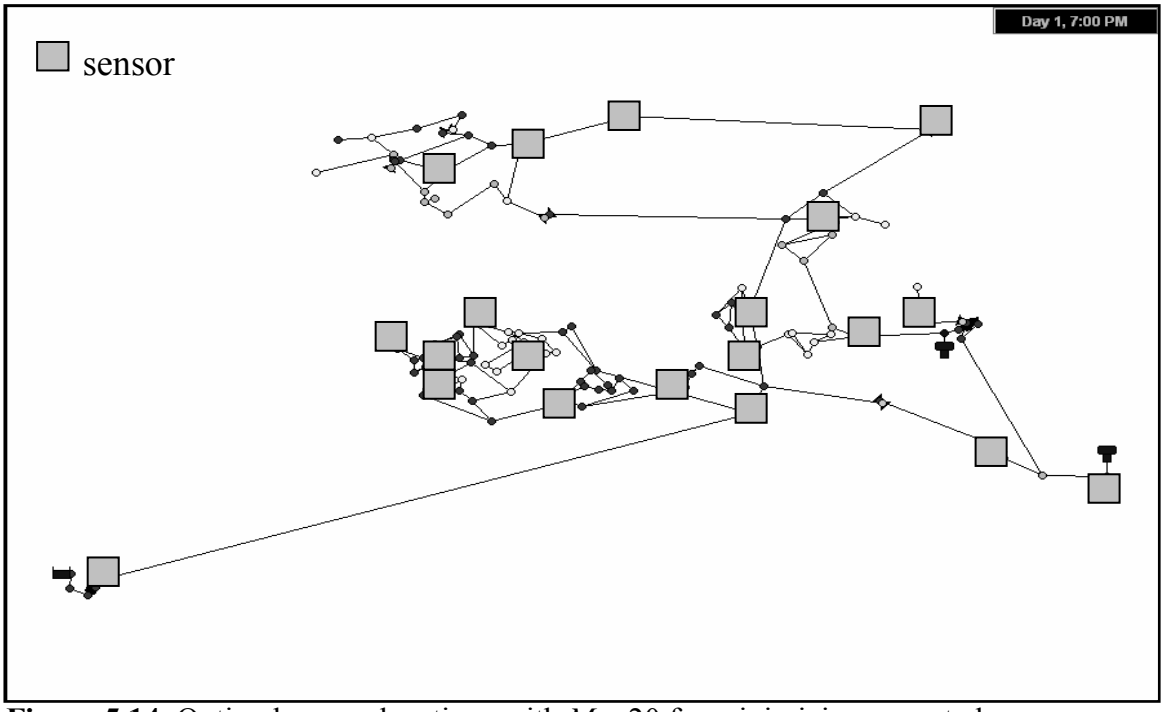


Figure 5.14 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the default attack case for BWSN Network 1.

Table 5.16 Comparison of optimal sensor locations for minimizing expected contaminated demand volume in the default attack case for BWSN Network 1 determined in this study with locations found in other works.

M	**Work**	**Sensor Nodes***	***Z_{vol}*** ** **(gal)**
5	This Study	17, 22, 68, 79, 102	3,864
	Berry et al. (2006c)	17, 21, 68, 79, 122	4,670
	Eliades and Polycarpou (2006)	17, 83, 101, 123, 126	11,415
	Krause et al. (2006)	17, 49, 68, 79, 102	3,887
20	This Study	4, 17, 21, 28, 30, 31, 34, 37, 46, 49, 68, 74, 79, 83, 90, 98, 102, 118, 122, 126	486
	Berry et al. (2006c)	3, 4, 17, 21, 25, 31, 34, 37, 46, 64, 68, 81, 82, 90, 98, 102, 116, 118, 122, 126	522
	Eliades and Polycarpou (2006)	10, 11, 14, 17, 19, 21, 30, 37, 45, 68, 74, 83, 90, 100, 102, 114, 118, 123, 124, 126	1,930
	Krause et al. (2006)	5, 17, 21, 29, 30, 31, 34, 37, 46, 49, 68, 74, 79, 83, 94, 97, 102, 118, 122, 126	518

* All sensor node numbers preceded by "JUNCTION-".

** Values computed using the software developed in this work.

Table 5.17 gives the sensor placement results for minimizing expected contaminant demand volume in the variant attack cases. As seen in comparing Figure 5.13 with Figures 5.15 and 5.16, the locations of sensor nodes do not deviate in location greatly from attack case to attack case, which is not unexpected given results presented previously for the other protection goals. Z_{vol} values of different attack cases differ from each other in expected manners. The higher Z_{vol} value for the case of a non-zero detection resolution is a consequence of the detection of contamination for fewer scenarios; in these additional scenarios of non-detection, large amounts of volume are contaminated and hence Z_{vol} increases. The substantially higher Z_{vol} value for the case of an increased injection duration is primarily due to the longer period of time allowed for water in the system to become contaminated under various hydraulic conditions such that contaminated water travels to more locations. The increase in Z_{vol} value when the contaminant hazard threshold is reduced to zero is the result of more demand volume being considered contaminated, but this increase is small for the sensors are placed upstream enough to detect contamination before contaminant concentration at locations throughout the system falls to within the range 0.0 to 0.3 mg/L. With an extra 3 hours of contaminant propagation after detection, it is not surprising that Z_{vol} for the response delay case is much higher than it is for default attack case. Greater Z_{vol} values for the two injection node cases are due to the increased opportunity for contaminated water to reach more places in system with an additional injection site. Table 5.18 indicates that the results found in this work for the variant attack cases also studied in BWSN are at least equal if not better than those of other works.

Table 5.17 Optimal sensor locations and performance measures for minimizing expected contaminated demand volume with $M = 5$ in the variant attack cases for BWSN Network 1.

Attack Case	Sensor Nodes*	Z_{lik} (%)	Z_{time} (min)	Z_{vol}** (gal)
default	17, 22, 68, 79, 102	66.9	2,256.9	3,864
sensor detection resolution = 0.3 mg/L	17, 22, 68, 79, 102	63.9	2,405.9	3,938
contaminant injection duration = 10 h	22, 68, 79, 102, 118	68.2	2,205.5	8,358
contaminant hazard threshold = 0.0 mg/L	17, 22, 68, 79, 102	66.9	2,256.9	3,887
response delay = 3 h	17, 26, 68, 79, 102	66.9	2,252.3	13,111
2 injection nodes at 1 time	17, 22, 68, 79, 102	88.5	976.9	4,742
2 injection nodes at 2 times	17, 22, 68, 79, 102	87.3	1,244.1	4,262

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

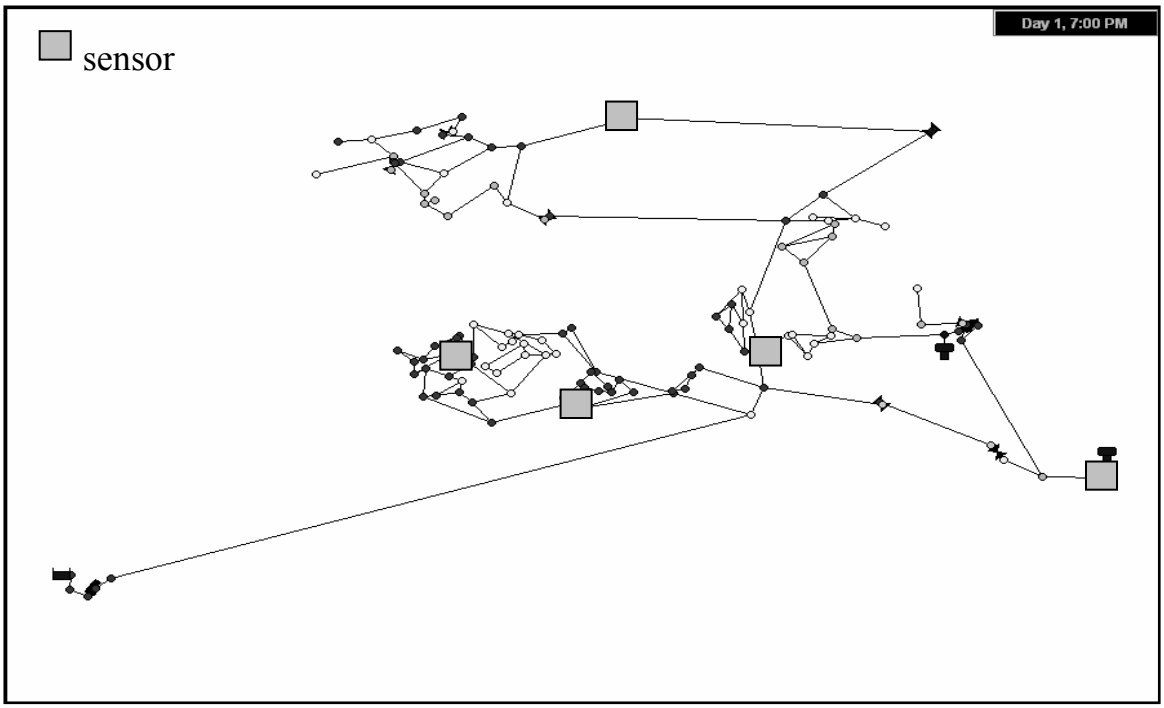


Figure 5.15 Optimal sensor locations with $M = 5$ for minimizing expected contaminated demand volume in the case of a 10-hour injection duration for BWSN Network 1.

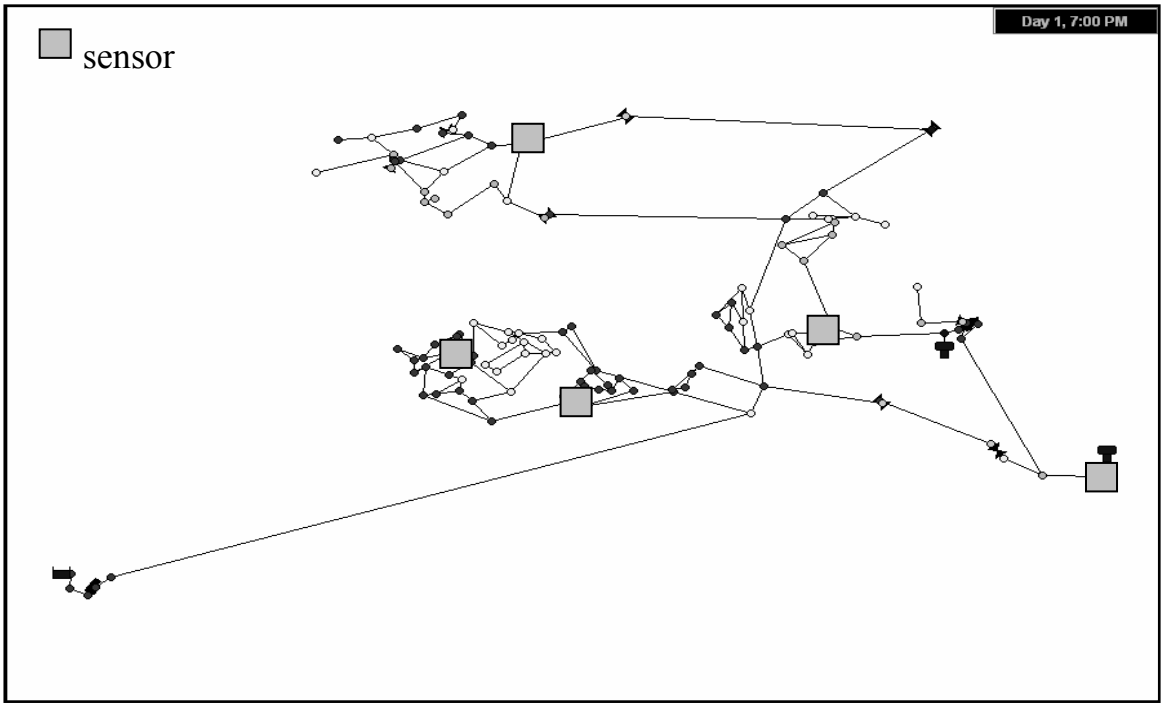


Figure 5.16 Optimal sensor locations with $M = 5$ for minimizing expected contaminated demand volume in the case of a 3-hour response delay for BWSN Network 1.

Table 5.18 Comparison of optimal sensor locations for minimizing expected contaminated demand volume with $M = 5$ in the variant attack cases for BWSN Network 1 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{vol}** (gal)
contaminant injection duration = 10 h	This Study	22, 68, 79, 102, 118	8,358
	Berry et al. (2006c)	20, 68, 79, 118, 122	9,423
	Krause et al. (2006)	17, 31, 81, 103, 118	10,535
response delay = 3 h	This Study	17, 26, 68, 79, 102	13,111
	Berry et al. (2006c)	17, 49, 68, 79, 103	13,127
	Krause et al. (2006)	17, 49, 68, 79, 103	13,127
2 injection nodes at 1 time	This Study	17, 22, 68, 79, 102	4,742
	Krause et al. (2006)	17, 49, 68, 97, 122	5,621

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

5.4.2 EVALUATION OF NODAL IMPORTANCE

To find the results presented above, ISSM was carried out with the aid of nodes ranked according to the small-system nodal importance function for minimizing expected contaminated demand volume in (4.28). Tables 5.19 and 5.20 imply that the nodal importance function is not as effective on average in narrowing the search for the optimal solution with respect to minimizing expected contaminated demand volume as the functions for maximizing Z_{lik} and minimizing Z_{time} were for finding the respective optimal solutions. However, the nodal importance function does allow the search domain to be narrowed significantly to approximately 30 to 35% of WDS nodes when 5 sensors are employed. The solution convergence plots in Figures 5.17 and 5.18 reinforce the need for a broader search domain for this single-objective problem versus those for the other two single-objective problems. It appears that a subset size of at least 30 is needed to find a solution with a Z_{vol} value reasonably close to that of the optimal solution for all attack cases. This decrease in narrowing ability highlights the drawback of multiple factors contributing to nodal importance for minimizing expected contaminated demand volume for small systems (i.e., nodal contaminated outflow, detection time, and importance with respect to maximizing detection likelihood), which can make some nodes somewhat less distinguishable from each other.

In the default attack case with 20 sensors employed, the drawback of high M/N ratio surfaces as it did in the problem of minimizing expected detection time with 20 sensors. Similar to that case, optimal sensor placement becomes more a function of detecting

contamination for more scenarios once the major protection is addressed in sensor placement.

A clear advantage of including the random unranked nodes in ISSM subsets is seen in the solution convergence plot in Figure 5.17. To elaborate, the solution plotted in that figure for $M = 5$ and subset size 16 contains 2 ranked nodes and 3 unranked nodes, which allows the solution to be superior to the one found for $M = 20$ and subset size 20. Thus, even though a subset size of 30 or more is needed to find a reliably close-to-optimal Z_{vol} value, the randomly included nodes in ISSM subsets provide a better chance for better solutions to be found when the ranked nodes in a subset do not include very good sensor node candidates.

Table 5.19 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for minimizing expected contaminated demand volume for the default attack case and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	42	32.6%
20	113	87.6%

Table 5.20 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for minimizing expected contaminated demand volume with $M = 5$ for the variant attack cases and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

Attack Case	Lowest Nodal Importance Rank	Proportion of Ranked WDS Nodes Containing Optimal Solution
default	42	32.6%
sensor detection resolution = 0.3 mg/L	40	31.0%
contaminant injection duration = 10 h	42	32.6%
contaminant hazard threshold = 0.0 mg/L	46	35.7%
response delay = 3 h	42	32.6%
2 injection nodes at 1 time	45	34.9%
2 injection nodes at 2 times	37	28.7%

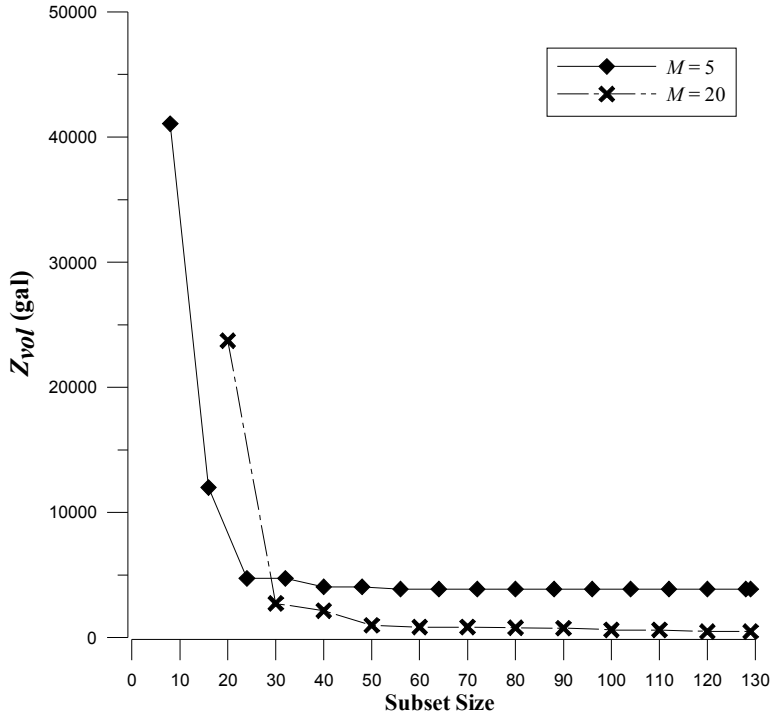


Figure 5.17 Solution convergence as ISSM is carried out for minimizing expected contaminated demand volume in the default attack case for BWSN Network 1.

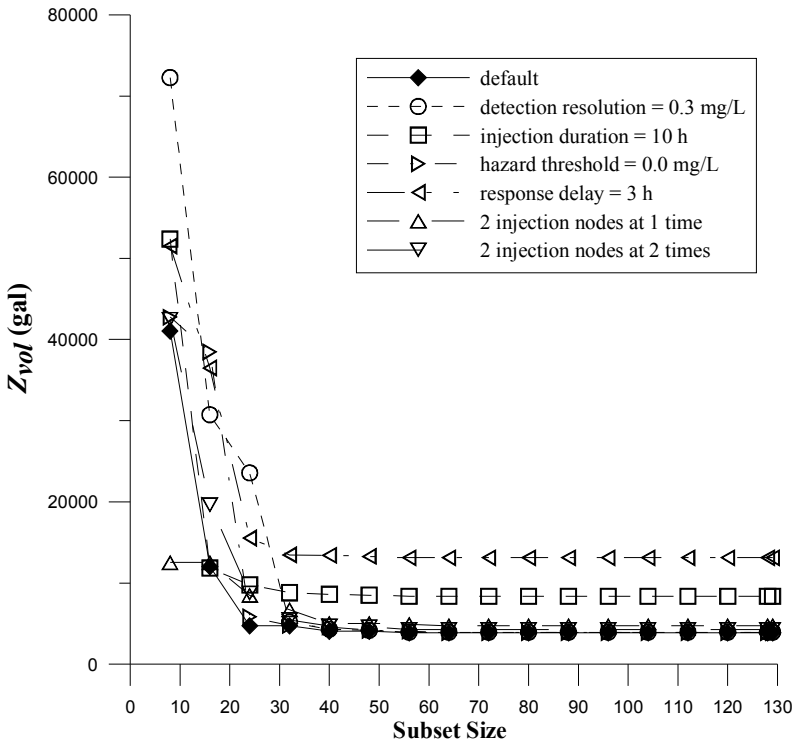


Figure 5.18 Solution convergence as ISSM is carried out for minimizing expected contaminated demand volume with $M = 5$ in the default and variant attack cases for BWSN Network 1.

5.5 MULTIOBJECTIVE PROBLEM: MAXIMIZING Z_{all}

The sensor placement findings for the three-prong multiobjective problem of this work are given and discussed here.

5.5.1 SENSOR NODE SOLUTIONS

Table 5.21 provides the optimal sensor node solutions for the three-prong multiobjective solution for the default attack case found in this work, and the locations of the sensor

nodes selected are depicted in Figures 5.19 and 5.20. The most apparent quality seen in these solutions is that they exactly match or nearly match those for minimizing Z_{time} presented above. The sensor nodes for $M = 5$ are exactly the same as those used to minimize Z_{time}, while the sensor nodes for $M = 20$ are only slightly rearranged relative to those used to minimize Z_{time}. These results are very reasonable, though, given the way the multiobjective optimization problem is formulated in this study with regard to minimizing protection tradeoffs between individual protection goals. For instance, the multiobjective solution sensor nodes with 5 sensors employed matching the sensor nodes for minimizing Z_{time} provide for full protection according to the minimizing Z_{time} protection goal. Since Z_{time} and Z_{lik} are strongly correlated with each other, these nodes also satisfy to a large degree the maximizing Z_{lik} goal. As detecting contamination sooner also benefits in the reducing of contaminated volume, there is not a great deal of sacrifice with regard to the minimizing Z_{vol} goal in selecting these sensor nodes. This minimization of tradeoff is represented with the high value for Z_{all} in the table for both $M = 5$ and $M = 20$ (i.e, $Z_{all} > 2.7$ out of 3).

Given the multiobjective problem formulation in this study, Pareto front space boundary values are needed to evaluate how well a particular multiobjective solution minimizes the protection sacrifices according to the three single-objective protection goals. To provide an indication of how these boundary values are determined, Table 5.22 gives the boundary values for BWSN Network 1 in the default attack case. Since these values are obtained from solving the three single-objective problems, the tables above where those values can be found are also given in Table 5.22 in order to illustrate the connection of the single-objective and multiobjective problems.

Table 5.21 Optimal sensor locations and performance measures for the three-prong multiobjective problem in the default attack case for BWSN Network 1.

M	Sensor Nodes*	Z_{all}	Z_{lik} (%)	Z_{time} (min)	Z_{vol} (gal)
5	11, 45, 83, 100, 118	2.72	88.3	1,436.8	16,145
20	10, 12, 19, 22, 34, 35, 39, 45, 68, 75, 79, 83, 97, 100, 102, 114, 117, 123, 124, 126	2.79	94.7	645.9	1,122

* All sensor node numbers preceded by "JUNCTION-".

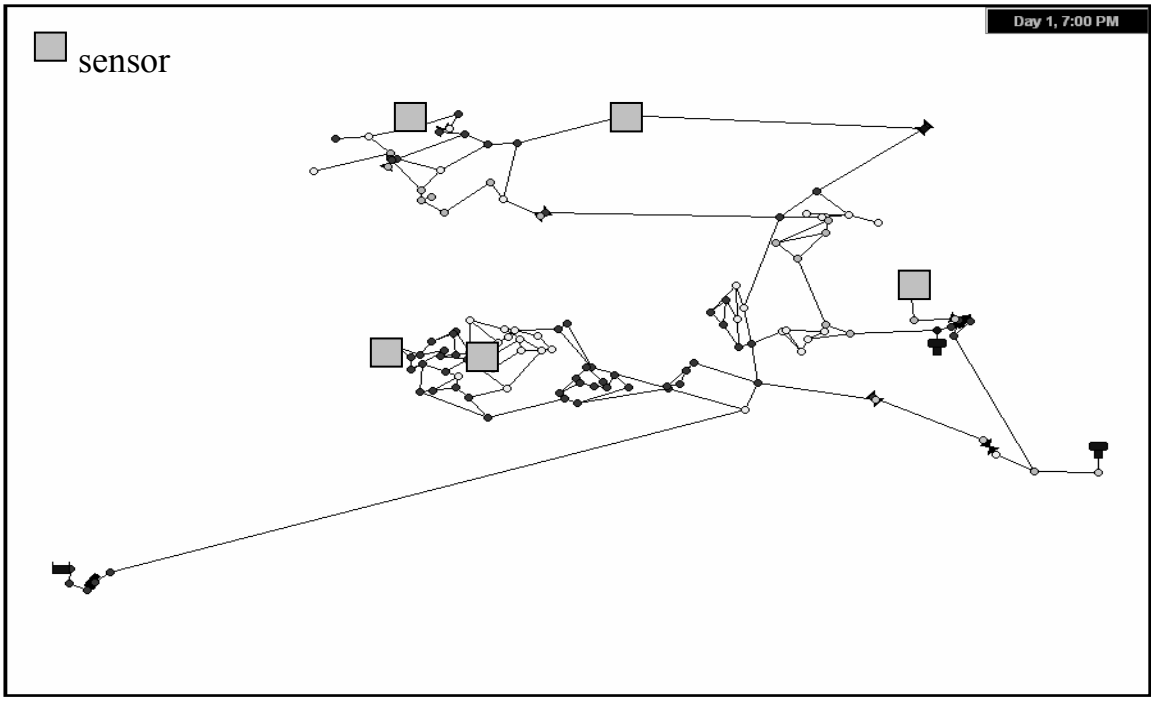


Figure 5.19 Optimal sensor locations with $M = 5$ for the three-prong multiobjective problem in the default attack case for BWSN Network 1.

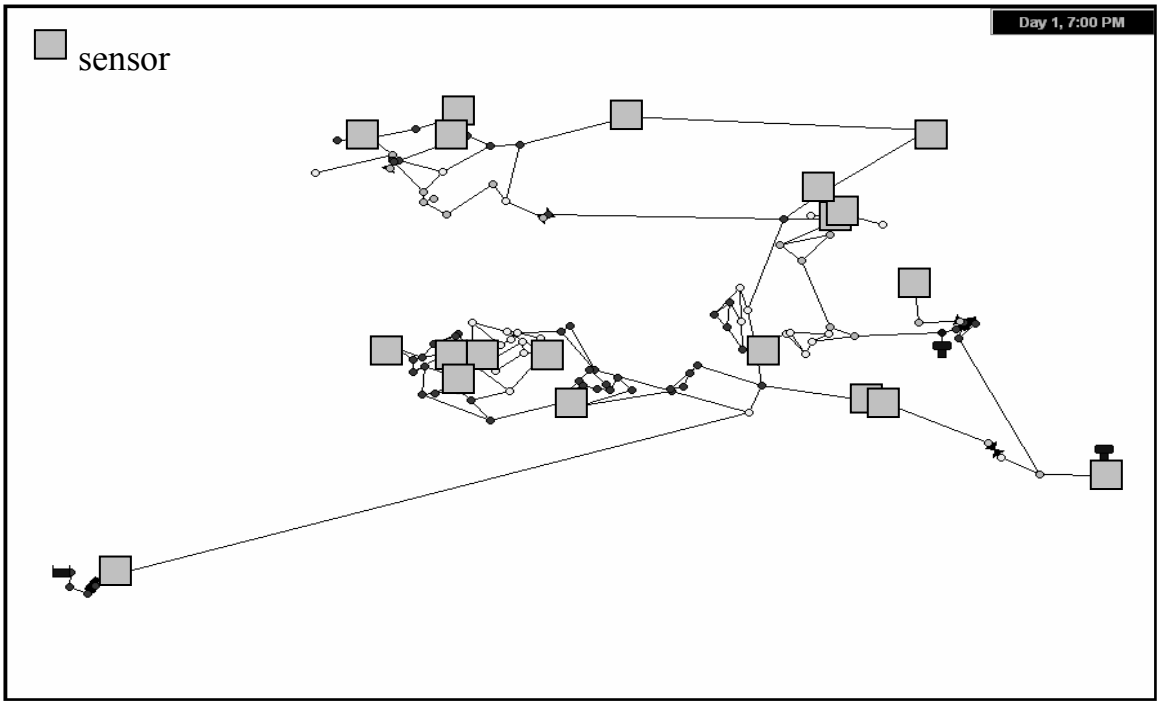


Figure 5.20 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the default attack case for BWSN Network 1.

Table 5.22 Pareto front space boundary values for the three-prong multiobjective program with $M = 5$ in the default attack case for BWSN Network 1. Accompanying each value is the number of the table above in which the value can be found.

Protection Goal Performance Measure	**Pareto Front Space Minimum Value** ***(Table Found)***	**Pareto Front Space Maximum Value** ***(Table Found)***
Z_{lik} (%)	66.9 *(5.15)*	89.7 *(5.1)*
Z_{time} (min)	1,436.8 *(5.9)*	2,256.9 *(5.15)*
Z_{vol} (gal)	3,864 *(5.15)*	59,701 *(5.1)*

Table 5.23 compares the multiobjective sensor nodes of this work with those of other works. However, these are not straightforward comparisons like those for the single-objective problems; the other works listed in the table included other performance

measures, defined the three protection goal performance measures differently, and/or weighted the protection goals in terms of priority in different manners. The solutions of other works are used in comparison only to ensure that other works have not provided superior results in terms of the multiobjective problem as defined in this work. Indeed, under this problem definition, the Z_{all} values for solutions of this work in the default attack case are equal or higher than those found with solutions from other works, as shown in Table 5.23.

Table 5.23 Comparison of optimal sensor locations for the three-prong multiobjective problem in the default attack case for BWSN Network 1 determined in this study with locations found in other works.

M	Work	Sensor Nodes*	Z_{all}**	Z_{lik}** (%)	Z_{time}** (min)	Z_{vol}** (gal)
5	This Study	11, 45, 83, 100, 118	2.72	88.3	1,436.8	16,145
	Aral et al. (2008)	10, 45, 83, 100, 118	2.66	88.9	1,507.5	16,150
		10, 68, 83, 118, 122	2.53	85.0	1,596.7	7,802
	Dorini et al. (2006)	10, 31, 45, 83, 118	2.51	86.4	1,642.5	8,873
	Eliades and Polycarpou (2006)	17, 31, 45, 83, 126	2.29	82.6	1,695.7	8,244
	Guan et al. (2006)	17, 31, 81, 98, 102	1.62	73.5	1,973.1	4,657
	Huang et al. (2006)	68, 81, 82, 97, 118	1.19	72.5	1,983.3	25,874
	Krause et al. (2006)	17, 31, 45, 83, 122	2.33	82.1	1,688.7	5,766
	Propato and Piller (2006)	17, 22, 68, 83, 123	2.23	79.9	1,691.5	5,561
	Wu and Walski (2006)	45, 68, 83, 100, 118	2.62	84.4	1,462.0	10,177
20	This Study	10, 12, 19, 22, 34, 35, 39, 45, 68, 75, 79, 83, 97, 100, 102, 114, 117, 123, 124, 126	2.79	94.7	645.9	1,122
	Dorini et al. (2006)	0, 10, 14, 17, 31, 34, 39, 45, 49, 68, 74, 82, 83, 90, 100, 102, 114, 122, 124, 128	2.28	92.3	803.3	979
	Eliades and Polycarpou (2006)	10, 11, 14, 17, 19, 21, 31, 35, 45, 68, 74, 83, 90, 100, 102, 114, 118, 123, 124, 126	2.53	94.7	659.5	2,131
	Guan et al. (2006)	4, 11, 17, 21, 27, 31, 34, 35, 46, 68, 75, 79, 82, 83, 98, 100, 102, 118, 122, 126	2.04	90.0	862.7	577
	Huang et al. (2006)	8, 11, 42, 46, 52, 68, 70, 75, 76, 82, 83, 95, 97, 99, 100, 109, 111, 117, 123, 126	1.51	90.8	861.5	3,149
	Krause et al. (2006)	10, 11, 17, 19, 21, 31, 34, 35, 45, 68, 76, 83, 90, 100, 114, 118, 122, 123, 124, 126	2.67	94.7	676.6	1,393
	Propato and Piller (2006)	11, 17, 34, 37, 38, 45, 49, 68, 76, 83, 90, 100, 102, 106, 114, 118, 123, 124, 125, 126	2.52	93.2	754.1	757
	Wu and Walski (2006)	10, 12, 19, 21, 34, 35, 40, 45, 68, 75, 80, 83, 98, 100, 102, 114, 118, 123, 124, 126	2.77	94.7	640.5	1,269

* All sensor node numbers preceded by "JUNCTION-".

** Values computed using the software developed in this work.

As indicated in Table 5.24, the preference for sensor nodes that provide for nearly or fully minimizing Z_{time} being selected for the three-prong multiobjective problem is also seen in studying the variant attack cases with 5 sensors employed. Figures 5.21 and 5.22 depict the locations of the sensor nodes for the two injection node cases—the cases for which the solution node set deviates somewhat from that found in the default attack case. The solutions perform as well as or better than those of other works under the multiobjective problem definition of this work, as implied data presented in Table 5.25.

Table 5.24 Optimal sensor locations and performance measures for the three-prong multiobjective problem with $M = 5$ in the variant attack cases for BWSN Network 1.

Attack Case	Sensor Nodes*	Z_{all}	Z_{lik} (%)	Z_{time} (min)	Z_{vol} (gal)
default	11, 45, 83, 100, 118	2.72	88.3	1,436.8	16,145
sensor detection resolution = 0.3 mg/L	11, 45, 83, 100, 118	2.72	87.2	1,485.8	16,623
contaminant injection duration = 10 h	11, 45, 83, 100, 118	2.72	88.4	1,408.9	22,848
contaminant hazard threshold = 0.0 mg/L	11, 45, 83, 100, 118	2.71	88.3	1,436.8	18,837
response delay = 3 h	11, 45, 83, 100, 118	2.61	88.3	1,436.8	37,833
2 injection nodes at 1 time	45, 68, 83, 118, 122	2.75	97.2	589.8	9,288
2 injection nodes at 2 times	46, 70, 83, 101, 118	2.63	95.7	831.7	12,497

* All sensor node numbers preceded by "JUNCTION-".

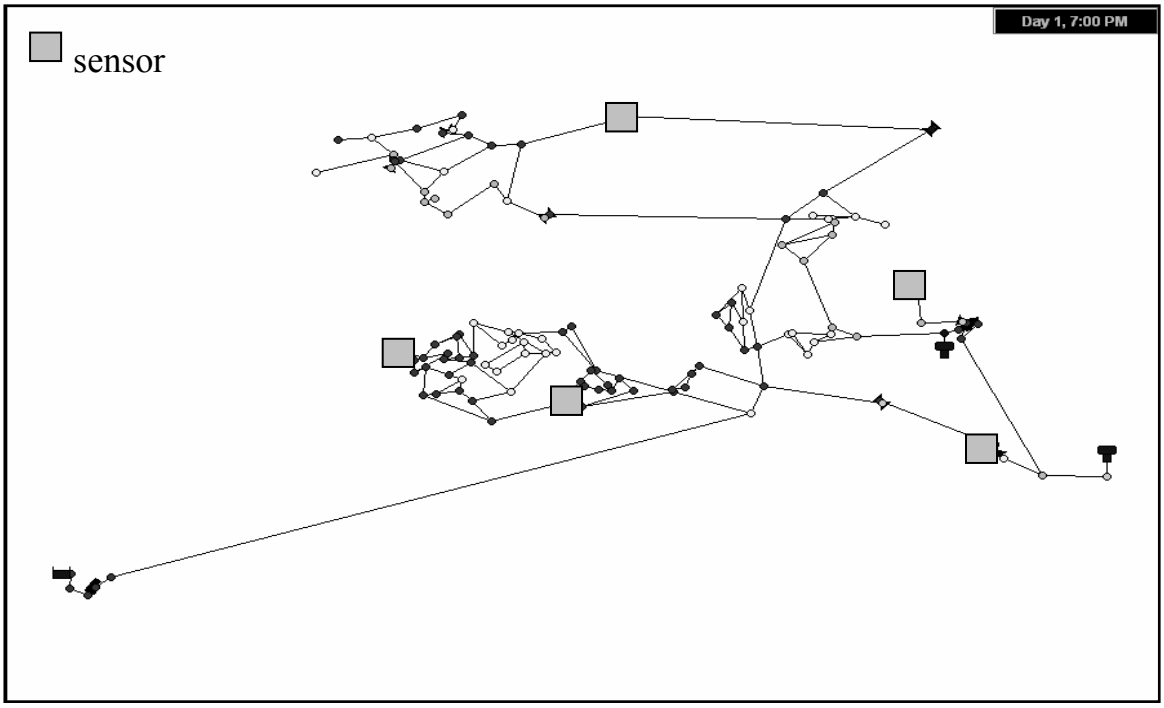


Figure 5.21 Optimal sensor locations with $M = 5$ for the three-prong multiobjective problem in the case of 2 nodes injected at 1 time for BWSN Network 1.

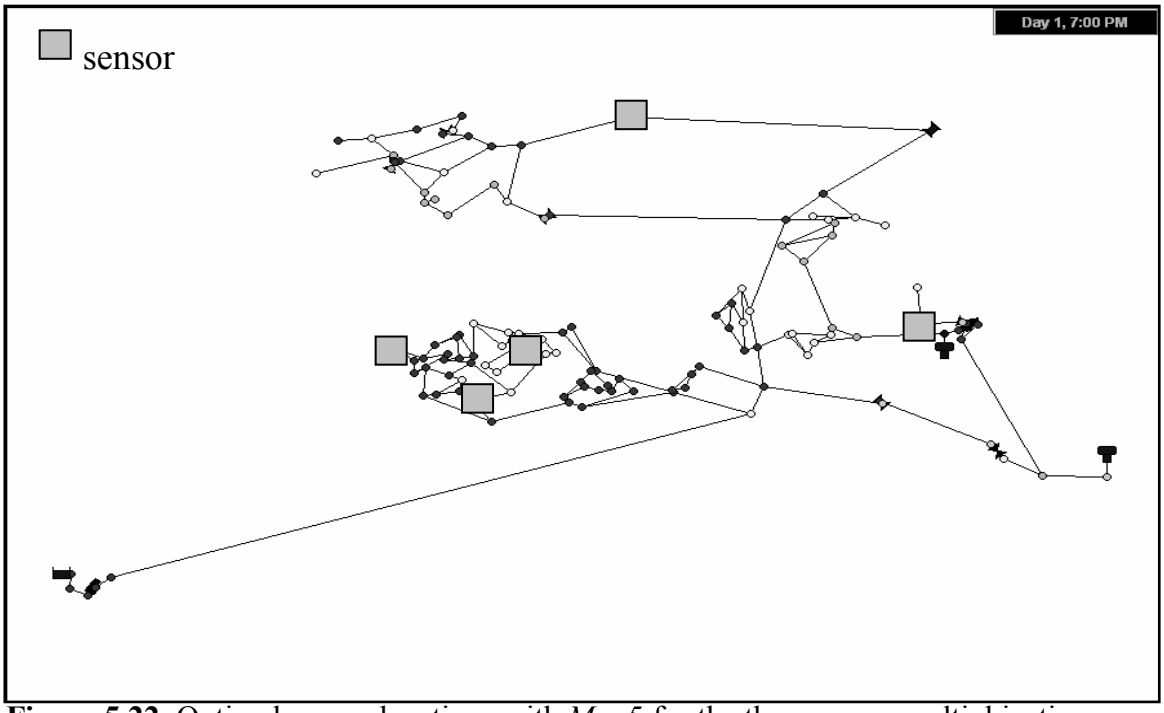


Figure 5.22 Optimal sensor locations with $M = 5$ for the three-prong multiobjective problem in the case of 2 nodes injected at 2 times for BWSN Network 1.

Table 5.25 Comparison of optimal sensor locations for the three-prong multiobjective problem with $M = 5$ in the variant attack cases for BWSN Network 1 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{all}**	Z_{lik}** (%)	Z_{time}** (min)	Z_{vol}** (gal)
contaminant injection duration = 10 h	This Study	11, 45, 83, 100, 118	2.72	88.4	1408.9	22,848
	Dorini et al. (2006)	10, 45, 49, 83, 118	2.40	86.0	1647.1	16,580
	Eliades and Polycarpou (2006)	17, 31, 45, 83, 126	2.19	82.7	1704.3	15,541
	Huang et al. (2006)	46, 70, 83, 100, 118	2.55	84.1	1449.7	16,843
	Krause et al. (2006)	17, 46, 83, 101, 126	2.22	83.5	1539.8	16,611
	Propato and Piller (2006)	17, 31, 83, 100, 120	2.46	81.8	1639.5	16,118
	Wu and Walski (2006)	21, 45, 83, 118, 123	2.42	84.4	1582.2	13,038
response delay = 3 h	This Study	11, 45, 83, 100, 118	2.61	88.3	1436.8	37,833
	Dorini et al. (2006)	11, 45, 49, 83, 118	2.59	85.7	1556.8	19,670
	Eliades and Polycarpou (2006)	17, 45, 68, 83, 126	2.28	82.6	1603.2	28,980
	Huang et al. (2006)	46, 83, 90, 100, 118	2.36	83.5	1558.6	30,178
	Krause et al. (2006)	49, 83, 101, 118, 122	2.44	81.9	1592.3	15,430
	Propato and Piller (2006)	49, 83, 101, 118, 123	2.49	82.7	1562.8	16,914
	Wu and Walski (2006)	17, 45, 83, 101, 126	2.33	84.0	1561.9	33,321
2 injection nodes at 1 time	This Study	45, 68, 83, 118, 122	2.75	97.2	589.8	9,288
	Dorini et al. (2006)	21, 68, 83, 117, 122	2.58	96.0	628.0	5,953
	Eliades and Polycarpou (2006)	10, 17, 45, 83, 126	2.37	98.3	729.7	19,013
	Huang et al. (2006)	46, 68, 83, 101, 118	2.71	96.6	576.3	10,091
	Krause et al. (2006)	17, 31, 45, 83, 122	2.45	96.6	694.2	7,323
	Propato and Piller (2006)	17, 21, 68, 83, 122	2.47	95.5	650.8	5,958
	Wu and Walski (2006)	68, 83, 101, 118, 122	2.73	96.8	579.0	8,990

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

5.5.2 ANALYSIS OF PERFORMANCE TRADEOFFS

In this section, the compromises in protection according to single-objective protection goals made in selecting certain sensor nodes for the three-prong multiobjective problem are quantified and evaluated. For this analysis, solutions with 5 sensors employed are examined as the tradeoffs are more apparent with fewer sensors in this small system. Table 5.26 presents the Pareto front space boundary values for each of the three performance measures, performance measure values corresponding to the multiobjective solution, and the cost of protection relative to each single-objective protection goal incurred by employing the optimal multiobjective sensor nodes with 5 sensors employed for the default attack case. Since the multiobjective sensor nodes match those for minimizing Z_{time}, the protection performance cost with respect to Z_{time} is 0.0% in Table 5.26. The protection cost in terms of Z_{lik} is only 6.1% and in terms of Z_{vol} is 22.0%. Performance costs for each protection goal performance measure are found in the same manner for the variant attack cases; these costs are provided in Table 5.27. Considering that no cost corresponding to any performance measure for any attack case exceeds 25%, the multiobjective optimization approach developed in this work appears to work well in preserving protection with respect to individual protection goals for BWSN Network 1.

Table 5.26 Performance tradeoffs for individual protection goals resulting from maximizing Z_{all} with $M = 5$ in the default attack case for BWSN Network 1.

Protection Goal Performance Measure	Pareto Front Space Minimum Value	Pareto Front Space Maximum Value	Value for Maximizing Z_{all}	Protection Performance Cost
Z_{lik}	66.9%	89.7%*	88.3%	6.1%
Z_{time}	1436.8 min*	2256.9 min	1436.8 min	0.0%
Z_{vol}	3863.7 gal*	59,701 gal	16,145 gal	22.0%

* Value represents maximum protection with respect to the particular single-objective protection goal in question.

Table 5.27 Performance tradeoffs for individual protection goals resulting from maximizing Z_{all} with $M = 5$ in the variant attack cases for BWSN Network 1.

Attack Case	Protection Performance Cost, Max Z_{lik}	Protection Performance Cost, Min Z_{time}	Protection Performance Cost, Min Z_{vol}
default	6.1%	0.0%	22.0%
sensor detection resolution = 0.3 mg/L	5.0%	0.0%	23.3%
contaminant injection duration = 10 h	6.9%	0.0%	21.5%
contaminant hazard threshold = 0.0 mg/L	6.4%	0.0%	22.6%
response delay = 3 h	6.4%	0.0%	32.2%
2 injection nodes at 1 time	16.1%	3.4%	6.0%
2 injection nodes at 2 times	24.1%	0.0%	12.7%

5.5.3 EVALUATION OF NODAL IMPORTANCE

To find the best multiobjective tradeoff solution for a given attack case, the nodal importance values and rankings for the three single objective problems are used in the Pareto dominance-based procedure outlined in Section 4.5.2.5 to determine multiobjective nodal importance rankings for a particular attack case. Figure 5.23 gives some indication of how this procedure is carried out for BWSN Network 1 in the default

attack case. Once multiobjective nodal importance rankings were determined, ISSM was carried out in the same manner as was done for the single-objective problems.

Tables 5.28 and 5.29 give the ranks of the lowest ranked nodes of optimal solutions for the three-prong multiobjective problem for the various attack cases. Upon first inspection of the data reported in the tables, it would seem that the multiobjective nodal importance approach is not effective in narrowing the search domain to a significant extent, even when 5 sensors are employed. However, when 5 sensors are employed in the one injection node attack cases, four of the five solution nodes are very highly ranked, while one solution node is ranked very low. This low-ranked node in every case is JUNCTION-11; it is very important in terms of maximizing Z_{lik} and minimizing Z_{time}, but it is too far downstream on its corresponding contaminant pathway to be important in terms of minimizing Z_{vol}. Therefore, it is lowly ranked for the multiobjective problem as defined in this study. Despite this optimal solution node being so lowly ranked, the multiobjective nodal importance approach does rank highly many nodes that are desirable for sensor placement even though they make for slightly suboptimal solutions. The solution convergence plots in Figures 5.24 and 5.25 illustrate this effectiveness for the cases with 5 sensors employed; the solution converges to good, though not optimal, solutions with ISSM when the subset size is in the neighborhood of roughly 30 to 60.

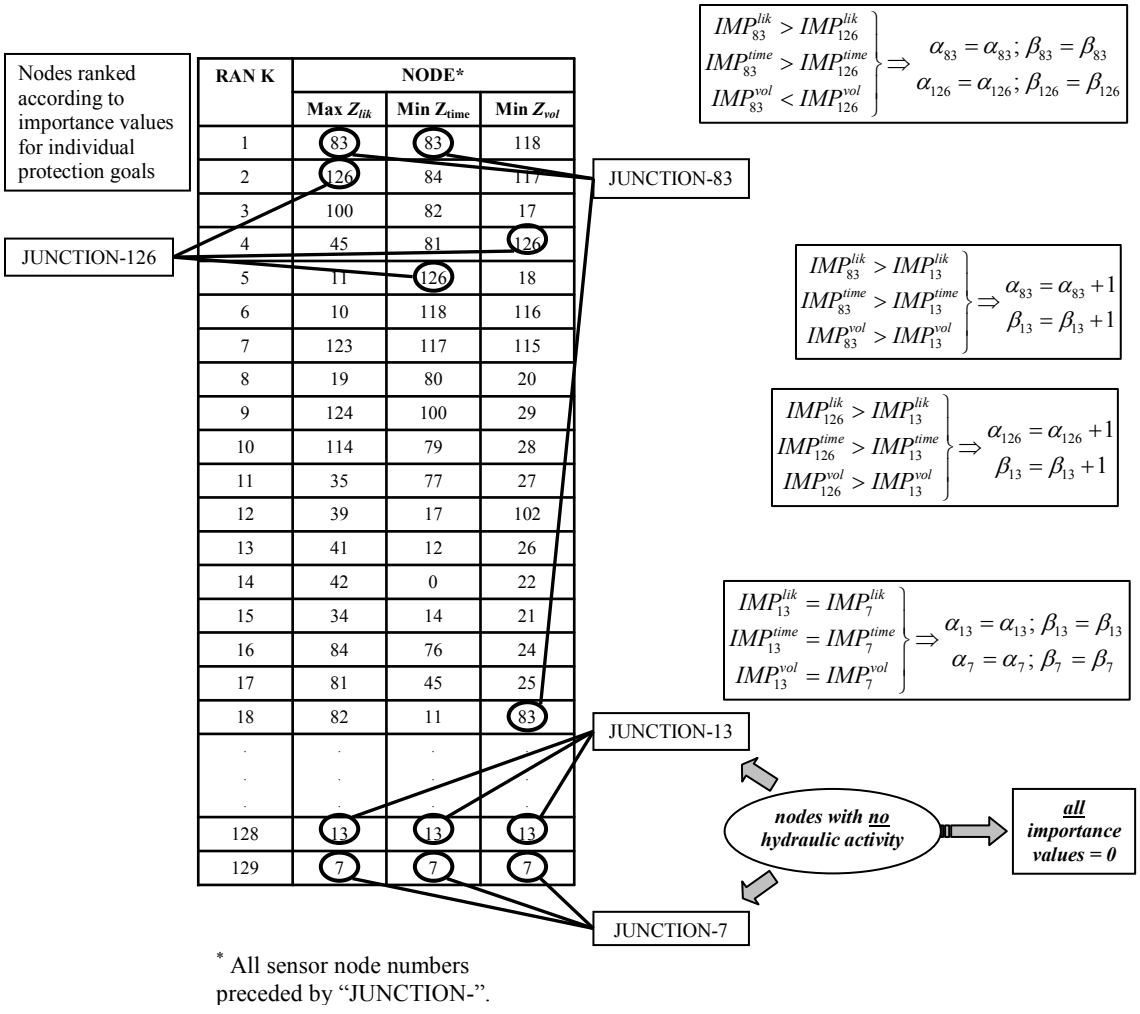
In the two injection node attack cases, the optimal sensor nodes are all relatively important in terms of minimizing Z_{vol} along with maximizing Z_{lik} and minimizing Z_{time}, so the lowest ranked nodes for those cases are much more highly ranked relative to the one

injection node cases. The protection costs given for the two injection node cases in Table 5.27 above support that the solution nodes in these cases are more evenly accommodating with respect to all three single-objective protection goals. As expected, given the more centralized nature of the sensor locations, the costs are lower with regard to minimizing Z_{vol} and higher with regard to maximizing Z_{lik}.

In the default attack case with 20 sensors employed, the drawback of a high M/N ratio is apparent as very low-ranked nodes are required to make up the entire set of optimal sensor nodes. These low-ranked nodes are important only with respect to maximizing detection likelihood, so they naturally would be considered much less desirable by the multiobjective nodal importance approach.

The solution convergence plot in Figure 5.24 reveals another benefit of the random inclusion by ISSM of unranked nodes in subsets. Despite the optimal solution containing low-ranked nodes for both $M = 5$ and $M = 20$, the random inclusion of these nodes in subsets allows the optimal solution to be found in these cases with subset sizes of 72 and 90, respectively. Thus, as long as the multiobjective nodal importance approach ranks highly most of the desirable nodes for the three-prong multiobjective problem, ISSM can find the optimal solution or at least a very near-optimal solution with a subset significantly smaller than the set of all WDS nodes. (It is noted that, for the solution convergence plots corresponding to the multiobjective problem, the convergence curves for particular attack cases and numbers of sensors employed should be evaluated

independently of each other since Z_{all} values for the cases are based on separate sets of boundary values. The curves are plotted together for the sake of concise illustration.)



RANK	NODE*		
	Max Z_{lik}	Min Z_{time}	Min Z_{vol}
1	83	83	118
2	126	84	117
3	100	82	17
4	45	81	126
5	11	126	18
6	10	118	116
7	123	117	115
8	19	80	20
9	124	100	29
10	114	79	28
11	35	77	27
12	39	17	102
13	41	12	26
14	42	0	22
15	34	14	21
16	84	76	24
17	81	45	25
18	82	11	83
.	.	.	.
.	.	.	.
.	.	.	.
128	13	13	13
129	7	7	7

$$\left.\begin{array}{l} IMP_{83}^{lik} > IMP_{126}^{lik} \\ IMP_{83}^{time} > IMP_{126}^{time} \\ IMP_{83}^{vol} < IMP_{126}^{vol} \end{array}\right\} \Rightarrow \begin{array}{c} \alpha_{83} = \alpha_{83};\ \beta_{83} = \beta_{83} \\ \alpha_{126} = \alpha_{126};\ \beta_{126} = \beta_{126} \end{array}$$

$$\left.\begin{array}{l} IMP_{83}^{lik} > IMP_{13}^{lik} \\ IMP_{83}^{time} > IMP_{13}^{time} \\ IMP_{83}^{vol} > IMP_{13}^{vol} \end{array}\right\} \Rightarrow \begin{array}{c} \alpha_{83} = \alpha_{83} + 1 \\ \beta_{13} = \beta_{13} + 1 \end{array}$$

$$\left.\begin{array}{l} IMP_{126}^{lik} > IMP_{13}^{lik} \\ IMP_{126}^{time} > IMP_{13}^{time} \\ IMP_{126}^{vol} > IMP_{13}^{vol} \end{array}\right\} \Rightarrow \begin{array}{c} \alpha_{126} = \alpha_{126} + 1 \\ \beta_{13} = \beta_{13} + 1 \end{array}$$

$$\left.\begin{array}{l} IMP_{13}^{lik} = IMP_{7}^{lik} \\ IMP_{13}^{time} = IMP_{7}^{time} \\ IMP_{13}^{vol} = IMP_{7}^{vol} \end{array}\right\} \Rightarrow \begin{array}{c} \alpha_{13} = \alpha_{13};\ \beta_{13} = \beta_{13} \\ \alpha_{7} = \alpha_{7};\ \beta_{7} = \beta_{7} \end{array}$$

Figure 5.23 Determining nodal importance rankings for the three-prong multiobjective problem in the default attack case for BWSN Network 1.

Table 5.28 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for the three-prong multiobjective problem for the default attack case and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	98	76.0%
20	121	93.8%

Table 5.29 Nodal importance ranks for the lowest-ranked BWSN Network 1 nodes of optimal solutions for the three-prong multiobjective problem with $M = 5$ for the variant attack cases and the proportions of ranked BWSN Network 1 nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	98	76.0%
sensor detection resolution = 0.3 mg/L	92	71.3%
contaminant injection duration = 10 h	99	76.7%
contaminant hazard threshold = 0.0 mg/L	96	74.4%
response delay = 3 h	98	76.0%
2 injection nodes at 1 time	44	34.1%
2 injection nodes at 2 times	60	46.5%

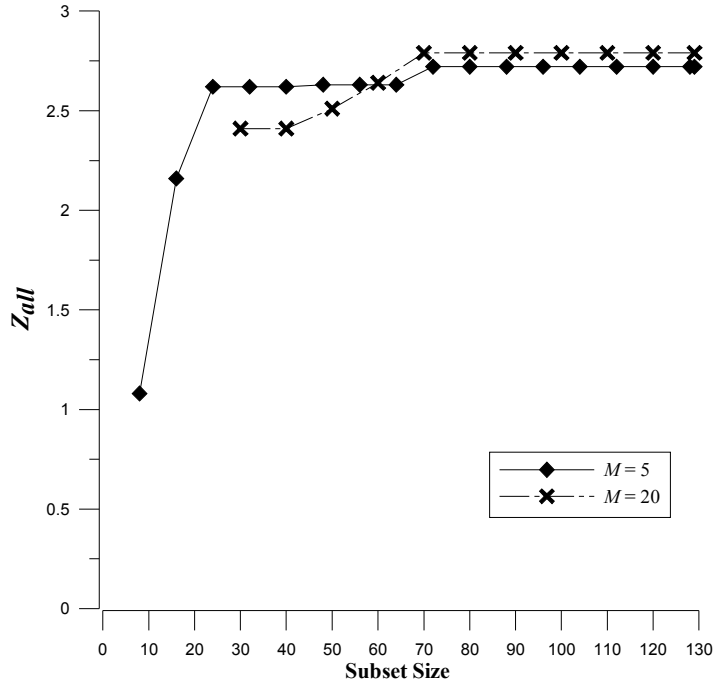


Figure 5.24 Solution convergence as ISSM is carried out for the three-prong multiobjective problem in the default attack case for BWSN Network 1. (Values of Z_{all} for solutions existing outside of the Pareto front space are not shown.)

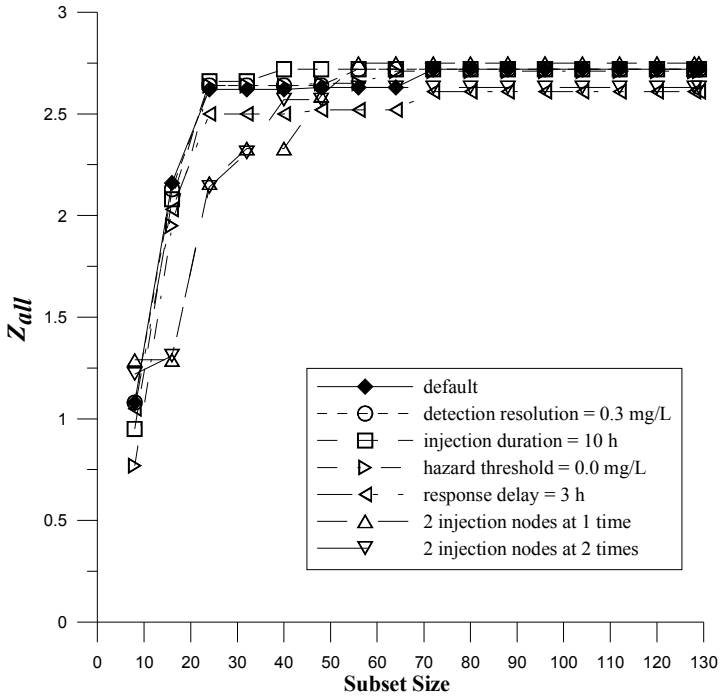


Figure 5.25 Solution convergence as ISSM is carried out for the three-prong multiobjective problem with $M = 5$ in the default and variant attack cases for BWSN Network 1.

5.6 COMPUTATIONAL EXPENSE NOTES

For this system, attack scenario generation required approximately 1 hour of runtime on a Dell Precision 690 with an Intel Xeon 5160 processor (2.99 GHz, 16 GB RAM) and created about 40 MB of data for use in the importance and optimization phases. For a given multiobjective problem run, performing nodal importance calculations for the three single-objective problems and the multiobjective problem together required less than 10 minutes. To solve all three single-objective problems and then the multiobjective problem, approximately 45 minutes of runtime were required. The computational costs for this work appear reasonable for this system, especially when compared to the expense incurred in other works employing more computationally demanding optimization methods such as that of Krause et al. (2006) which required over 1 GB of memory in some cases and multiple days of runtime with a comparable computational platform (3 GHz, 20 GB RAM). However, the GA model and library used in this work required strict convergence criteria and a relatively large population size of 100 to yield the results presented above in the times reported here. Runtimes may have even been more desirable with another GA model or library or even with another type of optimization solver.

CHAPTER 6

RESULTS AND ANALYSIS: BWSN NETWORK 2

In this chapter, sensor placement optimization results are presented and analyzed for BWSN Network 2 in the same manner as results were presented and analyzed in Chapter 5 for BWSN Network 1. The results are analyzed with respect to BWSN Network 2 alone and are compared in a general sense to optimization results for BWSN Network 1 in order to characterize similarities and differences in placement trends between results for systems of two very different sizes and configurations.

6.1 SINGLE-OBJECTIVE PROBLEM: MAXIMIZING Z_{lik}

Sensor placement results regarding the maximizing of detection likelihood for BWSN Network 2 are given and discussed in this section.

6.1.1 SENSOR NODE SOLUTIONS

Sensor placement results for maximizing detection likelihood in the default attack case for BWSN Network 2 are provided by Table 6.1. As expected, all sensor nodes in the default case are end-of-line nodes given the preference placed on end-of-line nodes by the respective nodal importance approach for this single-objective problem. There is more discussion below regarding the validated importance of end-of-line nodes in maximizing detection likelihood for this large system.

Z_{lik} values are much lower than their counterparts for BWSN Network 1. These differences stand to reason, though, as BWSN Network 2 is larger than BWSN Network 1 by orders of magnitude, and it cannot necessarily be expected that 5 or 20 sensors can cover the majority of the system and detect contamination for a majority of attack scenarios. The locations of selected sensor nodes with $M = 5$ and $M = 20$ are depicted in Figures 6.1 and 6.2, respectively. It is noted that the two nodes that appear close to each other in Figure 6.1 are actually rather hydraulically separated from each other and for the most part are not covering the same contaminant pathways. Figure 6.2 illustrates well the coverage of contaminant pathways provided by the sensor nodes selected; many of these nodes are located on the periphery of the system. In general, water in the system flows downstream on hydraulic pathways toward endpoints in the outer regions of the system, as in the case of BWSN Network 1. Other sensor nodes are noted to cover local neighborhoods of nodes in the system that are somewhat isolated from the major hydraulic pathways.

Table 6.1 Optimal sensor locations and performance measures for maximizing detection likelihood in the default attack case for BWSN Network 2.

M	**Sensor Nodes***	Z_{lik}** **(%)**	Z_{time} **(min)**	Z_{vol} **(gal)**
5	623, 1421, 1486, 2865, 3237	31.9	4,595.2	178,561
20	623, 877, 902, 1434, 1486, 2598, 2865, 3237, 3669, 3723, 3833, 4208, 4306, 5346, 7436, 7662, 8089, 8419, 9217, 11687	43.5	4,015.0	135,321

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

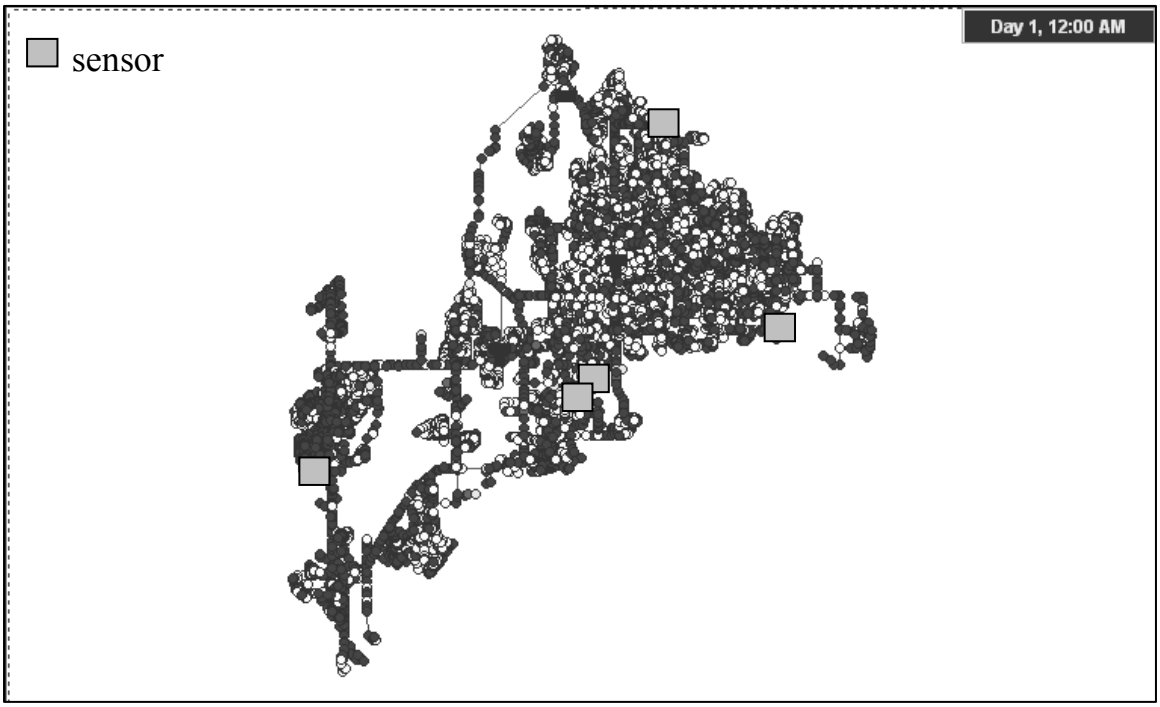


Figure 6.1 Optimal sensor locations with $M = 5$ for maximizing detection likelihood in the default attack case for BWSN Network 2.

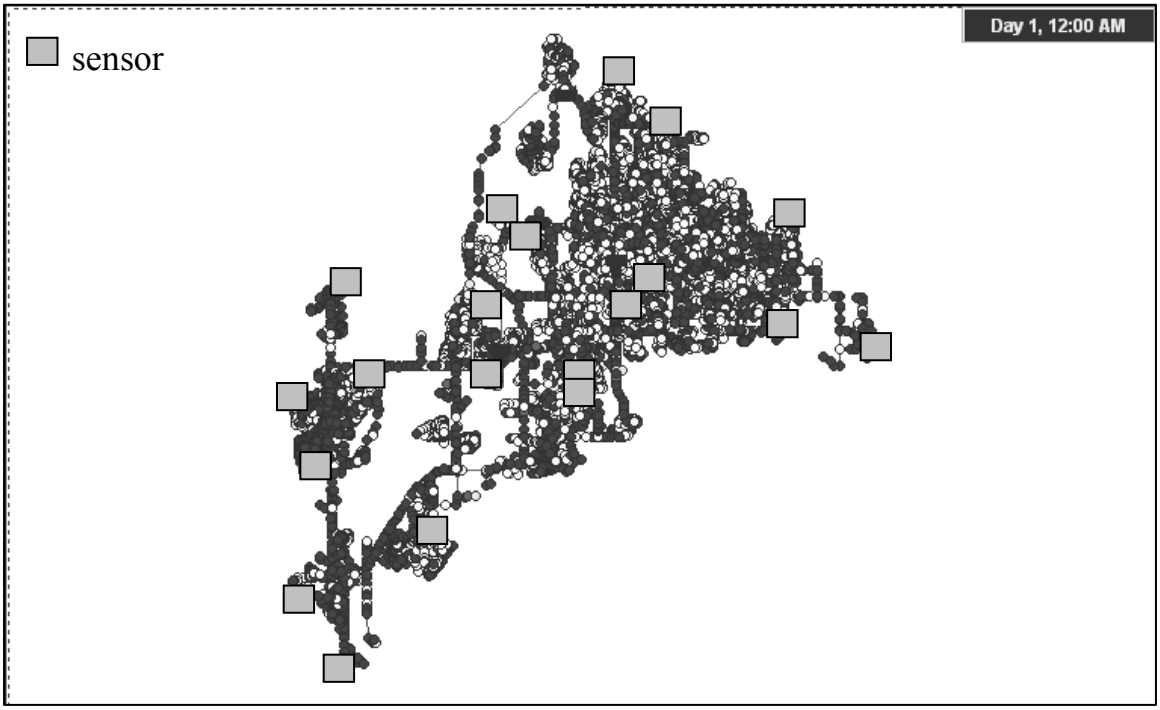


Figure 6.2 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the default attack case for BWSN Network 2.

Table 6.2 compares the sensor nodes chosen in this work for maximizing detection likelihood in the default attack case with those of other works. Z_{lik} values for this work are marginally better than those for the other works listed in the table, implying that the results are at least comparable if not better than those submitted in the other works.

Table 6.2 Comparison of optimal sensor locations for maximizing detection likelihood in the default attack case for BWSN Network 2 determined in this study with locations found in other works.

M	**Work**	**Sensor Nodes***	***Z*_lik**** **(%)**
5	This Study	623, 1421, 1486, 2865, 3237	31.9
	Berry et al. (2006c)	551, 1426, 1486, 2865, 3230	31.8
	Eliades and Polycarpou (2006)	610, 1486, 2865, 3678, 4359	31.7
	Krause et al. (2006)	1486, 3229, 3769, 3836, 10874	29.2
20	This Study	623, 877, 902, 1434, 1486, 2598, 2865, 3237, 3669, 3723, 3833, 4208, 4306, 5346, 7436, 7662, 8089, 8419, 9217, 11687	43.5
	Berry et al. (2006c)	623, 813, 902, 1436, 1486, 2598, 2865, 3230, 3669, 3723, 3833, 4208, 4306, 5346, 7436, 7665, 8089, 8419, 9223, 11687	43.4
	Eliades and Polycarpou (2006)	532, 579, 1426, 1486, 2865, 3231, 3679, 3836, 4234, 4359, 4511, 4609, 5087, 5585, 6922, 7440, 7670, 7858, 9044, 9787	40.8
	Krause et al. (2006)	542, 813, 1430, 1477, 1486, 1718, 2602, 3229, 3635, 3769, 3836, 4208, 4306, 5185, 5346, 7441, 7668, 9364, 10874, 11687	42.8

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

Table 6.3 provides the results of sensor placement optimization with 20 sensors employed for the variant attack cases applicable for maximizing detection likelihood. Sensor node locations for these attack cases are illustrated in Figures 6.3 through 6.6. With 20 sensors employed, true shifts in node location due to differences in attack conditions are apparent if they indeed occur. For this protection goal, there are subtle

differences in sensor placement between attack cases, but the areas covered by the sensor nodes are largely the same from attack case to attack case.

Of course, for the attack case of a 0.3 mg/L sensor detection resolution, end-of-line nodes are not assumed to account for entire contaminant pathways. Even though the optimization approach developed in this work did not put a preference on end-of-line nodes in the case of a non-zero detection resolution, it is noted that all nodes selected for this attack case in Table 6.3 are end-of-line nodes, likely due to the high uniqueness values of these nodes. Another interesting observation regarding the placement of sensors for the case of the 0.3 mg/L detection resolution is the need for an additional sensor in the top region of the system relative to the default attack case placement; this additional sensor is needed at that location to detect contaminant before the contaminant concentration falls below 0.3 mg/L.

Regarding the Z_{lik} values resulting from different attack cases, the pattern seen in studying maximization of detection likelihood for BWSN Network 1 is seen for this system as well. The Z_{lik} value for the case of the increased detection resolution is less than the value for the default attack case, the Z_{lik} value for increased injection duration is slightly higher, and Z_{lik} values for the two injection node cases are substantially higher.

Table 6.4 compares the results of this study for the variant attack cases with those of works with results available for comparison. As in the default attack case, this work provides nodes of equal or better protection performance with respect to other works.

Table 6.3 Optimal sensor locations and performance measures for maximizing detection likelihood with $M = 20$ in the variant attack cases for BWSN Network 2.

Attack Case	Sensor Nodes*	Z_{lik}** (%)	Z_{time} (min)	Z_{vol} (gal)
default	623, 877, 902, 1434, 1486, 2598, 2865, 3237, 3669, 3723, 3833, 4208, 4306, 5346, 7436, 7662, 8089, 8419, 9217, 11687	43.5	4,015.0	135,321
sensor detection resolution = 0.3 mg/L	623, 808, 1486, 2423, 2598, 2865, 3237, 3669, 3723, 3833, 4208, 4306, 5346, 6545, 7436, 7664, 8089, 8409, 9217, 11687	42.7	4,065.5	137,099
contaminant injection duration = 10 h	623, 877, 902, 1421, 1443, 1486, 2598, 2865, 3230, 3833, 4208, 4306, 5325, 7436, 7665, 8089, 8419, 9223, 9972, 11687	45.8	3,973.7	250,159
2 injection nodes at 1 time	623, 877, 902, 1436, 1486, 2598, 2865, 3237, 3676, 3723, 3833, 4208, 4306, 5346, 7436, 7662, 8089, 8419, 9217, 11687	67.3	2,921.0	229,107
2 injection nodes at 2 times	623, 865, 902, 1421, 1486, 2598, 2865, 3237, 3669, 3723, 3833, 4208, 4306, 5356, 7436, 7662, 8089, 8419, 9217, 11687	65.1	3,166.4	221,780

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

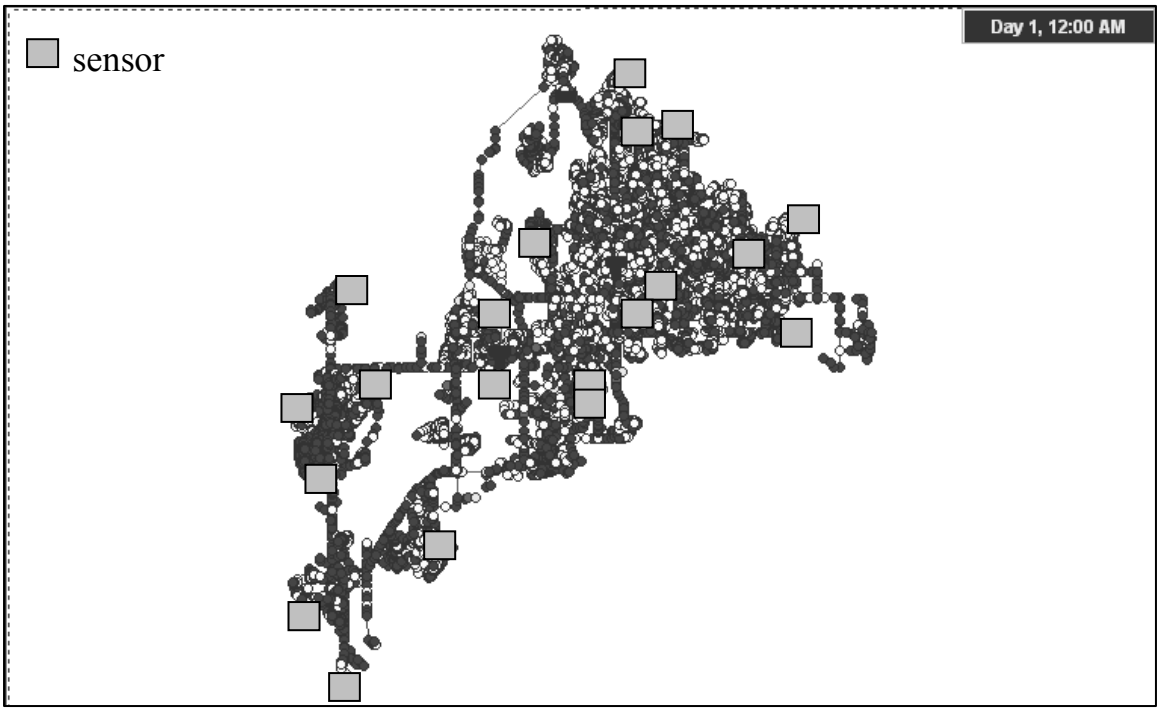


Figure 6.3 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of a detection resolution of 0.3 mg/L for BWSN Network 2.

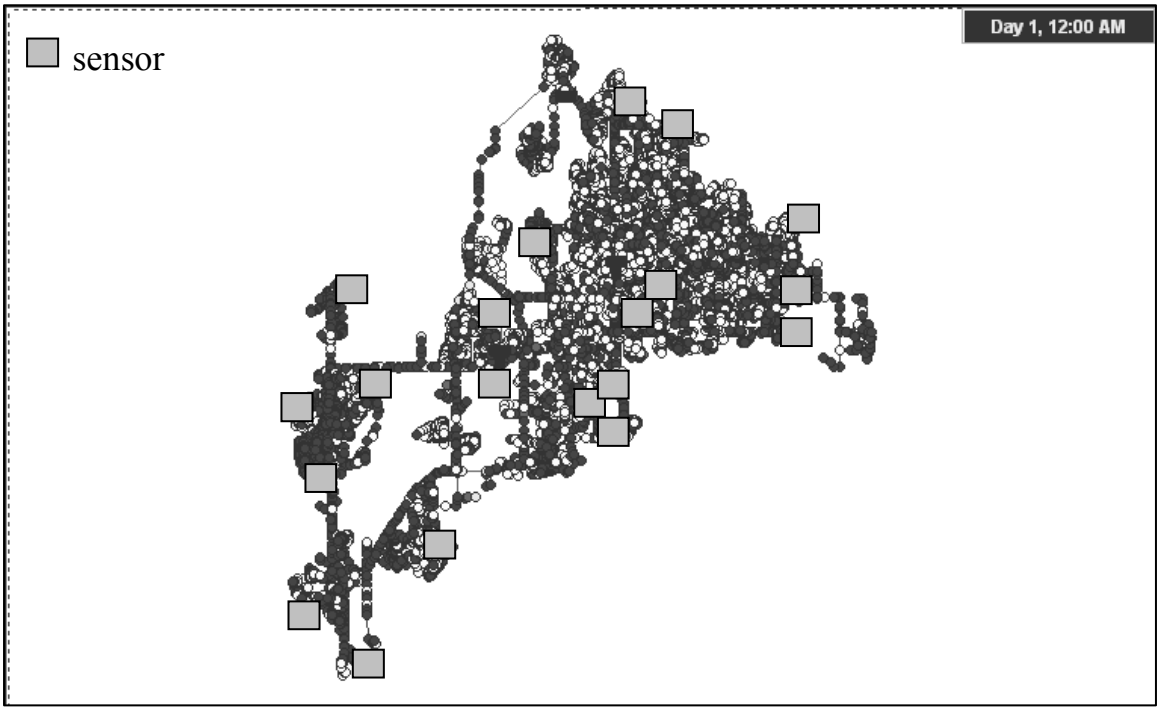


Figure 6.4 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of an injection duration of 10 h for BWSN Network 2.

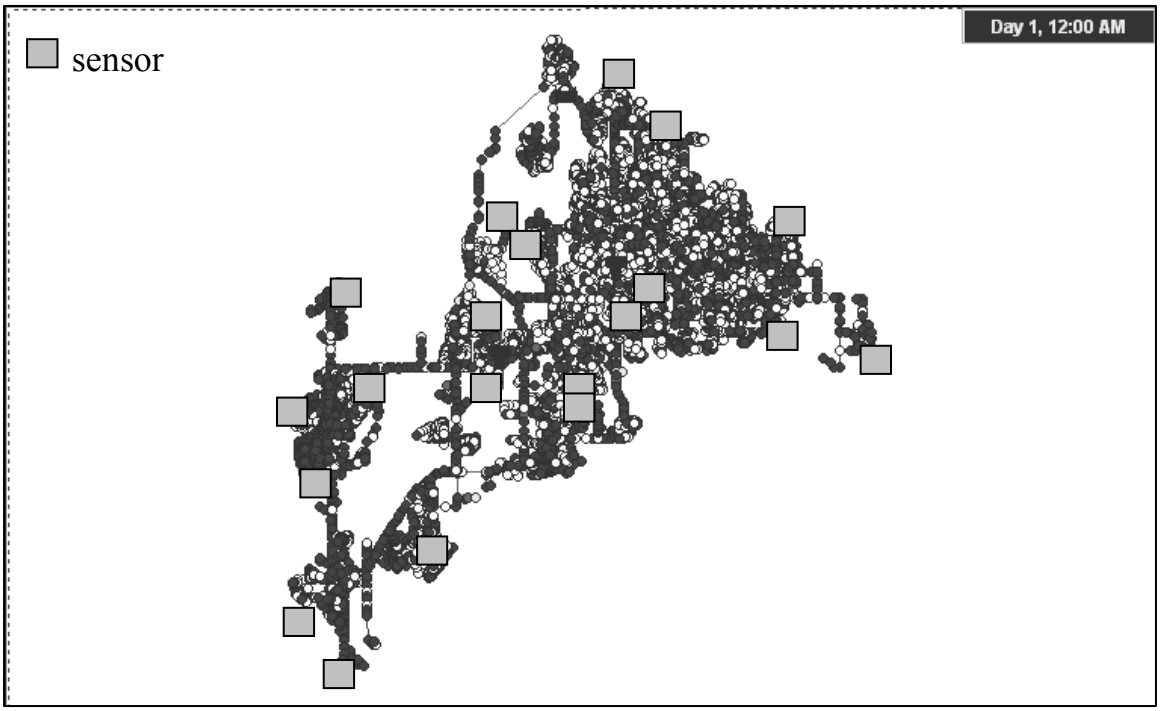


Figure 6.5 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of 2 nodes injected at 1 time for BWSN Network 2.

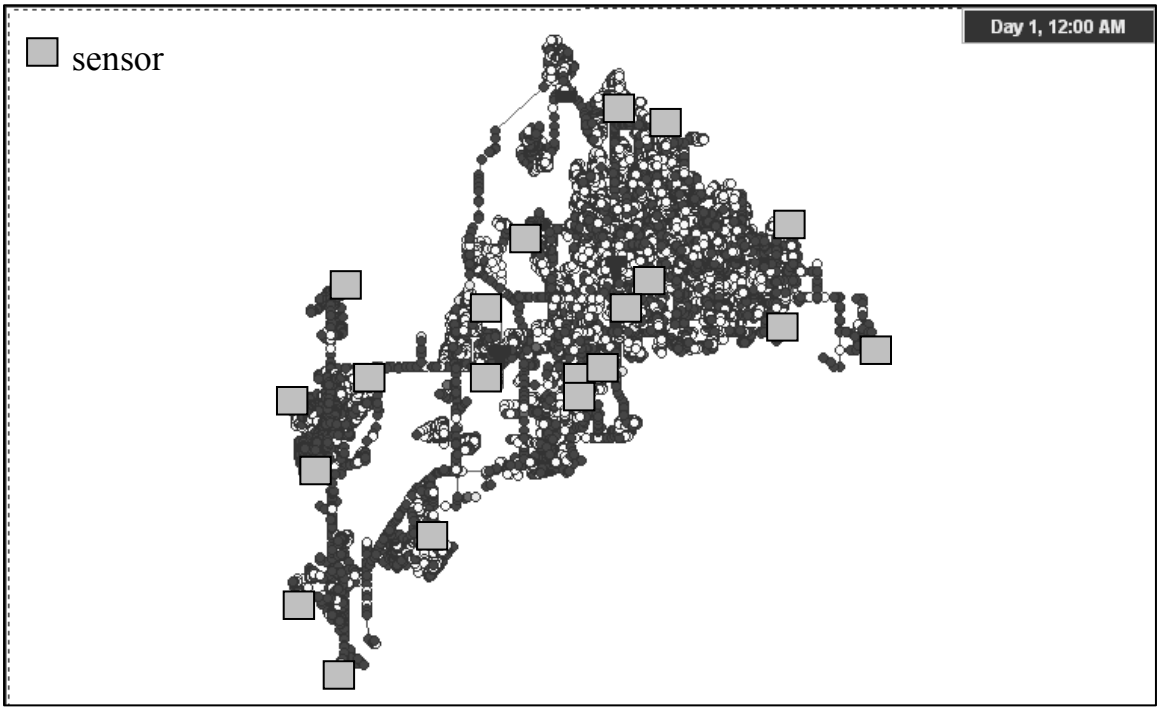


Figure 6.6 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of 2 nodes injected at 2 times for BWSN Network 2.

Table 6.4 Comparison of optimal sensor locations for maximizing detection likelihood with $M = 20$ in the variant attack cases for BWSN Network 2 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{lik}** (%)
contaminant injection duration = 10 h	This Study	623, 877, 902, 1421, 1443, 1486, 2598, 2865, 3230, 3833, 4208, 4306, 5325, 7436, 7665, 8089, 8419, 9223, 9972, 11687	45.8
	Berry et al. (2006c)	623, 813, 902, 1426, 1486, 2598, 2865, 3230, 3688, 3833, 4208, 4306, 6922, 7436, 7665, 8089, 8419, 9223, 11687, 12371	45.8
	Krause et al. (2006)	551, 813, 1436, 1486, 2598, 3231, 3669, 3723, 3833, 4208, 4306, 5326, 7436, 7665, 8089, 8419, 9223, 9364, 11687, 12371	45.4
2 injection nodes at 1 time	This Study	623, 877, 902, 1436, 1486, 2598, 2865, 3237, 3676, 3723, 3833, 4208, 4306, 5346, 7436, 7662, 8089, 8419, 9217, 11687	67.3
	Krause et al. (2006)	551, 813, 925, 1436, 1486, 2598, 3231, 3669, 3833, 4208, 4306, 5346, 7436, 7665, 8089, 8419, 9223, 9364, 11687, 12371	67.1

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

6.1.2 EVALUATION OF NODAL IMPORTANCE

Solutions given above were found by carrying out ISSM employing nodes ranked first by end-of-line node status then by values calculated with the large-system nodal importance function for maximizing detection likelihood in (4.21) for all attack cases assuming a zero sensor detection resolution. The ranking approach for this system is in contrast to the one used for BWSN Network 1 that made use of the nodal importance function in (4.22) and the sorting approach that includes both end-of-line and slow-outflow nodes.

The 2,140 end-of-line nodes in BWSN Network 2 represented an ample number of contaminant pathways that could be covered to maximize detection likelihood as validated by the results above, so the neglecting of slow-outflow nodes from the sorting approach for large systems seems to be an acceptable decision. In the case of the non-zero detection resolution, nodes were ranked by the value determined with (4.21) only.

The ISSM parameters used for optimization for this larger system were different than those for BWSN Network 1 due to the larger system size; these parameters are provided in Table 6.5. There is increased emphasis on ranked nodes in the ISSM subset for the larger system, which is affordable due to the larger size of the subset. The 10% of unranked nodes in the subset provides a more than adequate number of nodes to allow for unranked nodes to be included in best solutions for subsets if the potential exists.

Table 6.5 Parameters selected for use in ISSM for BWSN Network 2 for all protection goals and all attack cases.

M	**Initial Subset Size**	**Change in Subset Size**	**Proportion of Ranked Nodes in Subset**	**Proportion of Unranked Nodes in Subset**
5	300	300	90%	10%
20	300	300	90%	10%

As indicated by the ranks of the lowest-ranked nodes of optimal solutions in Tables 6.6 and 6.7, nodal importance allows for the optimal solutions for maximizing detection likelihood in this system to be isolated to very small fractions of WDS nodes. To supplement the data in Table 6.6 for the case of 5 sensors employed, it is noted that 4 of the 5 nodes in the optimal solution are ranked in the top 100, even though the fifth node

is ranked 518, further demonstrating the effectiveness of nodal importance in narrowing the search domain.

Even in the case of the 0.3 mg/L sensor detection resolution, the optimal solution is contained in proportions nearly equal in size to those of the other attack cases that were able to put preference on end-of-line nodes. Again, the solution nodes for this case are all end-of-line nodes as well. Thus, it seems that stratifying end-of-line nodes from the other WDS nodes in this case--even though the underlying assumption is not valid--may allow for the narrowing of the search domain to find a near-optimal—or even the optimal—solution.

The sensor nodes selected in this work are remarkably very close in location to those selected by Berry et al. (2006c), and many of the nodes selected by Berry et al. are end-of-line nodes. Berry et al. did not use such concepts as end-of-line status and uniqueness developed in this study to determine their "nodal impact coefficients" to solve this single-objective problem. The parallels between this study and their study validate somewhat the concepts used in this study to determine the desirability of a node as a sensor node candidate for maximizing detection likelihood, but in this study the desirability was quantified without an iterative, computationally-expensive process like that used by Berry et al.

In addition, it is argued that the nodal importance approach used in this study was key in finding a solution that is marginally better than that of Berry et al. in the default attack

case with 5 sensors employed. JUNCTION-623, submitted as one of the optimal sensor nodes in this work, is the endpoint of the hydraulic path that contains JUNCTION-551, an upstream junction submitted by Berry et al. as an optimal sensor node. In scenario data generation, an injection node was selected downstream of JUNCTION-551 for two optimization scenarios; JUNCTION-551 could not detect contamination for these scenarios given its location, whereas JUNCTION-623 could as it is located at the end of the contaminant transport path. Thus, it seems that putting preference on end-of-line nodes in general ensures maximum coverage, even though in this case the difference between placing the sensor at an end-of-line versus a node that is not an end-of-line node is only marginal.

Berry et al. claimed that they could not perform optimization in the two injection node cases due to computational expense associated with the determining of nodal impact coefficients, but in this study there was no problem in finding solutions using ISSM with nodal importance, and the solution for the case of 2 nodes injected at 1 time seems to marginally outperform the solution submitted by Krause et al. (2006). The feasibility of performing optimization in these two injection node cases adds to the value demonstrated in the nodal importance approach for maximizing detection likelihood.

Solution convergence plots for this single-objective problem are given in Figures 6.7 and 6.8. All convergence curves imply that convergence occurs with only a few ISSM iterations for all attack cases, and even ISSM subsets much smaller than those that would contain all optimal sensor nodes could provide near-optimal solutions.

Table 6.6 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for maximizing detection likelihood in the default attack case and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	518	4.1%
20	808	6.5%

Table 6.7 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for maximizing detection likelihood with $M = 20$ in the variant attack cases and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	808	6.5%
sensor detection resolution = 0.3 mg/L	788	6.3%
contaminant injection duration = 10 h	870	6.9%
2 injection nodes at 1 time	811	6.5%
2 injection nodes at 2 times	813	6.5%

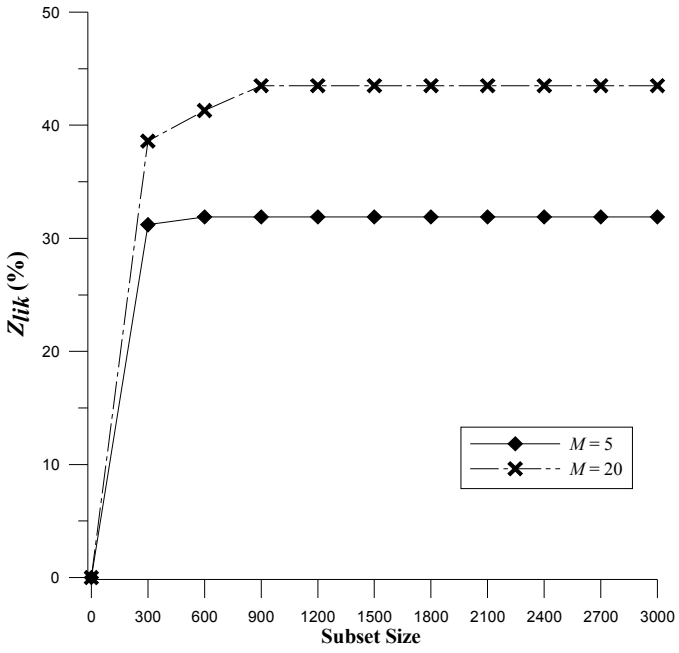


Figure 6.7 Solution convergence as ISSM is carried out for maximizing detection likelihood in the default attack case for BWSN Network 2. (Only solution values for subset sizes of 3,000 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 3,000.)

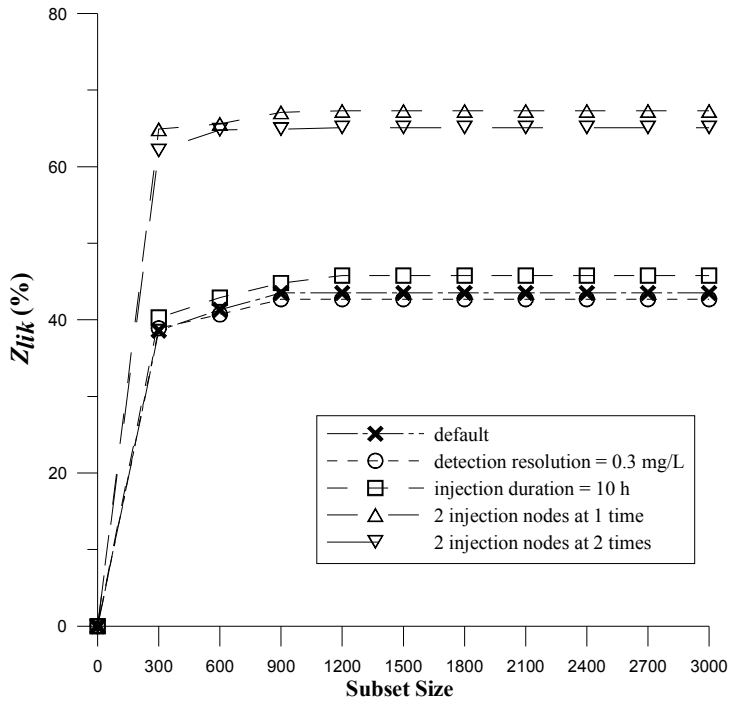


Figure 6.8 Solution convergence as ISSM is carried out for maximizing detection likelihood with $M = 20$ in the default and variant attack cases for BWSN Network 2. (Only solution values for subset sizes of 3,000 or less are plotted as solutions for all cases reached best performance by subset size = 3,000.)

6.2 SINGLE-OBJECTIVE PROBLEM: MINIMIZING Z_{time}

The optimization results for minimizing expected detection time for BWSN Network 2 are given and discussed below.

6.2.1 SENSOR NODE SOLUTIONS

Table 6.8 provides sensor placement results for minimizing expected detection time for BWSN Network 2 in the default attack case. It is observed that the selected sensor nodes provide for a drastic reduction in Z_{time} value, especially when 20 sensors are employed, when compared to the Z_{time} values in Table 6.1 above. Z_{vol} values are also reduced significantly versus the Z_{vol} values in Table 6.1, which is expected as sensors are placed more upstream for minimizing expected detection time relative to sensors placed for maximizing detection likelihood. Z_{lik} values in Table 6.8 are quite close to those in Table 6.1 due to the correlation between Z_{time} and Z_{lik}, especially given that in a system with low maximum Z_{lik} values where Z_{time} reduction is very dependent on minimizing instances of "infinite" detection time resulting from non-detection. The locations of the sensor nodes selected are illustrated in Figures 6.9 and 6.10. Table 6.9 compares the results for minimizing expected detection time in the default attack case for this study with those of other studies. The results for this study are marginally better than those of Krause et al. (2006) and significantly better than those of Eliades and Polycarpou (2006).

For the case 5 sensors employed, some node locations are in the same locations as those for maximizing detection likelihood (e.g., JUNCTION-623 and JUNCTION-1486), while

other sensor nodes are placed upstream often at nodes where flows from different areas of the system intersect (e.g, JUNCTION-10874). The compromise between maximizing coverage and detecting contamination farther upstream is more evident in Figure 6.10 for the case of 20 sensors employed. As shown in that figure, some sensor nodes are located in the outer regions of the system to maximize coverage, while there are more sensor nodes located toward the center of the system relative to Figure 6.2, implying a shift upstream of some sensor nodes to detect contamination sooner. This shift was also seen in minimizing expected detection time for BWSN Network 1.

Table 6.8 Optimal sensor locations and performance measures for minimizing expected detection time in the default attack case for BWSN Network 2.

M	**Sensor Nodes***	Z_{lik} **(%)**	Z_{time}** **(min)**	Z_{vol} **(gal)**
5	623, 1486, 3229, 3770, 10874	30.3	4,366.8	99,528
20	551, 877, 1486, 1904, 1917, 2598, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10874, 12404	42.8	3,660.9	47,899

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

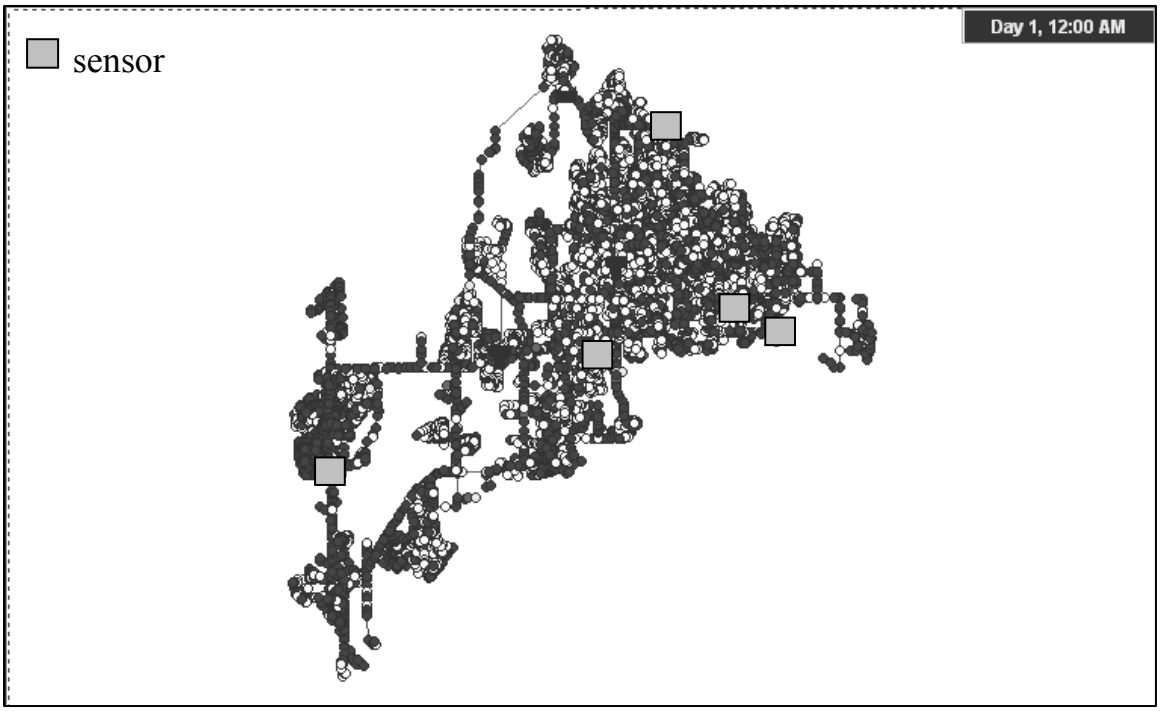


Figure 6.9 Optimal sensor locations with $M = 5$ for minimizing expected detection time in the default attack case for BWSN Network 2.

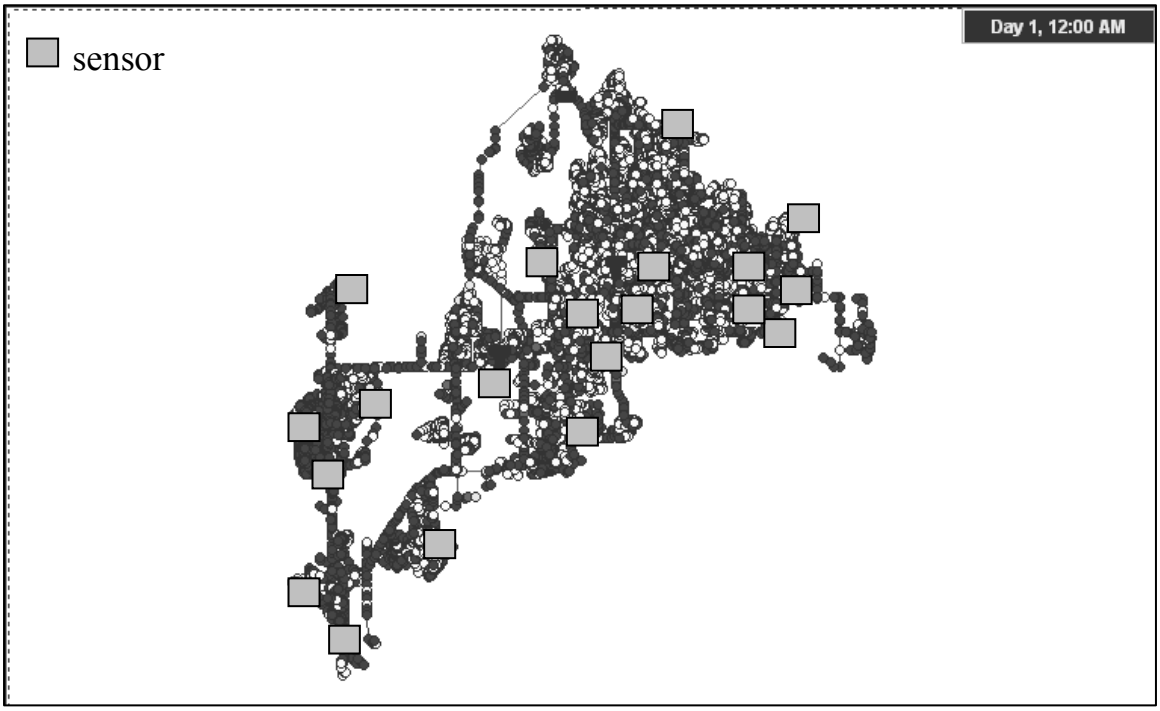


Figure 6.10 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the default attack case for BWSN Network 2.

Table 6.9 Comparison of optimal sensor locations for minimizing expected detection time in the default attack case for BWSN Network 2 determined in this study with locations found in other works.

M	**Work**	**Sensor Nodes***	Z_{time}** **(min)**
5	This Study	623, 1486, 3229, 3770, 10874	4,366.8
	Eliades and Polycarpou (2006)	532, 1486, 3836, 4359, 8445	4,446.8
	Krause et al. (2006)	1486, 3770, 3836, 7485, 10874	4,370.6
20	This Study	551, 877, 1486, 1904, 1917, 2598, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10874, 12404	3,660.9
	Eliades and Polycarpou (2006)	375, 532, 579, 1426, 1486, 3229, 3836, 4234, 4359, 4511, 4609, 5087, 5585, 6922, 7670, 7858, 8403, 8629, 9360, 9787	3,928.7
	Krause et al. (2006)	540, 877, 1486, 1904, 2602, 3229, 3635, 3770, 3836, 4247, 4306, 5346, 7441, 7485, 7667, 9210, 9364, 10874, 11167, 11687	3,665.0

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

Table 6.10 gives the sensor placement results minimizing expected detection time in the applicable variant attack cases, and Figures 6.11 through 6.14 indicate the locations of the selected sensor nodes in the system. Although some sensor nodes differ between the different attack cases, the areas where sensors are located do not vary greatly from case to case. Z_{time} values vary in expected manners from case to case. The decreased performance in the case of increased sensor detection resolution and increased performance in the cases of a longer injection duration and two injection nodes are primarily due to differences in network coverage ability in the particular cases. That is the reason why performance for this protection goal generally varies in the same way that detection likelihood varies among attack cases. Comparing the variant attack case results to the available results from Krause et al. (2006) for the cases of a 10-hour injection

duration and 2 nodes injected at 1 time in Table 6.11 shows that results from this work are comparable if not better than those already documented in literature.

Table 6.10 Optimal sensor locations and performance measures for minimizing expected detection time with $M = 20$ in the variant attack cases for BWSN Network 2.

Attack Case	Sensor Nodes*	Z_{lik} (%)	Z_{time}** (min)	Z_{vol} (gal)
default	551, 877, 1486, 1904, 1917, 2598, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10874, 12404	42.8	3,660.9	47,899
sensor detection resolution = 0.3 mg/L	516, 540, 623, 808, 1486, 1904, 1917, 2598, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 8332, 8492, 9223, 9364, 10874	42.1	3,725.8	58,665
contaminant injection duration = 10 h	495, 636, 877, 1477, 1486, 1806, 2602, 3229, 3357, 3635, 3770, 3836, 4247, 4306, 6922, 7441, 7664, 9364, 10716, 11687	45.3	3,516.6	67,210
2 injection nodes at 1 time	636, 813, 1229, 1486, 2602, 3229, 3357, 3629, 3770, 3836, 4247, 4306, 6922, 7441, 7667, 8942, 9364, 10720, 11304, 11687	66.6	2,366.7	52,765
2 injection nodes at 2 times	636, 813, 1229, 1486, 1514, 2602, 3229, 3357, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8942, 9364, 10421, 10720, 12404	64.7	2,614.4	55,525

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

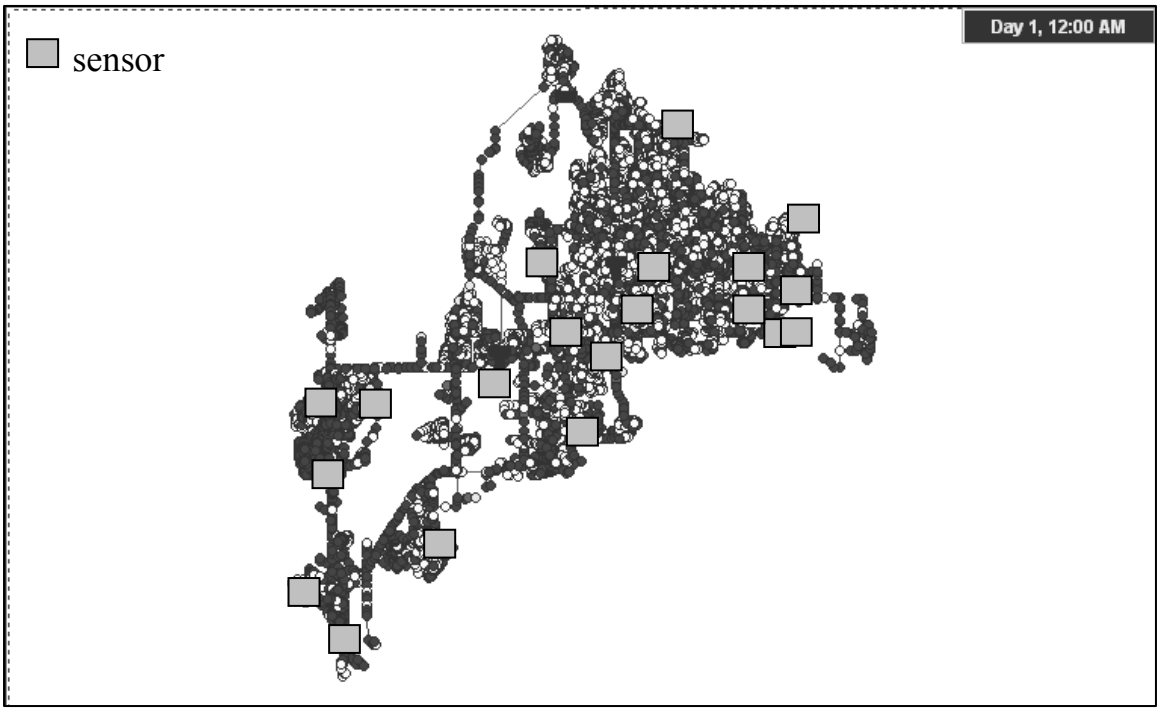


Figure 6.11 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of a detection resolution of 0.3 mg/L for BWSN Network 2.

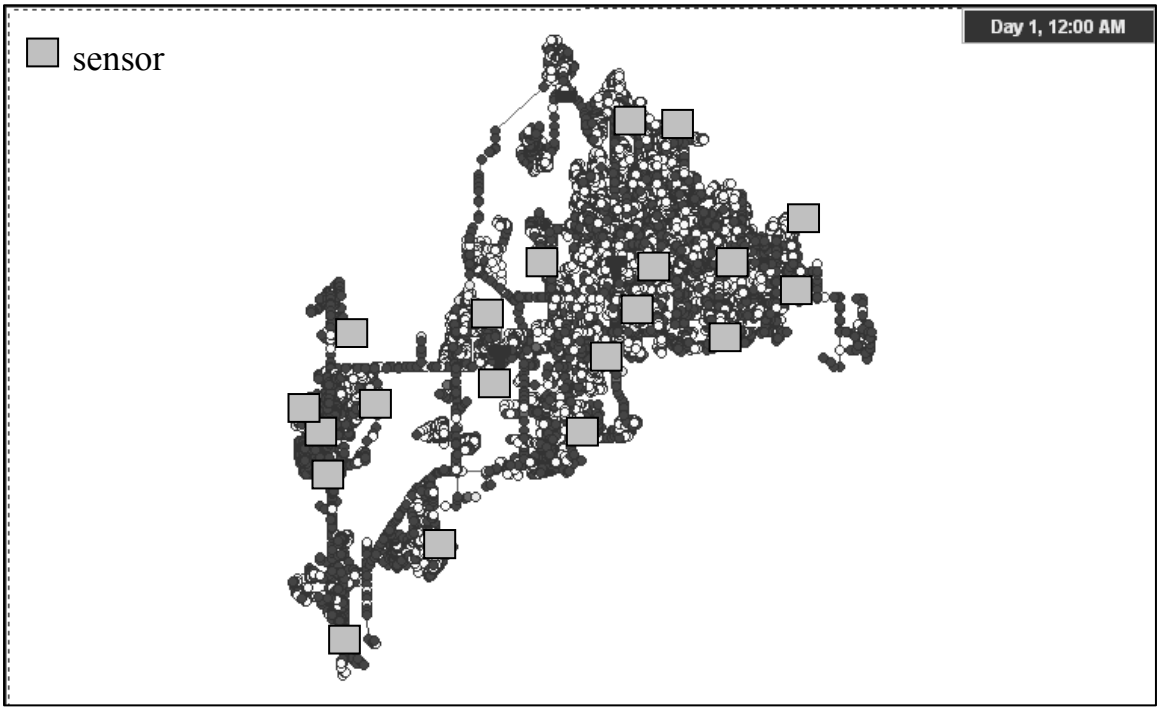


Figure 6.12 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of a 10-hour injection duration for BWSN Network 2.

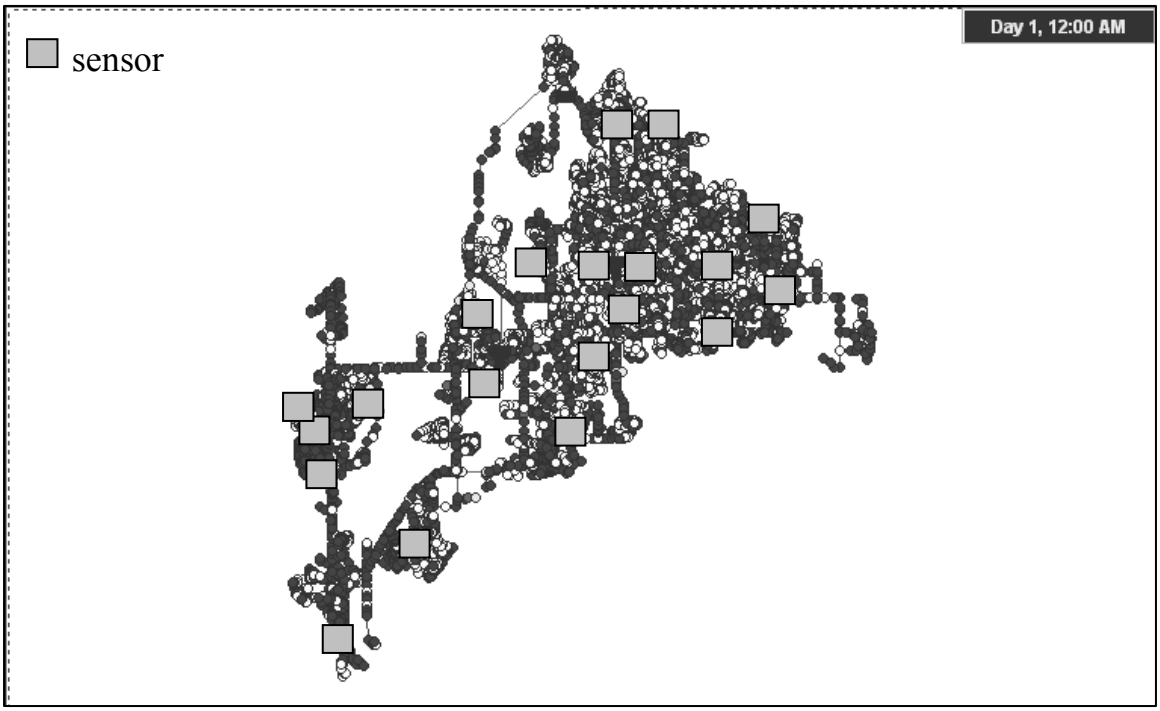


Figure 6.13 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of 2 nodes injected at 1 time for BWSN Network 2.

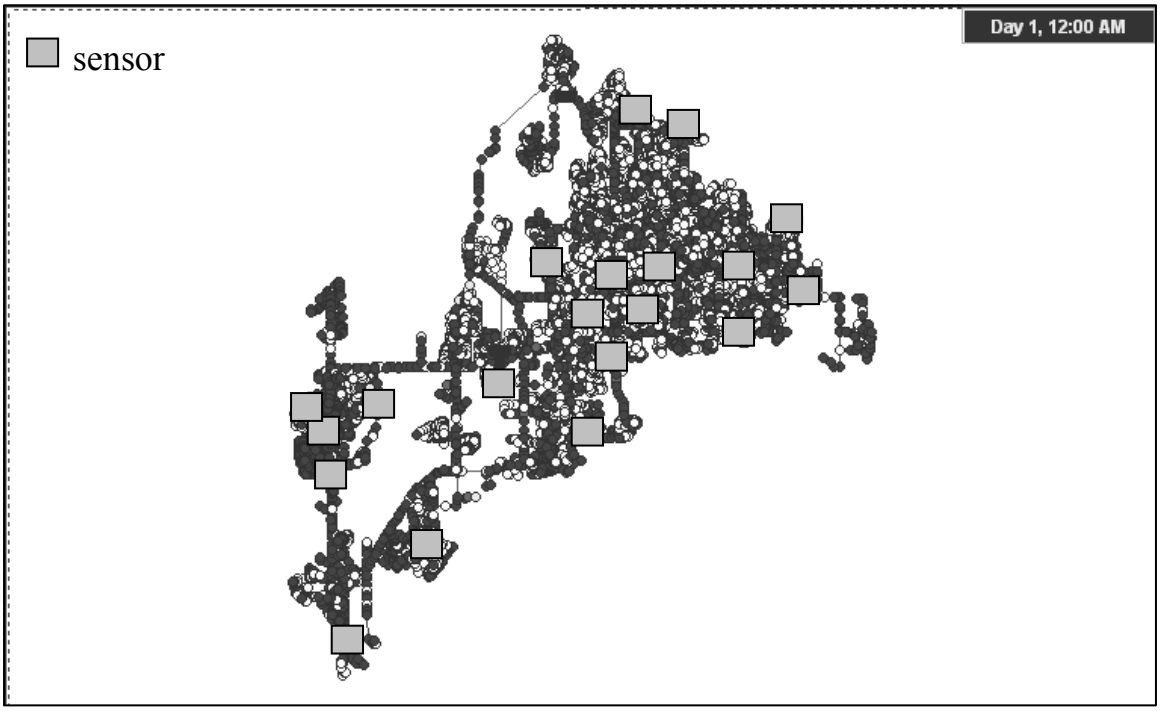


Figure 6.14 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of 2 nodes injected at 2 times for BWSN Network 2.

Table 6.11 Comparison of optimal sensor locations for minimizing expected detection time with $M = 20$ in the variant attack cases for BWSN Network 2 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{time}** (min)
contaminant injection duration = 10 h	This Study	495, 636, 877, 1477, 1486, 1806, 2602, 3229, 3357, 3635, 3770, 3836, 4247, 4306, 6922, 7441, 7664, 9364, 10716, 11687	3,516.6
	Krause et al. (2006)	636, 877, 1229, 1430, 1486, 2602, 3229, 3357, 3635, 3770, 3836, 4115, 4247, 4306, 5346, 7441, 7666, 9210, 9364, 11687	3,530.4
2 injection nodes at 1 time	This Study	636, 813, 1229, 1486, 2602, 3229, 3357, 3629, 3770, 3836, 4247, 4306, 6922, 7441, 7667, 8942, 9364, 10720, 11304, 11687	2,366.7
	Krause et al. (2006)	636, 877, 1229, 1486, 2602, 3229, 3357, 3635, 3770, 3836, 4115, 4240, 4306, 5346, 7441, 7664, 9210, 9364, 11304, 11687	2,368.2

* All sensor node numbers preceded by "JUNCTION-".

** Values computed using the software developed in this work.

6.2.2 EVALUATION OF NODAL IMPORTANCE

The results discussed above were found by employing ISSM with the aid of nodes ranked by the nodal importance function in (4.23). Considering the ranks of the lowest ranked solution nodes given in Table 6.12, it appears the nodal importance approach for minimizing expected detection time is effective in narrowing the search for the optimal solution in the default attack case for this large system. The solution for 5 sensors employed was isolated to less than 8% of WDS nodes, and for 20 sensors employed the solution was isolated to less than 16% of WDS nodes. With regard to the variant attack cases, Table 6.13 shows that the degree of narrowing was practically the same for the case of a 0.3 mg/L sensor detection resolution, but for the 10-hour injection duration and two injection node cases, finding the optimal solution requires a broader proportion of WDS nodes. It is suspected that the reason for lower ranks for some sensor nodes in these cases is the increased opportunities for some nodes to experience contamination making them more desirable in terms of importance but also less distinguishable from other nodes that are important. Despite these larger proportions, the search domain is narrowed considerably (i.e., less than 30%) for all attack cases.

Examining the solution convergence plots in Figures 6.15 and 6.16 also reveals points about the effectiveness of nodal importance for this single-objective problem.
When 5 sensors are employed, the lowest-ranked optimal sensor node is ranked 943, but a subset size of 600 (540 ranked nodes, 60 unranked nodes) allows ISSM to find a very good solution close to the optimal. For the one injection node cases with 20 sensors

employed, ISSM can find near-optimal solutions with subset sizes of about 2,000. Solution convergence was much more gradual for the two injection node cases, highlighting the need in those cases for subset sizes of at least 3,000 to find even a near-optimal solution.

Table 6.12 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for minimizing expected detection time in the default attack case and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	943	7.5%
20	1,935	15.5%

Table 6.13 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for minimizing expected detection time with $M = 20$ in the variant attack cases and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	1,935	15.5%
sensor detection resolution = 0.3 mg/L	1,788	14.3%
contaminant injection duration = 10 h	2,790	22.3%
2 injection nodes at 1 time	2,831	22.6%
2 injection nodes at 2 times	3,732	29.8%

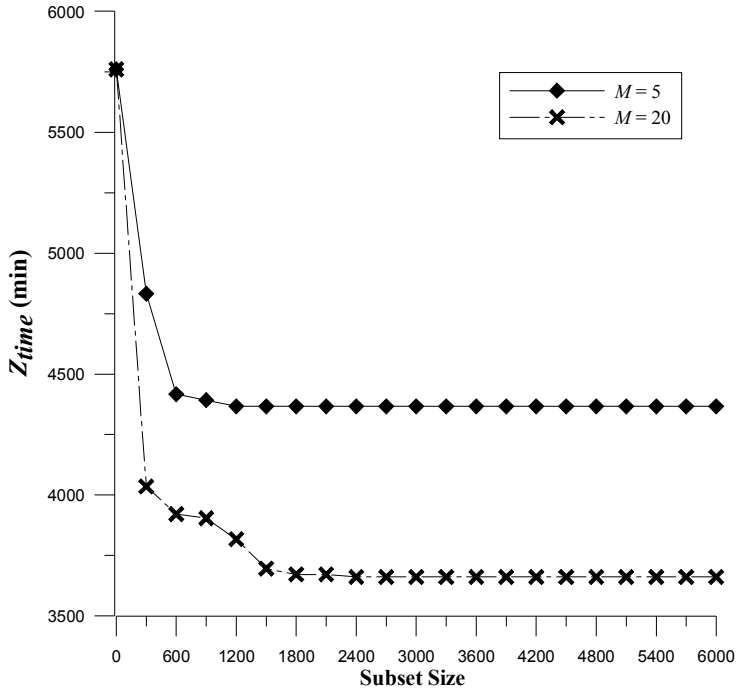


Figure 6.15 Solution convergence as ISSM is carried out for minimizing expected detection time in the default attack case for BWSN Network 2. (Only solution values for subset sizes of 6,000 or less are plotted as solution reached best performance for both M = 5 and M = 20 by subset size = 6,000.)

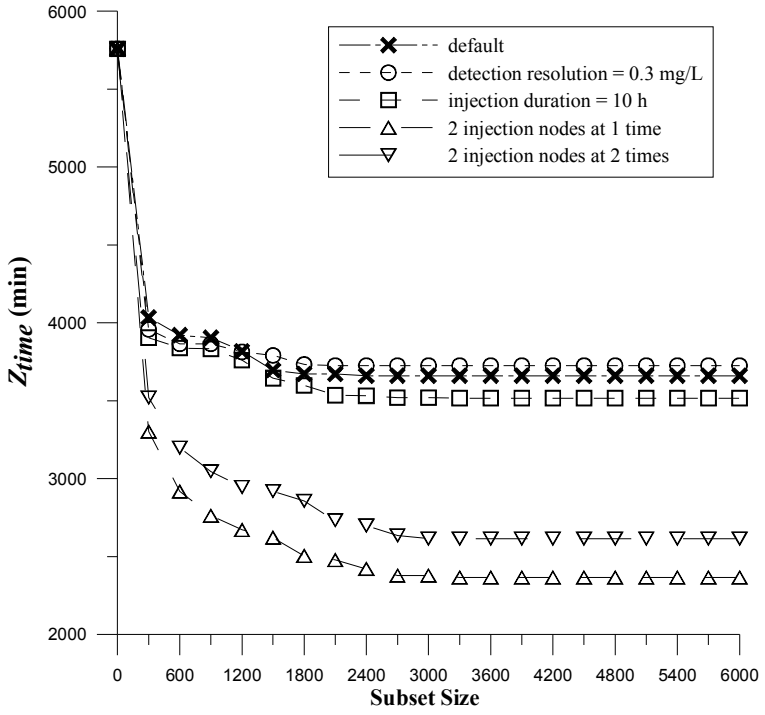


Figure 6.16 Solution convergence as ISSM is carried out for minimizing expected detection time with $M = 20$ in the default and variant attack cases for BWSN Network 2. (Only solution values for subset sizes of 6,000 or less are plotted as solutions for all cases reached best performance by subset size = 6,000.)

6.3 SINGLE-OBJECTIVE PROBLEM: MINIMIZING Z_{vol}

Sensor placement findings for minimizing expected contaminated demand volume are provided and discussed in this section.

6.3.1 SENSOR NODE SOLUTIONS

Sensor nodes selected to minimize expected contaminated demand volume in the default attack case for BWSN Network 2 are given in Table 6.14, and their locations in the system are depicted in Figures 6.17 and 6.18. Compared with the Z_{vol} values in Tables 6.1 and 6.8, the Z_{vol} values in Table 6.14 seem significantly reduced with sensors located at the nodes indicated. To achieve the low volumes reported in Table 6.14, some protection was sacrificed in terms of minimizing Z_{time} and significant protection was sacrified in terms of maximizing Z_{lik}. As detecting contamination sooner is key to minimizing Z_{vol}, it is not surprising that the sacrifices with respect to Z_{time} are not extensive and that some sensor nodes for minimizing Z_{vol} and minimizing Z_{time} match (e.g., JUNCTION-10874). Figures 6.17 and 6.18 show remarkable shifts of sensor nodes from the outer regions of the system where sensors are placed to maximize coverage to more central locations that are more upstream and on or near major transmission lines so that contamination is detected sooner and major amounts of water volume can be protected. These trends in sensor placement were observed in general for minimizing expected contaminated demand volume for BWSN Network 1 as well.

Table 6.14 Optimal sensor locations and performance measures for minimizing expected contaminated demand volume in the default attack case for BWSN Network 2.

M	Sensor Nodes*	Z_{lik} (%)	Z_{time} (min)	Z_{vol}** (gal)
5	3357, 4684, 10874, 11184, 11304	24.1	4,545.6	59,263
20	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	33.5	3,994.4	12,732

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

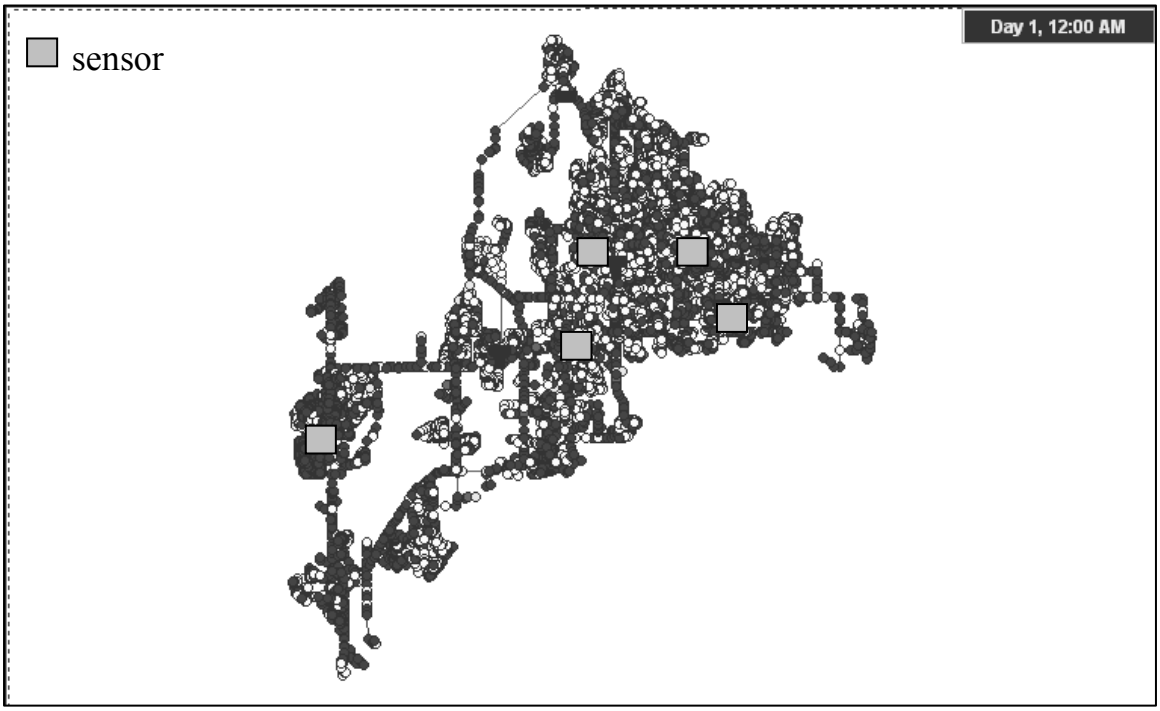


Figure 6.17 Optimal sensor locations with $M = 5$ for minimizing expected contaminated demand volume in the default attack case for BWSN Network 2.

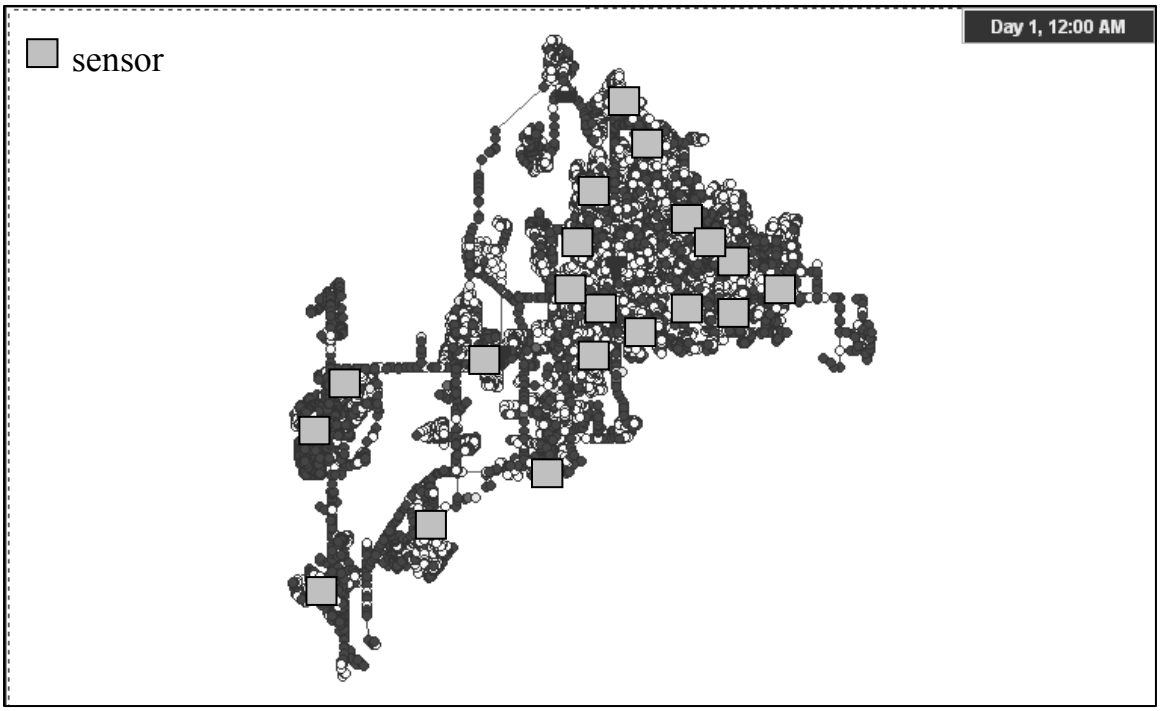


Figure 6.18 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the default attack case for BWSN Network 2.

Table 6.15 gives encouraging comparisons of the results of this work for this single-objective problem in the default attack case with results of other works, although in some cases the results of another work are marginally better than those of this work. With 5 sensors employed, the nodes selected match those selected by Berry et al. (2006c), which are the best performing nodes of the sets listed in the table. However, when 20 sensors are employed, the solution for this work is slightly less performing than that of Berry et al., but is superior to those provided by the other works listed. The difference in Z_{vol} between this work and the work of Berry et al. is only about 600 gal—mostly negligible given the potential for many more thousands of gallons of water that could become contaminated given even subtle differences in sensor placement from the nodes selected in this study. Therefore, even though it appears that the optimization method of this work

did not find the true optimal solution when 20 sensors were employed, the solution found performs very well. The reasons for the suboptimal solution are explored in discussion below regarding the effectiveness of nodal importance for this single-objective problem.

Table 6.15 Comparison of optimal sensor locations for minimizing expected contaminated demand volume in the default attack case for BWSN Network 2 determined in this study with locations found in other works.

M	**Work**	**Sensor Nodes***	Z_{vol}** **(gal)**
5	This Study	3357, 4684, 10874, 11184, 11304	59,263
	Berry et al. (2006c)	3357, 4684, 10874, 11184, 11304	59,263
	Eliades and Polycarpou (2006)	532, 1486, 3836, 4359, 8445	102,293
	Krause et al. (2006)	339, 3357, 4651, 10614, 11271	70,514
20	This Study	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	12,732
	Berry et al. (2006c)	636, 1917, 3357, 3573, 3770, 4132, 4240, 4594, 5114, 6583, 6700, 7652, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	12,153
	Eliades and Polycarpou (2006)	532, 1426, 1486, 3231, 3836, 4234, 4359, 4511, 4609, 5087, 5585, 6922, 7670, 7858, 8403, 8629, 9360, 9787, 10885, 12167	55,012
	Krause et al. (2006)	339, 636, 778, 1478, 2073, 3357, 4032, 4084, 4240, 4651, 4793, 6700, 7652, 9142, 9722, 10614, 10874, 11167, 11271, 12258	13,647

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.

Table 6.16 gives the sensor nodes selected in this work to minimize expected contaminated demand volume for the variant attack cases, and Figures 6.19 through 6.23 depict the locations of these nodes in the system for cases that have different sensor node

sets than that for the default attack case. As seen for the other protection goals, sensor node locations do not vary greatly as attack conditions change.

Z_{vol} values vary in the expected manners relative to the default attack case. A slight adjustment in the sensor placement scheme is warranted for the case of a higher detection resolution as some nodes selected for the default attack case would not be able to detect contaminant with the higher resolution and would lead to more scenarios with very high contaminated volumes due to non-detection; the sensor node rearrangement allows for preservation of coverage to prevent these instances of very high contaminated volumes at the cost of some performance in terms of the overall Z_{vol} value. The increases in Z_{vol} for the cases of increased injection duration, response delay, and two injection nodes are expected due to the increased opportunity for water to become contaminated. The biggest change in Z_{vol} value comes as a result of a 0.0 mg/L hazard threshold, apparently due to a large amount of volume that is considered contaminated when the hazard threshold is reduced. This observation implies that Z_{vol} can be rather sensitive to contaminant hazard threshold.

Table 6.16 Optimal sensor locations and performance measures for minimizing expected contaminated demand volume with $M = 20$ in the variant attack cases for BWSN Network 2.

Attack Case	Sensor Nodes*	Z_{lik} (%)	Z_{time} (min)	Z_{vol}** (gal)
default	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	33.5	3,994.4	12,732
sensor detection resolution = 0.3 mg/L	636, 1917, 3357, 3453, 3770, 4132, 4253, 4594, 5114, 6583, 6700, 8060, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	33.7	3,982.1	16,094
contaminant injection duration = 10 h	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7652, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	36.7	3,846.1	28,467
contaminant hazard threshold = 0.0 mg/L	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5109, 6309, 6700, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	33.6	3,999.9	33,638
response delay = 3 h	636, 1917, 3357, 3453, 3770, 4132, 4594, 4793, 6309, 6700, 7336, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	33.2	4,009.5	23,238
2 injection nodes at 1 time	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	56.5	2,754.4	19,963
2 injection nodes at 2 times	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 8055, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	55.8	2,907.4	19,662

* All sensor node numbers preceded by "JUNCTION-".
** Performance measure for protection goal of interest.

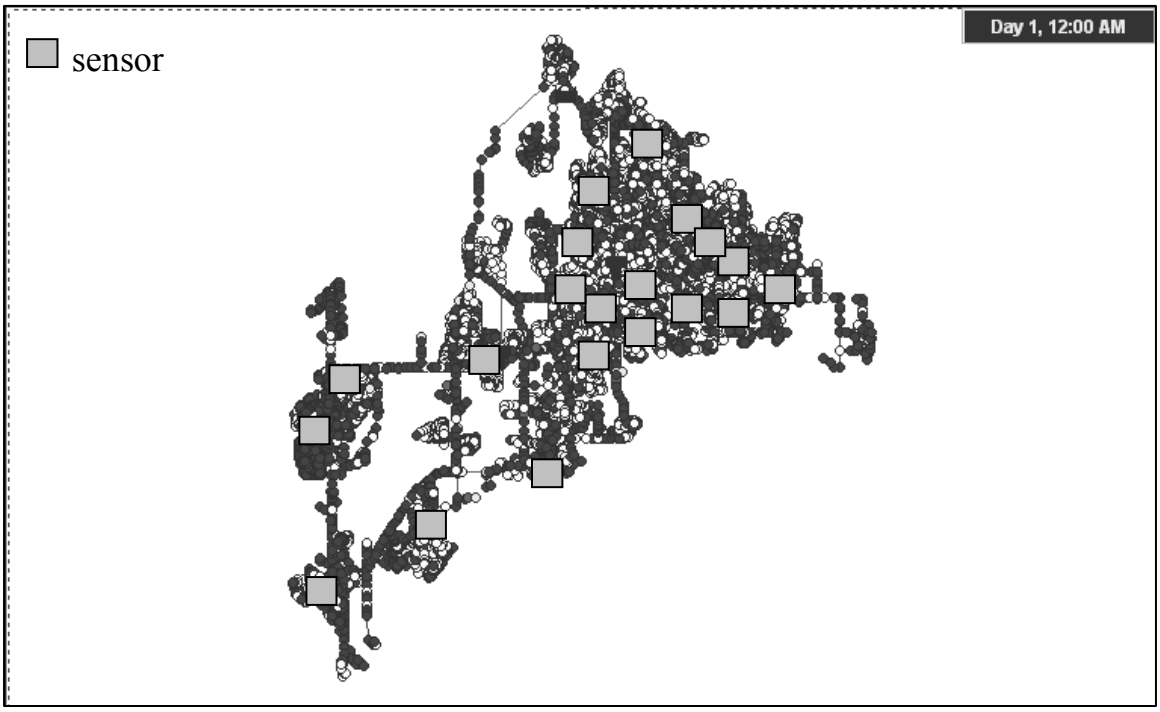


Figure 6.19 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a detection resolution of 0.3 mg/L for BWSN Network 2.

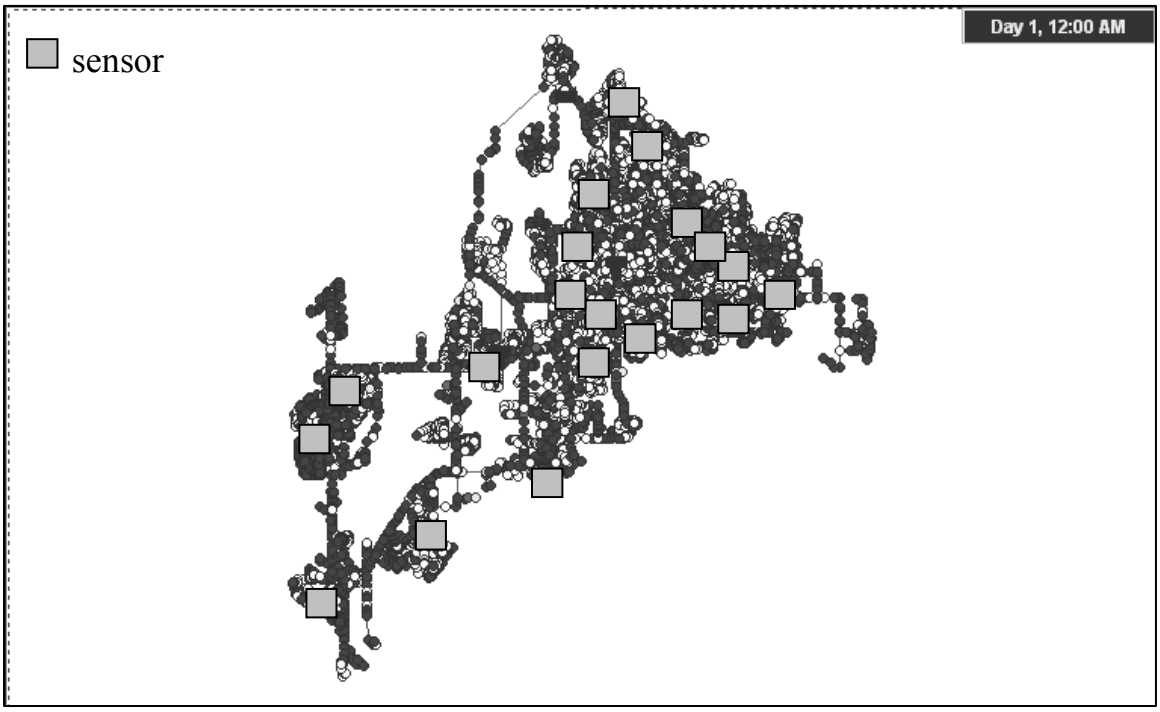


Figure 6.20 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a 10-hour injection duration for BWSN Network 2.

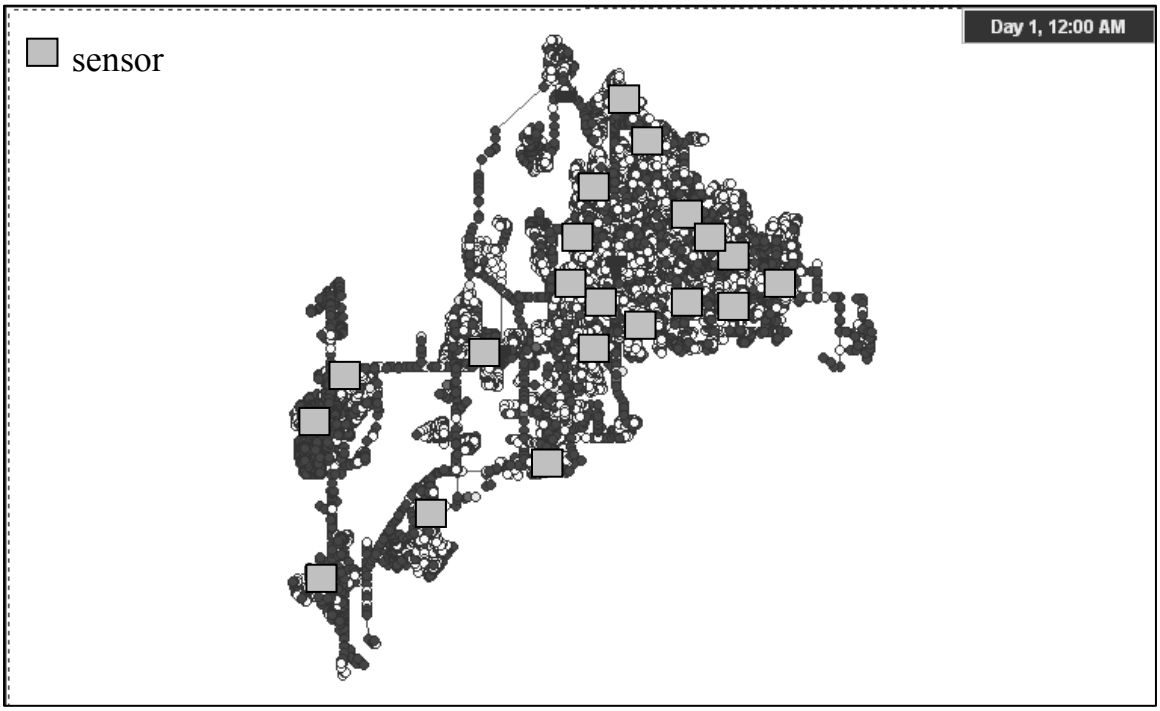


Figure 6.21 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a hazard threshold of 0.0 mg/L for BWSN Network 2.

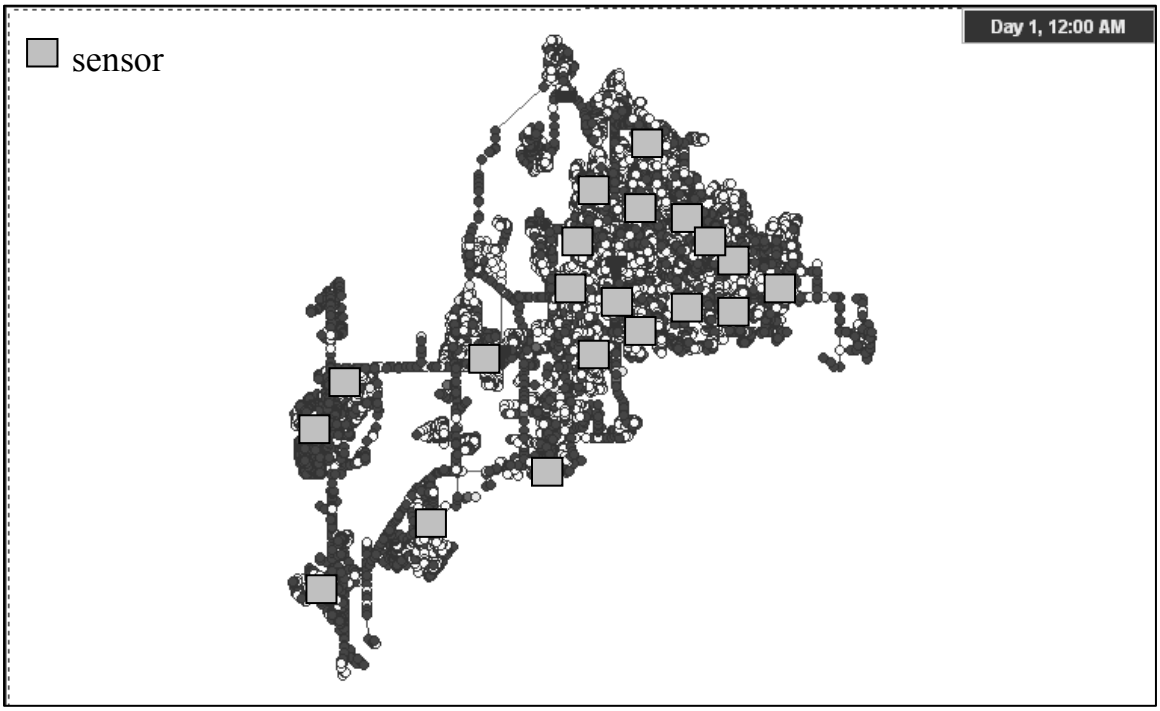


Figure 6.22 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a 3-hour response delay for BWSN Network 2.

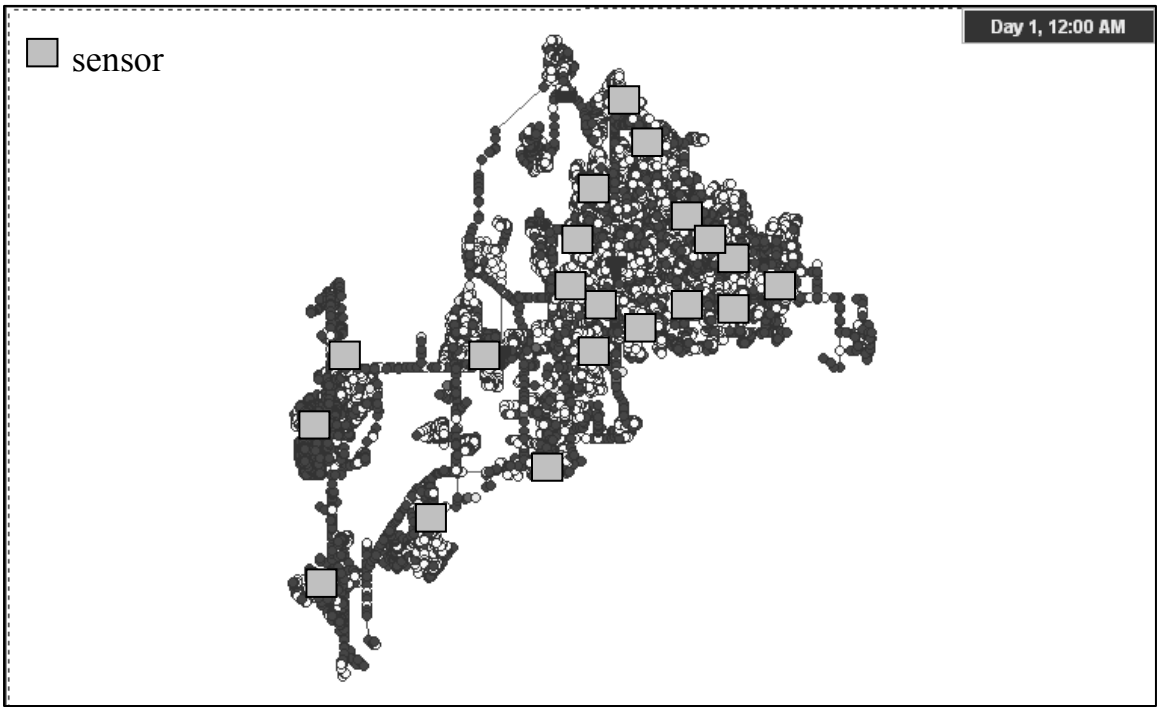


Figure 6.23 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of 2 nodes injected at 2 times for BWSN Network 2.

Table 6.17 compares the results of this work for the variant attack cases with available results from literature. For both the case of the 10-hour injection duration and the 3-hour response delay, the results of this study are slightly less performing than those of Berry et al. (2006c), but perform very well relative to those of Krause et al. (2006), as in the default attack case. As stated above, Berry et al. could not perform optimization for the two injection node cases, so they have no results for comparison for those cases. The results of this study for those cases are superior to those submitted by Krause et al. In sum, even though Berry et al. submitted marginally better results versus the results of this work for multiple cases, the optimization method of this work can be carried out for all cases and provide near-optimal results.

Table 6.17 Comparison of optimal sensor locations for minimizing expected contaminated demand volume with $M = 20$ in the variant attack cases for BWSN Network 2 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{vol}** (gal)
contaminant injection duration = 10 h	This Study	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7652, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	28,467
	Berry et al. (2006c)	636, 1917, 3357, 3573, 3770, 4132, 4240, 4594, 5114, 6583, 6700, 7652, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	27,326
	Krause et al. (2006)	339, 636, 3357, 3573, 4032, 4084, 4132, 4240, 4684, 5121, 6583, 6981, 8850, 9142, 9364, 10614, 10874, 11304, 11167, 12258	29,051
response delay = 3 h	This Study	636, 1917, 3357, 3453, 3770, 4132, 4594, 4793, 6309, 6700, 7336, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	23,238
	Berry et al. (2006c)	636, 1917, 3357, 3573, 3770, 4032, 4132, 4793, 6309, 6700, 7336, 8852, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	22,956
	Krause et al. (2006)	339, 636, 1115, 3328, 4032, 4132, 4684, 4793, 6309, 6637, 6700, 7256, 7336, 8852, 8907, 8999, 9142, 10874, 11167, 11271	28,505
2 injection nodes at 1 time	This Study	636, 1522, 1917, 3357, 3453, 3770, 4132, 4594, 5114, 6583, 6700, 7653, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	19,963
	Krause et al. (2006)	339, 636, 3357, 3573, 4032, 4084, 4132, 4684, 5121, 6309, 6981, 7652, 9142, 9722, 10614, 10874, 11167, 11304, 12258, 12481	21,643

* All sensor node numbers preceded by "JUNCTION-".

** Values computed using the software developed in this work.

6.3.2 EVALUATION OF NODAL IMPORTANCE

To obtain the sensor node solutions presented above, ISSM was carried out making use of nodal importance values calculated with the large-system nodal importance function for minimizing expected contaminated demand volume in (4.27). As indicated by Table 6.18, the nodal importance function effectively narrowed the optimal solution when 5 sensors are employed to 7% of WDS nodes. The nodal importance function also allowed the near-optimal solution given above for the case of 20 sensors employed to be isolated to approximately the top one-fifth of ranked nodes. Solutions for the other attack cases are also narrowed to less than one-quarter of nodes according to the data given in Table 6.19. Solution convergence plots in Figures 6.24 and 6.25 show overall gradual convergence to the best solution for all cases.

Upon examination of sensor nodes selected in this work and those submitted by Berry et al. (2006c) for all cases to determine the reason for the slightly suboptimal sensor nodes of this work, it is observed that two nodes selected by Berry et al. that allowed for their lower Z_{vol} values to be found were fairly lowly ranked according to the values found with the importance function in (4.27). It would be impractical to expect ISSM to consistently find the complete set of optimal sensor nodes with the larger subsets required to include these two nodes. Thus, the large-system importance function for minimizing expected contaminated demand volume has drawbacks, but it is encouraging that so many of the optimal sensor nodes are highly ranked. Future study may find minor adjustments to the function such that those two nodes more appropriately ranked.

Table 6.18 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for minimizing expected contaminated demand volume in the default attack case and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

M	Lowest Nodal Importance Rank	Proportion of Ranked WDS Nodes Containing Optimal Solution
5	871	7.0%
20	2,694	21.5%

Table 6.19 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for minimizing expected contaminated demand volume with $M = 20$ in the variant attack cases and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

Attack Case	Lowest Nodal Importance Rank	Proportion of Ranked WDS Nodes Containing Optimal Solution
default	2,694	21.5%
sensor detection resolution = 0.3 mg/L	2,712	21.6%
contaminant injection duration = 10 h	2,978	23.8%
contaminant hazard threshold = 0.0 mg/L	2,482	19.8%
response delay = 3 h	2,694	21.5%
2 injection nodes at 1 time	2,864	22.9%
2 injection nodes at 2 times	2,821	22.5%

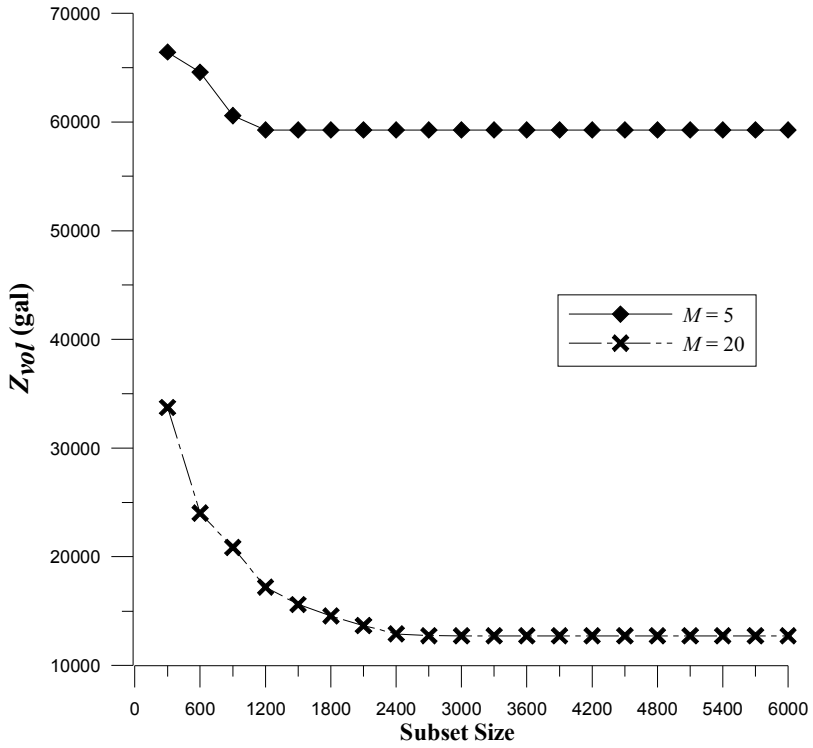


Figure 6.24 Solution convergence as ISSM is carried out for minimizing expected contaminated demand volume in the default attack case for BWSN Network 2. (Only solution values for subset sizes of 6,000 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 6,000.)

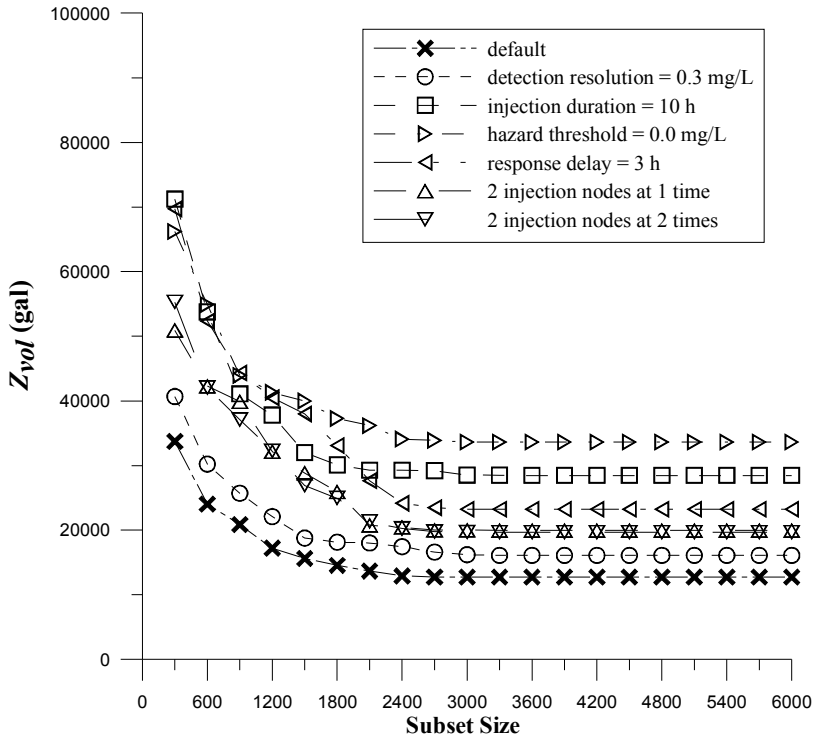


Figure 6.25 Solution convergence as ISSM is carried out for minimizing expected contaminated demand volume with $M = 20$ in the default and variant attack cases for BWSN Network 2. (Only solution values for subset sizes of 6,000 or less are plotted as solutions for all cases reached best performance by subset size = 6,000.)

6.4 MULTIOBJECTIVE PROBLEM: MAXIMIZING Z_{all}

Optimization results for the three-prong multiobjective problem are given and discussed below.

6.4.1 SENSOR NODE SOLUTIONS

Table 6.20 presents the sensor nodes found in this work for the three-prong multiobjective problem in the default attack case for BWSN Network 2, and Figures 6.26 and 6.27 illustrate the locations of these sensor nodes in the system. As was seen in examining the results for BWSN Network 1 in Chapter 5, sensor node results for maximizing Z_{all} for BWSN Network 2 tend to preserve most if not all of the protection with respect to minimizing expected detection time while compromising larger amounts of protection in terms of maximizing detection likelihood and minimizing expected contaminated demand volume. The multiobjective results of this work are compared with other multiobjective results from other works in Table 6.21. Again, other works considered other factors in their multiobjective approaches, but the comparison serves to ensure that other sets of nodes submitted in literature are not superior to those found in this work. For both $M = 5$ and $M = 20$, Z_{all} values for the nodes selected in this study are equal or better than the values calculated using nodes submitted by other works.

Table 6.20 Optimal sensor locations and performance measures for the three-prong multiobjective problem in the default attack case for BWSN Network 2.

M	Sensor Nodes*	Z_{all}	Z_{lik} (%)	Z_{time} (min)	Z_{vol} (gal)
5	551, 1486, 3770, 8336, 10874	2.42	29.9	4,366.9	97,510
20	551, 850, 1486, 1904, 1917, 2602, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10703, 10874	2.55	42.1	3,679.1	43,952

* All sensor node numbers preceded by "JUNCTION-".

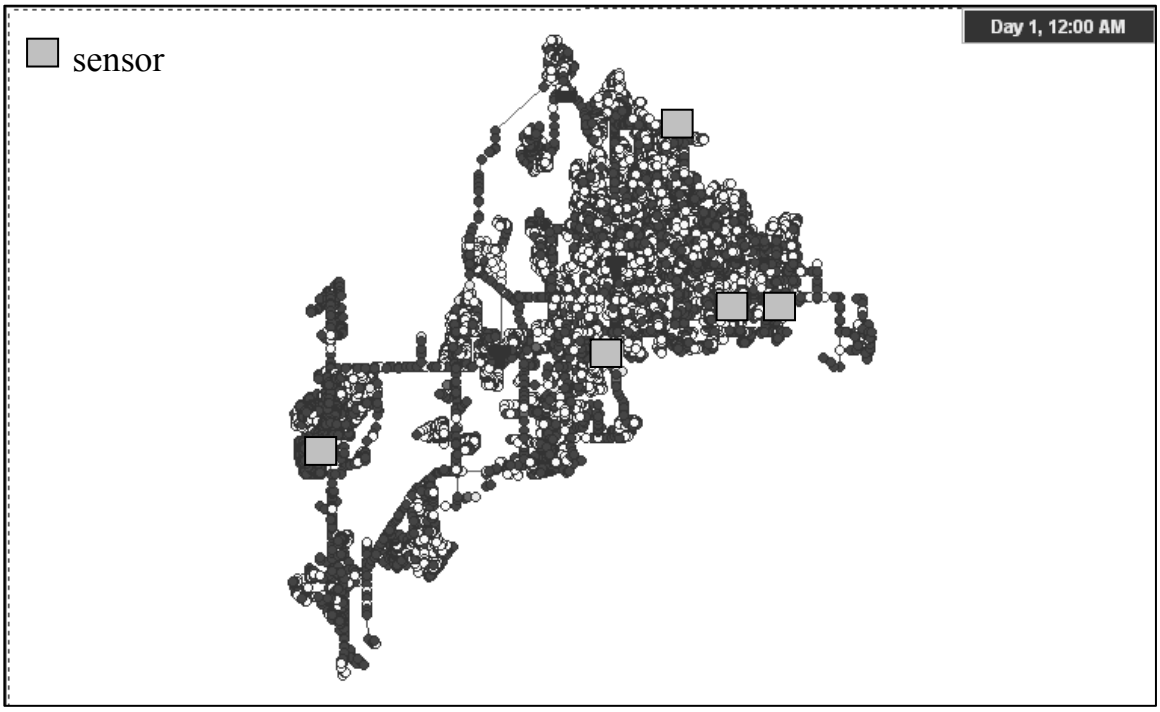


Figure 6.26 Optimal sensor locations with $M = 5$ for the three-prong multiobjective problem in the default attack case for BWSN Network 2.

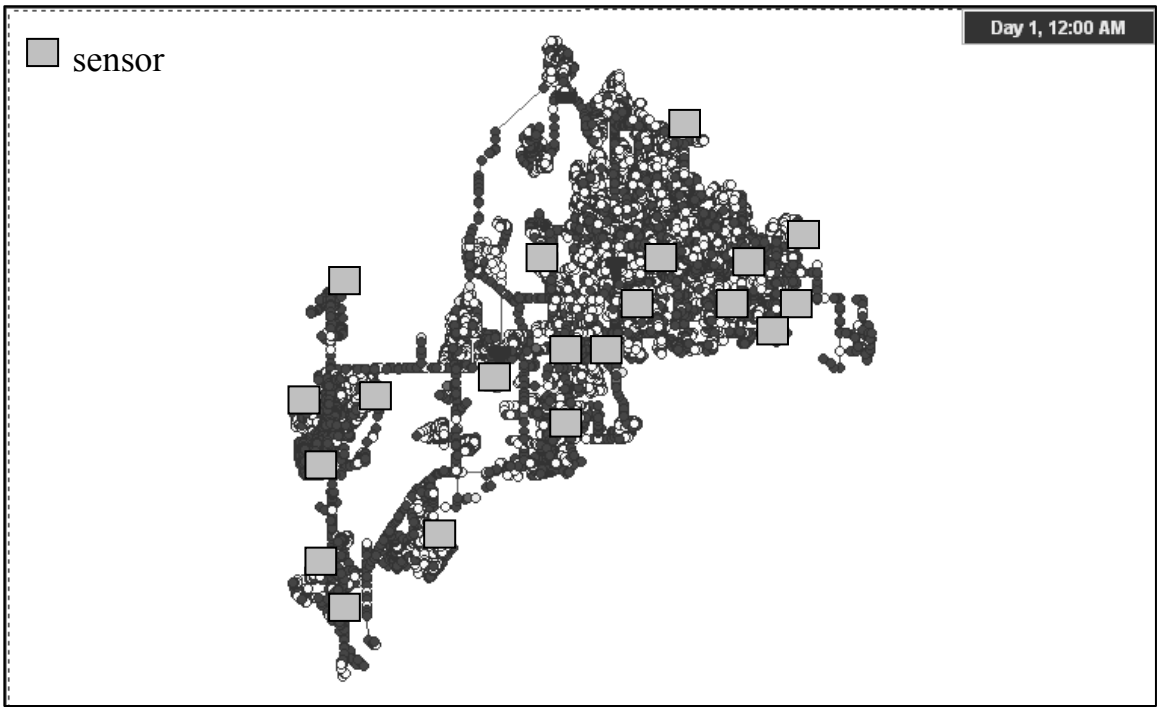


Figure 6.27 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the default attack case for BWSN Network 2.

Table 6.21 Comparison of optimal sensor locations for the three-prong multiobjective problem in the default attack case for BWSN Network 2 determined in this study with locations found in other works.

M	Work	Sensor Nodes*	Z_{all}**	Z_{lik}** (%)	Z_{time}** (min)	Z_{vol}** (gal)
5	This Study	551, 1486, 3770, 8336, 10874	2.42	29.9	4,366.9	97,510
	Dorini et al. (2006)	636, 3585, 4684, 9364, 10387	1.96	28.6	4,456.6	85,342
	Eliades and Polycarpou (2006)	532, 1486, 3357, 4359, 4609	1.80	28.4	4,459.0	100,660
	Guan et al. (2006)	321, 3770, 4084, 4939, 7762	***	21.2	4,696.4	75,824
	Huang et al. (2006)	3355, 5088, 5430, 9005, 9550	***	21.5	4,709.7	87,978
	Krause et al. (2006)	3357, 4684, 10874, 11184, 11304	1.22	24.1	4,545.6	59,263
	Wu and Walski (2006)	3709, 4957, 6583, 8357, 9364	2.42	30.0	4,368.7	98,195
20	This Study	551, 850, 1486, 1904, 1917, 2602, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10703, 10874	2.55	42.1	3,679.1	43,952
	Dorini et al. (2006)	647, 928, 1478, 1872, 2223, 2848, 3573, 4650, 5076, 5366, 6835, 7422, 8336, 8402, 9204, 9364, 10874, 11271, 11528, 12377	1.61	38.2	3,882.8	42,117
	Eliades and Polycarpou (2006)	532, 1426, 1486, 1976, 3231, 3679, 3836, 4234, 4359, 4609, 5087, 5585, 6922, 7670, 7858, 8629, 9360, 9787, 10885, 12167	1.50	39.2	3,916.8	55,315
	Guan et al. (2006)	174, 311, 1486, 1905, 2589, 2991, 3548, 3757, 3864, 4184, 4238, 5091, 6995, 7145, 7689, 8826, 9308, 9787, 10614, 12086	***	29.3	4,244.0	32,381
	Huang et al. (2006)	73, 108, 1028, 1112, 1437, 2526, 3180, 4036, 4648, 5363, 5826, 5879, 6581, 8439, 8580, 8841, 9363, 9616, 10216, 10385	***	32.7	4,133.0	37,217
	Krause et al. (2006)	1478, 1904, 2579, 3229, 3357, 3635, 3747, 3836, 4032, 4132, 4240, 4684, 6700, 8834, 8999, 9142, 9364, 10874, 11184, 11304	2.08	38.1	3,802.5	15,430
	Wu and Walski (2006)	871, 1334, 2589, 3115, 3640, 3719, 4247, 4990, 5630, 6733, 7442, 7714, 8387, 8394, 9778, 10290, 10522, 10680, 11151, 11519	2.04	39.8	3,781.6	38,201

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.
*** Solution falls outside of Pareto front space.

Table 6.22 gives the multiobjective sensor placement results found in this study for the variant attack cases, and these results further show that maximizing Z_{all} tends to preserve protection according to minimizing Z_{time}. Figures 6.28 through 6.32 illustrate the locations of the sensor nodes selected for the variant cases that have different sensor node locations than those of the default attack case. Table 6.23 shows that the results of this study when 20 sensors are employed for the applicable variant cases are superior to results of other works if evaluated under the multiobjective problem formulation developed in this work.

Table 6.22 Optimal sensor locations and performance measures for the three-prong multiobjective problem with $M = 20$ in the variant attack cases for BWSN Network 2.

Attack Case	Sensor Nodes*	Z_{all}	Z_{lik} (%)	Z_{time} (min)	Z_{vol} (gal)
default	551, 850, 1486, 1904, 1917, 2602, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10703, 10874	2.55	42.1	3,679.1	43,952
sensor detection resolution = 0.3 mg/L	551, 636, 850, 1486, 1904, 1917, 2602, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 8311, 8403, 9210, 9364, 10703, 10874	2.58	41.5	3,725.8	50,446
contaminant injection duration = 10 h	495, 636, 850, 1486, 1514, 1806, 2602, 3229, 3357, 3573, 3770, 3836, 4247, 4306, 7441, 7664, 9022, 9364, 10716, 11687	2.60	44.7	3,535.9	64,757
contaminant hazard threshold = 0.0 mg/L	551, 877, 1486, 1806, 1904, 1917, 2602, 3229, 3573, 3770, 3836, 4247, 4306, 7441, 7667, 8444, 9212, 9364, 10703, 10874	2.64	41.9	3,695.6	60,083
response delay = 3 h	551, 850, 1486, 1904, 1917, 2602, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10703, 10874	2.47	42.1	3,679.1	69,052
2 injection nodes at 1 time	636, 850, 1229, 1486, 1514, 2602, 3229, 3357, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8942, 9364, 10720, 11304, 11687	2.74	66.4	2,378.4	52,111
2 injection nodes at 2 times	636, 850, 1229, 1486, 1514, 2602, 3229, 3357, 3629, 3770, 3836, 4247, 4306, 4806, 7441, 7667, 8942, 9364, 10703, 10720	2.74	64.2	2,616.0	51,807

* All sensor node numbers preceded by "JUNCTION-".

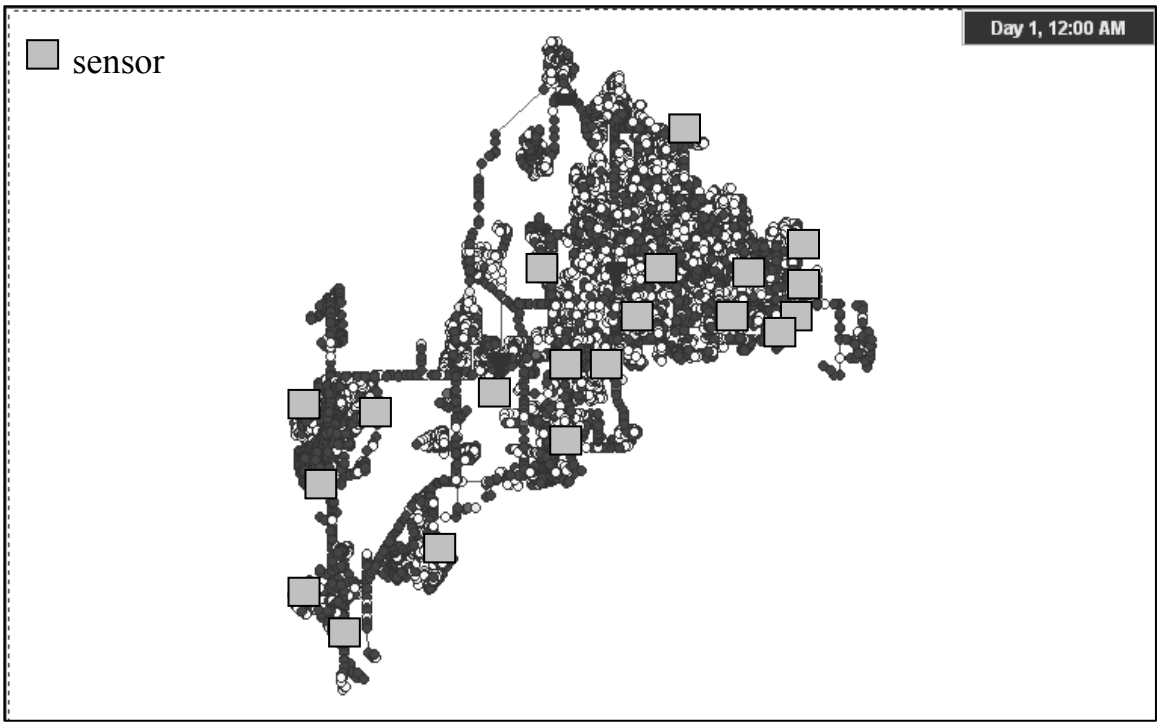


Figure 6.28 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a detection resolution of 0.3 mg/L for BWSN Network 2.

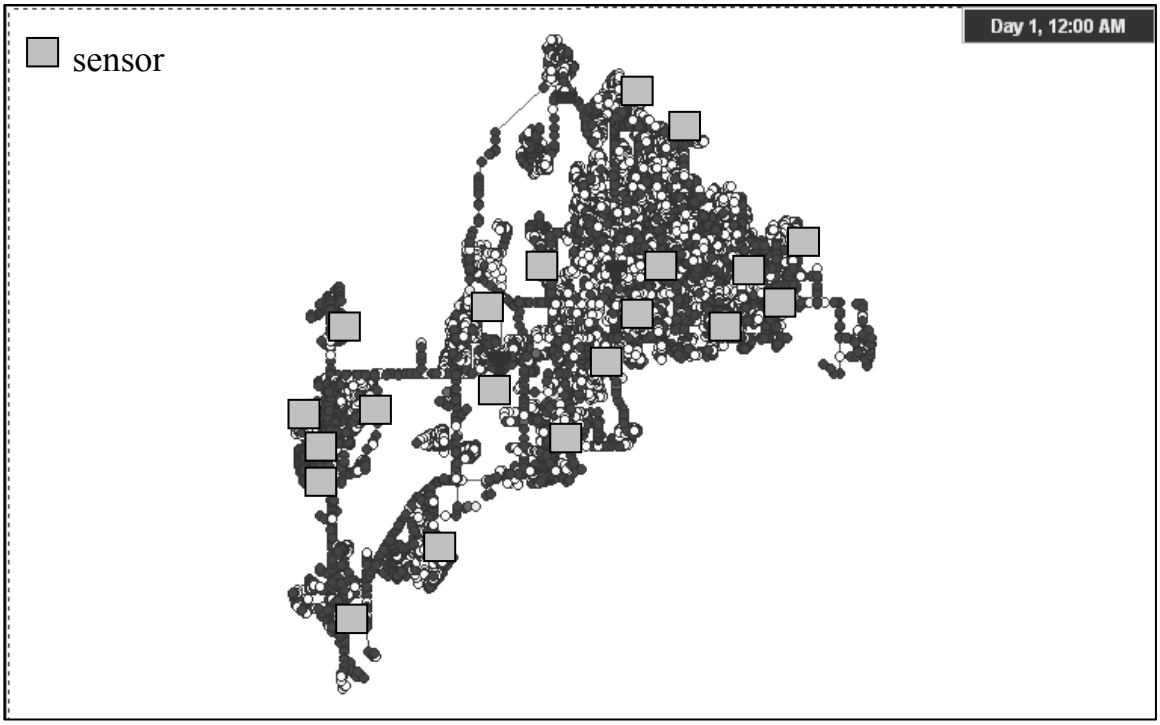


Figure 6.29 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a 10-hour injection duration for BWSN Network 2.

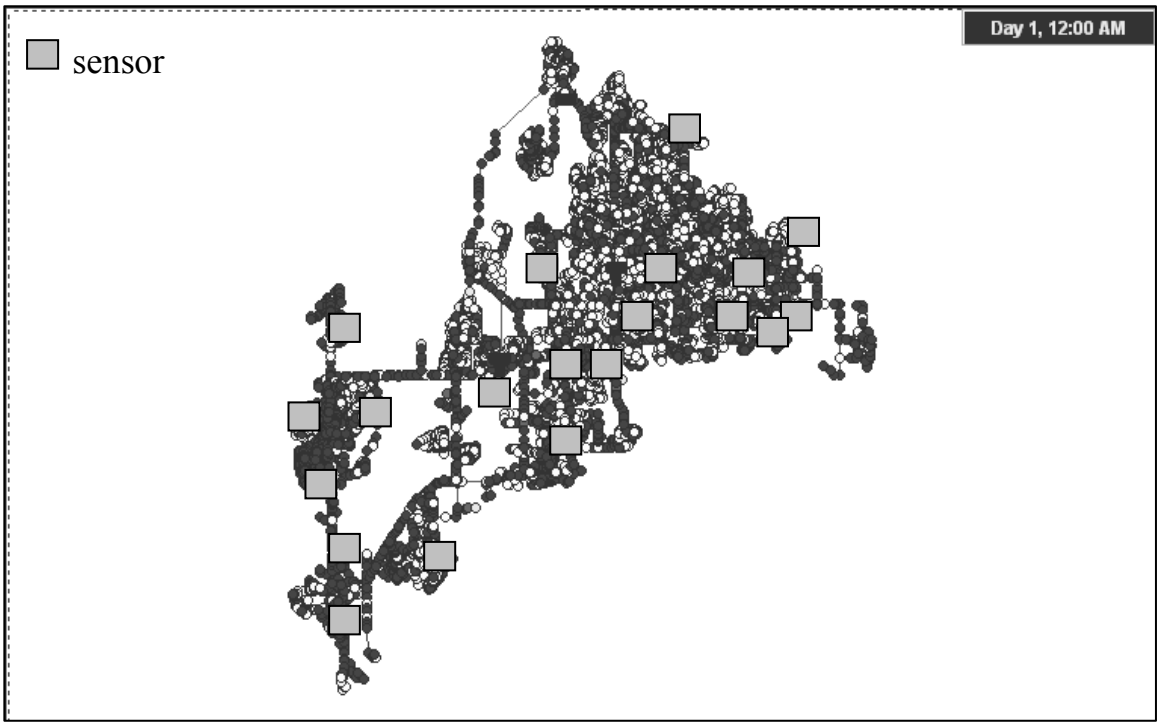


Figure 6.30 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a hazard threshold of 0.0 mg/L for BWSN Network 2.

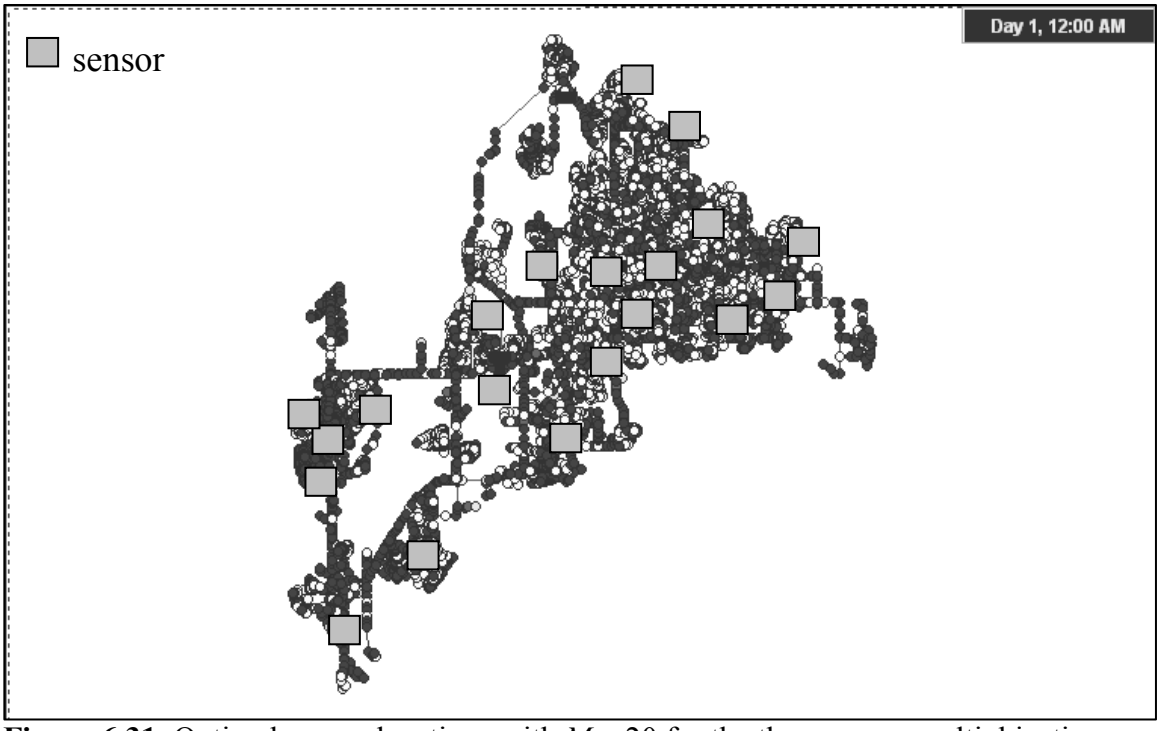


Figure 6.31 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of 2 nodes injected at 1 time for BWSN Network 2.

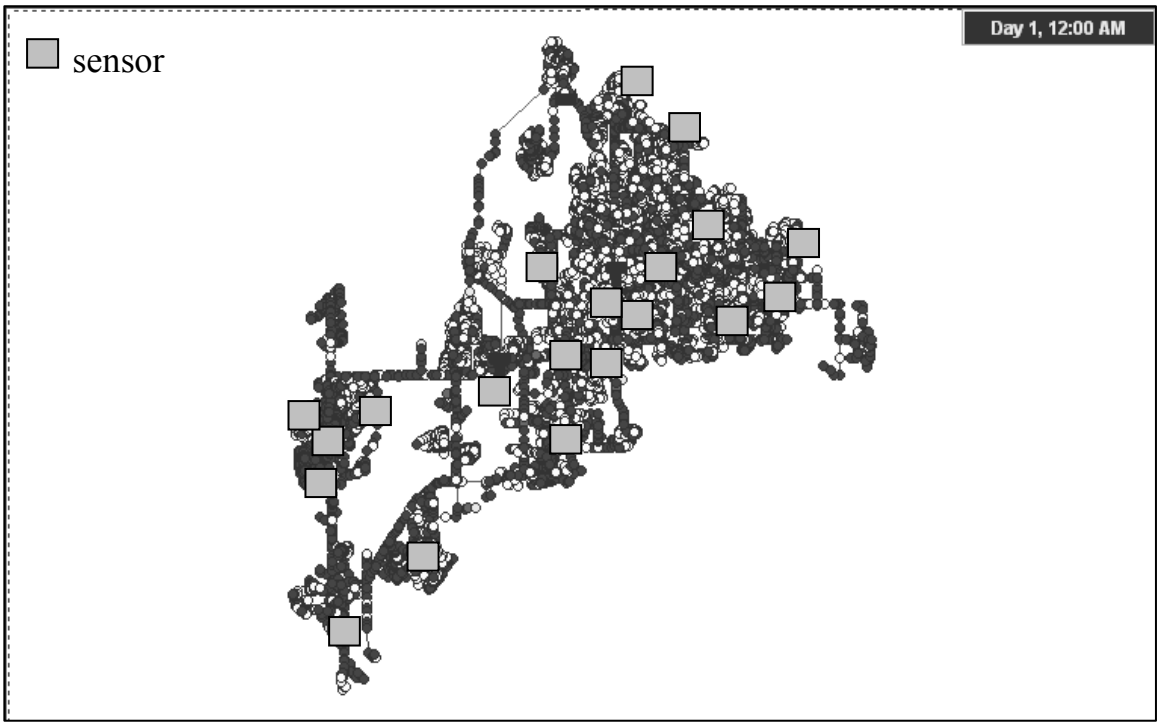


Figure 6.32 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of 2 nodes injected at 2 times for BWSN Network 2.

Table 6.23 Comparison of optimal sensor locations for the three-prong multiobjective problem with $M = 20$ in the variant attack cases for BWSN Network 2 determined in this study with locations found in other works.

Attack Case	Work	Sensor Nodes*	Z_{all}	Z_{lik}** (%)	Z_{time}** (min)	Z_{vol}** (gal)
contaminant injection duration = 10 h	This Study	495, 636, 850, 1486, 1514, 1806, 2602, 3229, 3357, 3573, 3770, 3836, 4247, 4306, 7441, 7664, 9022, 9364, 10716, 11687	2.60	44.7	3,535.9	64,757
	Dorini et al. (2006)	637, 1170, 1478, 1872, 2223, 2848, 3573, 4685, 4985, 5409, 6429, 7422, 8336, 8709, 9210, 9364, 10407, 10874, 11528, 12352	1.35	40.1	3,788.3	62,395
	Eliades and Polycarpou (2006)	532, 1426, 1486, 1976, 3231, 3679, 3836, 4234, 4359, 4609, 5087, 5585, 6922, 7670, 7858, 8629, 9360, 9787, 10885, 12167	***	41.4	3,850.0	105,212
	Krause et al. (2006)	1478, 1904, 2579, 3229, 3357, 3635, 3747, 3836, 4032, 4132, 4240, 4684, 7441, 8999, 9142, 9364, 10874, 10944, 11184, 11304	2.09	41.4	3,650.6	32,629
	Wu and Walski (2006)	623, 872, 1081, 1486, 2313, 3115, 3640, 3718, 4375, 4621, 4990, 5481, 7443, 7711, 8294, 8392, 9778, 10190, 10522, 10680	1.97	42.8	3,673.1	66,319
response delay = 3 h	This Study	551, 850, 1486, 1904, 1917, 2602, 3229, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8089, 8492, 9212, 9364, 10703, 10874	2.47	42.1	3,679.1	69,052
	Dorini et al. (2006)	637, 1428, 1478, 1999, 2527, 3328, 3573, 4684, 4985, 5391, 6837, 7638, 8336, 9187, 9196, 9677, 10208, 10874, 11528, 12352	1.50	36.8	3,908.1	44,311
	Eliades and Polycarpou (2006)	532, 1426, 1486, 3231, 3836, 4234, 4359, 4511, 4609, 5087, 5585, 6922, 7670, 7858, 8403, 8629, 9360, 9787, 10885, 12167	1.25	38.2	3,945.3	81,299
	Krause et al. (2006)	339, 636, 3357, 3629, 4032, 4132, 4240, 4684, 6583, 6700, 7336, 7652, 8999, 9142, 9722, 10614, 10874, 11167, 11304, 12258	1.25	34.3	3,961.6	24,535
	Wu and Walski (2006)	871, 1081, 2313, 3115, 3322, 3640, 3719, 3780, 4209, 4990, 5630, 6737, 7442, 7908, 8408, 9364, 9779, 10494, 10680, 12389	1.99	39.6	3,787.0	60,401
2 injection nodes at 1 time	This Study	636, 850, 1229, 1486, 1514, 2602, 3229, 3357, 3629, 3770, 3836, 4247, 4306, 7441, 7667, 8942, 9364, 10720, 11304, 11687	2.74	66.4	2,378.4	52,111
	Dorini et al. (2006)	636, 1170, 1478, 1729, 2241, 3573, 3770, 4594, 4985, 5387, 7422, 8272, 8336, 8714, 9030, 9364, 10186, 10874, 11528, 12377	1.89	61.5	2,625.3	41,527
	Eliades and Polycarpou (2006)	375, 532, 1426, 1486, 1976, 2865, 3235, 3679, 3836, 4234, 4359, 4609, 5087, 5585, 6922, 7670, 7858, 8629, 9787, 10670	1.76	63.6	2,681.3	89,043
	Huang et al. (2006)	73, 108, 630, 889, 2252, 2526, 2544, 3947, 4648, 5086, 5363, 6581, 8246, 8439, 8719, 9027, 9363, 9999, 10214, 10358	***	51.1	3,155.6	59,228
	Krause et al. (2006)	1478, 1904, 2579, 3229, 3357, 3635, 3770, 3836, 4032, 4132, 4240, 7441, 7668, 8999, 9142, 9364, 10874, 10944, 11184, 11304	2.38	63.1	2,482.8	25,586
	Wu and Walski (2006)	872, 1081, 1422, 2595, 3115, 3318, 3640, 3782, 4375, 4990, 5630, 6734, 7442, 7713, 8394, 9778, 10522, 10680, 11151, 12389	2.18	63.9	2,534.4	62,456

* All sensor node numbers preceded by "JUNCTION-".
** Values computed using the software developed in this work.
*** Solution falls outside of Pareto front space.

6.4.2 ANALYSIS OF PERFORMANCE TRADEOFFS

The quantified tradeoffs in protection with respect to individual protection goals made in employing multiobjective sensor node solutions for the various attack cases are indicated in Tables 6.24 and 6.25. As observed above and indicated by the performance cost values in the tables below, the solutions to maximize Z_{all} preserve a great deal of protection with regard to minimizing Z_{time} and make larger compromises in terms of the other two protection goals. One minor exception is seen in the attack case of the 0.0 mg/L contaminant hazard threshold, for which the compromises in terms of minimizing Z_{time} and minimizing Z_{vol} were both small but almost equal.

Table 6.24 Performance tradeoffs for individual protection goals resulting from maximizing Z_{all} with $M = 20$ in the default attack case for BWSN Network 2.

Protection Goal Performance Measure	Pareto Front Space Minimum Value	Pareto Front Space Maximum Value	Value for Maximizing Z_{all}	Protection Performance Cost
Z_{lik}	33.5%	43.5%*	42.1%	13.9%
Z_{time}	3,660.9 min*	4,015.0 min	3,679.1 min	5.1%
Z_{vol}	12,732 gal*	135,321 gal	43,952 gal	25.5%

* Value represents maximum protection with respect to the particular protection goal in question.

Table 6.25 Performance tradeoffs for individual protection goals resulting from maximizing Z_{all} with $M = 20$ in the variant attack cases for BWSN Network 2.

Attack Case	Protection Performance Cost, Z_{lik}	Protection Performance Cost, Z_{time}	Protection Performance Cost, Z_{vol}
default	13.9%	5.1%	25.5%
sensor detection resolution = 0.3 mg/L	13.3%	1.2%	28.4%
contaminant injection duration = 10 h	12.6%	5.9%	22.0%
contaminant hazard threshold = 0.0 mg/L	16.4%	9.8%	9.5%
response delay = 3 h	13.6%	5.1%	34.1%
2 injection nodes at 1 time	8.3%	2.1%	15.4%
2 injection nodes at 2 times	9.4%	0.3%	15.9%

6.4.3 EVALUATION OF NODAL IMPORTANCE

Tables 6.26 and 6.27 indicate the effectiveness of the multiobjective nodal importance approach for this large system by reporting the lowest-ranked nodes in solutions reported above and proportions of ranked WDS nodes that contain the optimal solutions. In the default attack case with 5 sensors employed, the optimal solution was isolated to just over 11% of WDS nodes using the multiobjective nodal importance approach. In the attack cases with 20 sensors employed, the proportion of nodes containing the optimal solution range between roughly 15% and 30%.

The solution convergence plots in Figures 6.33 and 6.34 show that reasonable progress toward finding the optimal solution occurs with each iteration of ISSM in most cases. One case that is peculiar in terms of rate of convergence is that of the 0.0 mg/L hazard

threshold for which many iterations of ISSM were required to achieve solutions close to optimal. On the whole, though, it appears that ISSM making use of the multiobjective nodal importance approach developed in this study provides well-performing tradeoff solutions for this large system and is able to narrow the search domain for those solutions to substantial degrees.

Table 6.26 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for the three-prong multiobjective problem in the default attack case and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	1,385	11.1%
20	2,602	20.8%

Table 6.27 Nodal importance ranks for the lowest-ranked BWSN Network 2 nodes of optimal solutions for the three-prong multiobjective problem with $M = 20$ in the variant attack cases and the proportions of ranked BWSN Network 2 nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	2,602	20.8%
sensor detection resolution = 0.3 mg/L	1,814	14.5%
contaminant injection duration = 10 h	3,496	27.9%
contaminant hazard threshold = 0.0 mg/L	3,260	26.0%
response delay = 3 h	2,602	20.8%
2 injection nodes at 1 time	3,642	29.1%
2 injection nodes at 2 times	2,942	23.5%

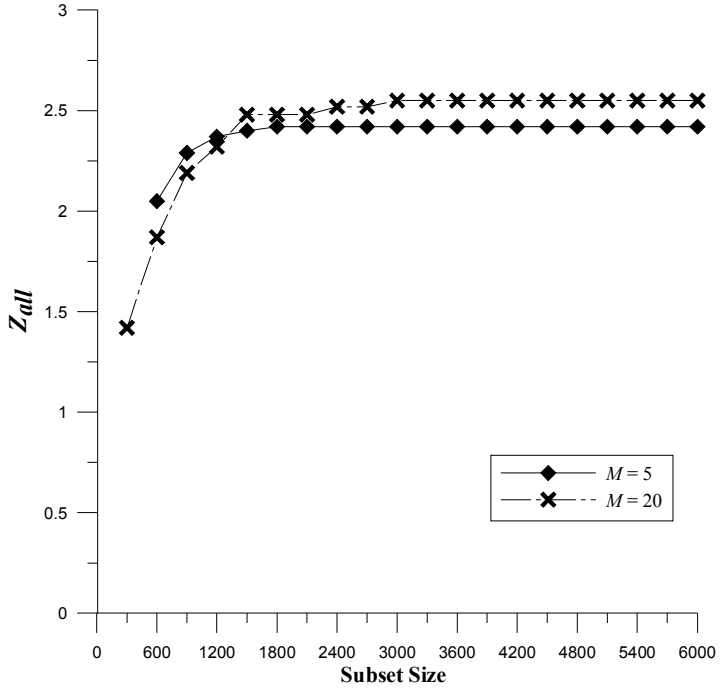


Figure 6.33 Solution convergence as ISSM is carried out for the three-prong multiobjective problem in the default attack case for BWSN Network 2. (Only solution values for subset sizes of 6,000 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 6,000; negative Z_{all} values are not plotted.)

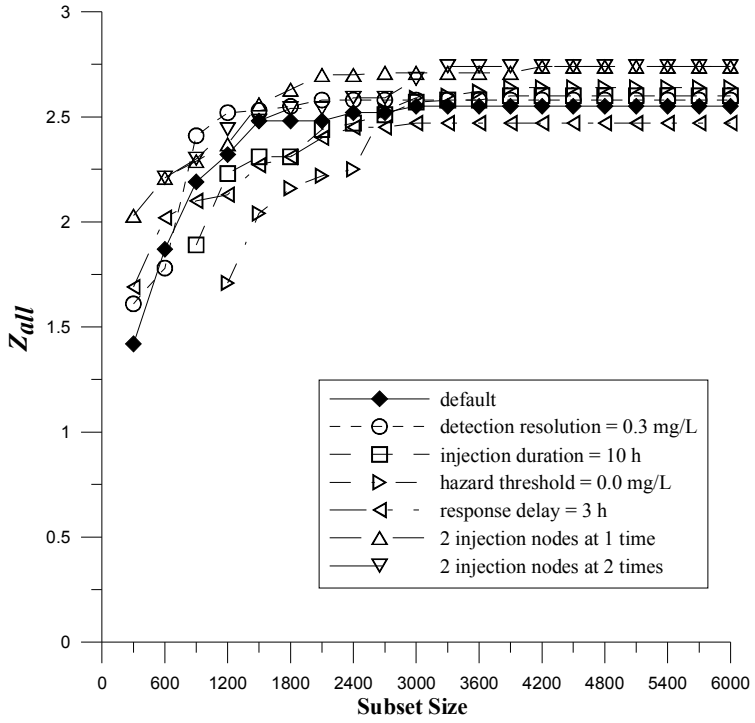


Figure 6.34 Solution convergence as ISSM is carried out for the three-prong multiobjective problem with $M = 20$ in the default and variant attack cases for BWSN Network 2. (Only solution values for subset sizes of 6,000 or less are plotted as solutions for all cases reached best performance by subset size = 6,000; negative Z_{all} values are not plotted.)

6.5 COMPUTATIONAL EXPENSE NOTES

As expected, the computational resources required in sensor placement optimization for BWSN Network 2 are significantly greater than those required for BWSN Network 1. Scenario data generation took approximately 18 hours and produced about 1 GB of data for the importance and optimization phases. The importance and optimization cases for all single-objective problems and the multiobjective problem together took roughly between 3 and 5 hours to carry out under the ISSM stopping criteria specified for this study; longer runtimes were associated with cases requiring broader searches for optimal solutions. Again, these times may have been shorter if a different optimization solver had been used. The computational demands for this study seem very manageable, though, given the memory required (e.g., 15 to 30 GB for Krause et al. (2006)) and runtimes required (e.g., multiple days for Berry et al. (2006c) and Krause et al. (2006)) for the methods of other works documented.

CHAPTER 7

RESULTS AND ANALYSIS: TOMS RIVER WDS

Results of sensor placement optimization for the Toms River WDS are presented in this chapter. The results are analyzed independently as well as with respect to the other two study systems. Comparisons between the results for the Toms River WDS and those for BWSN Network 2 are especially important in order to validate the optimization methods of this study for use in monitoring large systems of most interest. Sensor placement optimization has not been performed for this system prior to this study, so there are no other results from other works with which to compare to the results of this study.

7.1 SINGLE-OBJECTIVE PROBLEM: MAXIMIZING Z_{lik}

The sensor placement results for the maximizing of detection likelihood for the Toms River WDS are given and discussed below.

7.1.1 SENSOR NODE SOLUTIONS

Table 7.1 presents the sensor node solutions for maximizing detection likelihood in the default attack case for the Toms River WDS, and Figures 7.1 and 7.2 indicate where the nodes are located in the system. As indicated by the data reported in Table 7.1, the optimization method of this study seems very effective in maximizing detection likelihood. For instance, with only 5 sensors employed, contamination is detected for

over half of the attack scenarios (i.e., 51.8%) given the sensor locations listed in the table. These sensors apparently provide the major coverage for the system. With 20 sensors employed, local neighborhoods of nodes are able to be covered, allowing for a higher detection likelihood. In comparing these results with the counterpart results for BWSN Network 2, it is observed that Z_{lik} values are higher for the Toms River WDS even though the system has more nodes than does BWSN Network 2 and is of comparable size to that of BWSN Network 2, giving some measure of validation that the results in Table 7.1 are indeed good. Figures 7.1 and 7.2 show the general spread of sensor locations toward the periphery of the system where many hydraulic pathways end; this behavior in sensor placement was seen for maximizing detection likelihood for both BWSN Networks 1 and 2. It is also noted that the Z_{vol} values in Table 7.1 are much higher than those in Table 6.1 for BWSN Network 2, indicating that the potential volume that could be contaminated for this system is much higher than that for BWSN Network 2.

Table 7.2 gives the sensor node solutions for the variant attack cases. In short, the sensor node locations do not vary greatly from attack case to attack case, and Z_{lik} values vary in the expected manners discussed previously for BWSN Networks 1 and 2. Figures 7.3 through 7.6 depict the sensor node locations for the variant attack cases.

Table 7.1 Optimal sensor locations and performance measures for maximizing detection likelihood in the default attack case for the Toms River WDS.

M	Sensor Nodes	Z_{lik}^{*} (%)	Z_{time} (min)	Z_{vol} (gal)
5	1091, 9669, 10681, 12301, 13107	51.8	4,035.3	548,469
20	1091, 1920, 2959, 3715, 5005, 5359, 7722, 8119, 8180, 8421, 9669, 10681, 10711, 10776, 11753, 13289, 14107, 14536, 14712, 15111	59.4	3,367.0	227,484

* Performance measure for protection goal of interest.

Figure 7.1 Optimal sensor locations with $M = 5$ for maximizing detection likelihood in the default attack case for the Toms River WDS.

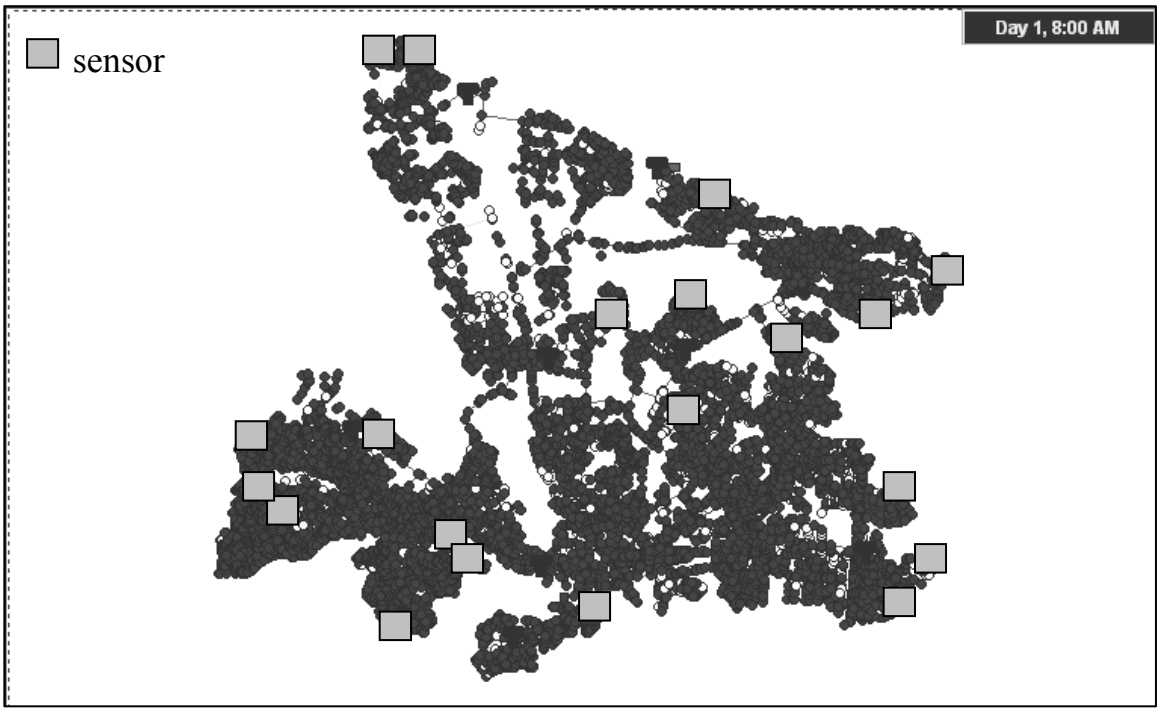


Figure 7.2 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the default attack case for the Toms River WDS.

Table 7.2 Optimal sensor locations and performance measures for maximizing detection likelihood with $M = 20$ in the variant attack cases for the Toms River WDS.

Attack Case	Sensor Nodes	Z_{lik}* (%)	Z_{time} (min)	Z_{vol} (gal)
default	1091, 1920, 2959, 3715, 5005, 5359, 7722, 8119, 8180, 8421, 9669, 10681, 10711, 10776, 11753, 13289, 14107, 14536, 14712, 15111	59.4	3367.0	227,484
sensor detection resolution = 0.3 mg/L	1091, 1799, 1920, 3715, 4854, 5359, 6759, 7722, 7924, 8588, 9144, 9451, 9669, 10477, 10711, 10776, 11753, 13289, 14536, 15111	57.7	3340.4	176,052
contaminant injection duration = 10 h	56, 1091, 1920, 2959, 3715, 5359, 7722, 8119, 8180, 8421, 9669, 10711, 10776, 11231, 11753, 13289, 14107, 14536, 14712, 15111	60.8	3256.1	291,195
2 injection nodes at 1 time	56, 1091, 1920, 3662, 5005, 5359, 7722, 8094, 8119, 8421, 9464, 10711, 10776, 11753, 13289, 13335, 14536, 14712, 14860, 15111	82.5	2158.9	257,169
2 injection nodes at 2 times	56, 1091, 5005, 5359, 7722, 8094, 8104, 8119, 8421, 9289, 9464, 10711, 10776, 11753, 13289, 13335, 14536, 14712, 14860, 15111	83.4	2333.9	263,217

* Performance measure for protection goal of interest.

Figure 7.3 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of a detection resolution of 0.3 mg/L for the Toms River WDS.

Figure 7.4 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of a 10-hour injection duration for the Toms River WDS.

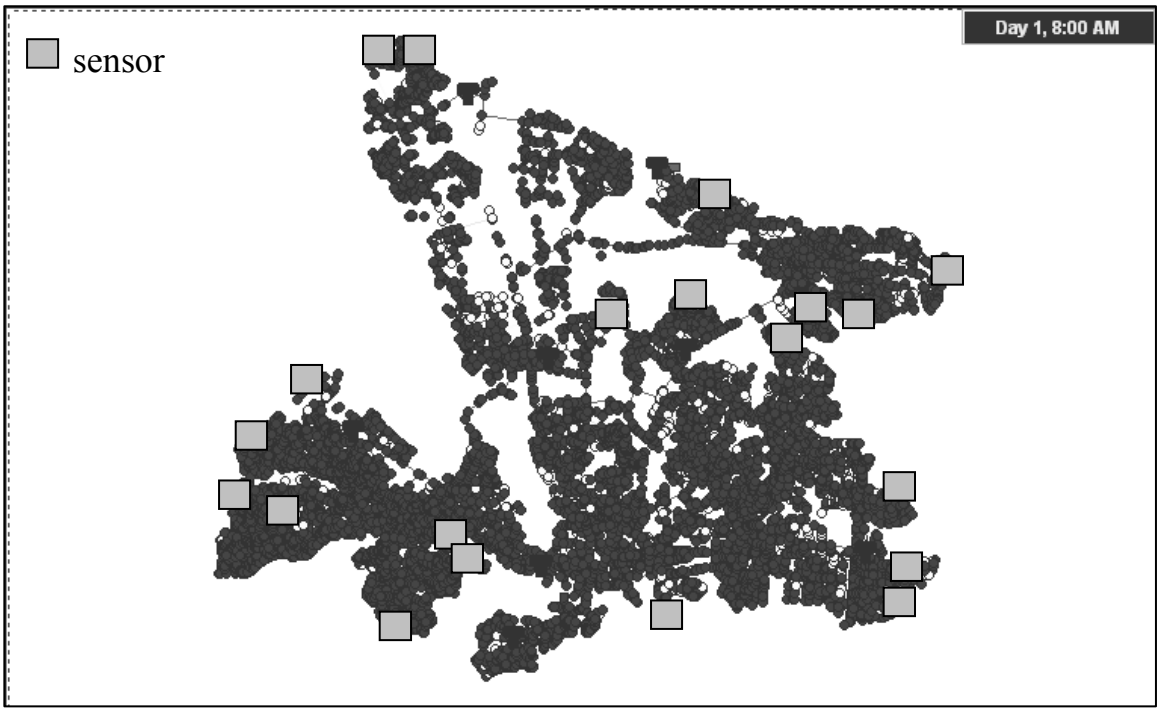


Figure 7.5 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of 2 nodes injected at 1 time for the Toms River WDS.

Figure 7.6 Optimal sensor locations with $M = 20$ for maximizing detection likelihood in the case of 2 nodes injected at 2 times for the Toms River WDS.

7.1.2 EVALUATION OF NODAL IMPORTANCE

To obtain the solutions above for all cases where a zero sensor detection resolution is assumed, ISSM was carried out making use of nodal ranking done by end-of-line node status first and nodal importance value found by (4.21) second, as was done for BWSN Network 2. ISSM parameters used were those also used for BWSN Network 2 (Table 6.5). The Toms River WDS has 1,998 end-of-line nodes—enough end-of-line nodes to allow for many possible contaminant pathways to be covered without considering slow-outflow nodes in the sorting procedure. Of course, nodes are ranked according to nodal importance value only in the case of a 0.3 mg/L detection resolution.

The nodal importance approach seems to be as effective for this system in narrowing the search for the optimal nodes to maximize detection likelihood as it was for the other two study systems. The optimal sensor nodes in the default attack case with 5 sensors employed are able to be isolated to less than 3% of WDS nodes, given the data reported in Table 7.3. Tables 7.3 and 7.4 indicate that the search can be narrowed to approximately 5 to 8% of WDS nodes when 20 sensors are employed in all attack cases studied. The effect of the higher detection resolution on nodal importance is more apparent for this system. A broader proportion of WDS nodes is needed to contain the optimal solution in this case because nodes cannot be stratified by end-of-line node status, as implied by the proportion for this case reported in Table 7.4.

The solution convergence plots in Figures 7.7 and 7.8 indicate that near-optimal solutions in all cases can be found with ISSM with one or two iterations, and only a few more iterations beyond the initial iteration can yield the optimal solution. In sum, the nodal importance concepts for maximizing detection likelihood for a large system appear to be very effective in finding very well-performing solutions and narrowing the search for these solutions substantially for the Toms River WDS.

Table 7.3 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for maximizing detection likelihood in the default attack case and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	431	2.8%
20	780	5.2%

Table 7.4 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for maximizing detection likelihood with $M = 20$ in the variant attack cases and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	780	5.2%
sensor detection resolution = 0.3 mg/L	1,153	7.7%
contaminant injection duration = 10 h	816	5.4%
2 injection nodes at 1 time	998	6.7%
2 injection nodes at 2 times	932	6.2%

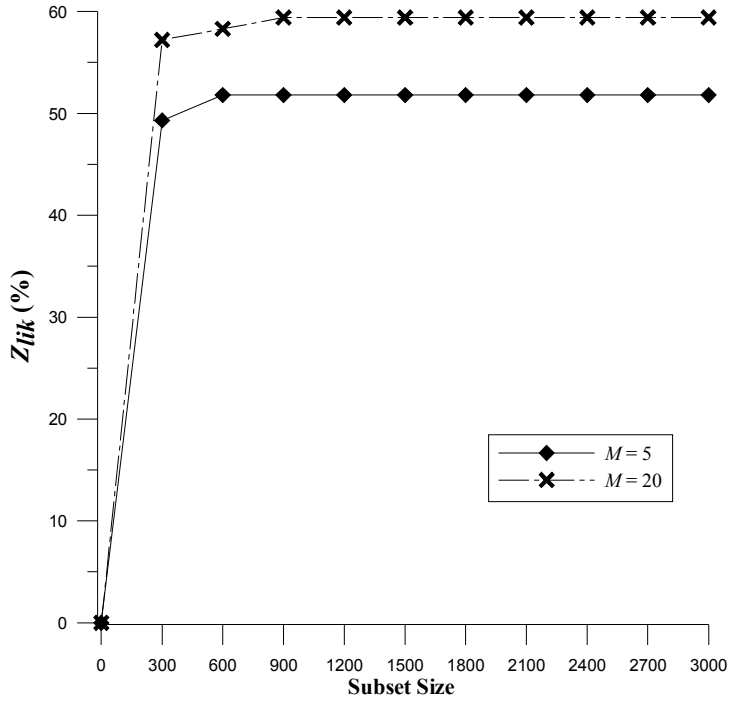


Figure 7.7 Solution convergence as ISSM is carried out for maximizing detection likelihood in the default attack case for the Toms River WDS. (Only solution values for subset sizes of 3,000 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 3,000.)

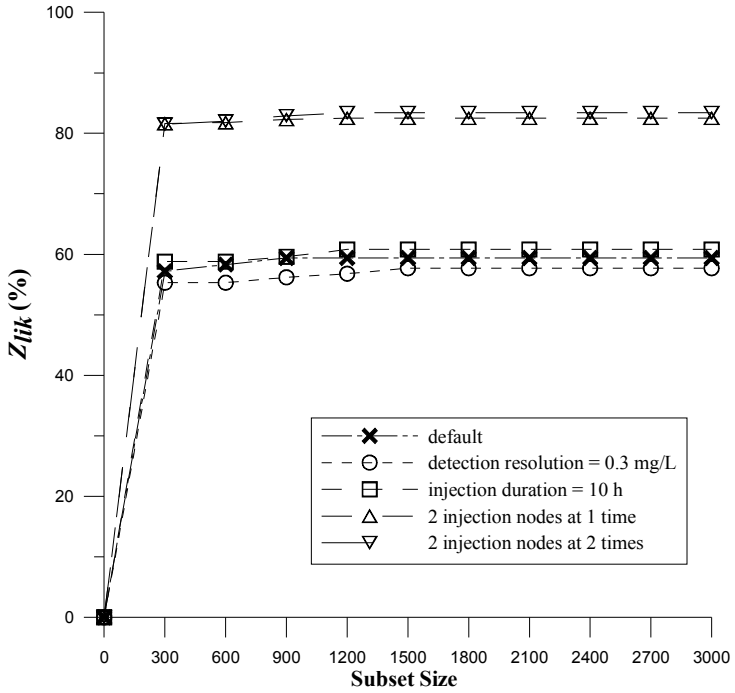


Figure 7.8 Solution convergence as ISSM is carried out for maximizing detection likelihood with $M = 20$ in the default and variant attack cases for the Toms River WDS. (Only solution values for subset sizes of 3,000 or less are plotted as solutions for all cases reached best performance by subset size = 3,000.)

7.2 SINGLE-OBJECTIVE PROBLEM: MINIMIZING Z_{time}

Optimization results with respect to minimizing expected detection time are given and discussed below.

7.2.1 SENSOR NODE SOLUTIONS

Sensor nodes selected for minimizing expected detection time in the default attack time for the Toms River WDS are presented in Table 7.5. It seems that the optimization approach of this work substantially reduced Z_{time} values for both $M = 5$ and $M = 20$ based on a comparison of Z_{time} values in Table 7.1 versus those in Table 7.5. The results in Table 7.5 also seem desirable when compared to the results for minimizing expected detection time for BWSN Network 2, given that the Z_{time} values in Table 7.5 are less than their counterparts in Table 6.8. As observed in the study of the other two study systems, there is only a small cost in detection likelihood when minimizing expected detection time for the Toms River WDS, and minimizing Z_{time} reduces Z_{vol} significantly relative to Z_{vol} values found when maximizing detection likelihood.

Figures 7.9 and 7.10 show the shifts of some sensor nodes from the periphery of the system where contaminant pathway endpoints are located toward the center of the system (i.e., upstream from contaminant pathway endpoints). The shift is most apparent in Figure 7.9 for the 5-sensor case; one sensor is needed much closer to the center of the system to detect contamination sooner as opposed to the sensor node layout in Figure 7.1 that has all 5 sensors covering contaminant pathway endpoints. It was observed in

Chapter 6 that sensor nodes selected for minimizing expected detection time for BWSN Network 2 often were nodes that intercepted flow from multiple contaminant pathways; that placement pattern is seen for this system as well with such sensor nodes as Junction 382 and Junction 10130 that receive flows from multiple contaminant pathways.

Table 7.6 gives the sensor nodes selected for the applicable variant attack cases, and Figures 7.11 through 7.14 depict their locations in the system. As for the other two study systems, there are not significant adjustments in sensor node locations from attack case to attack case. Also, Z_{time} varies in the same manner under different attack conditions relative to the default attack case as it did for the other two systems.

Table 7.5 Optimal sensor locations and performance measures for minimizing expected detection time in the default attack case for the Toms River WDS.

M	**Sensor Nodes**	Z_{lik} **(%)**	Z_{time}^{*} **(min)**	Z_{vol} **(gal)**
5	382, 5081, 6333, 10130, 10682	49.6	3,428.3	102,922
20	382, 2678, 3740, 4399, 5081, 5593, 6333, 7688, 7710, 7920, 9215, 9660, 9806, 10130, 10412, 10682, 12926, 14202, 14496, 14746	54.5	3,062.9	40,841

* Performance measure for protection goal of interest.

Figure 7.9 Optimal sensor locations with $M = 5$ for minimizing expected detection time in the default attack case for the Toms River WDS.

Figure 7.10 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the default attack case for the Toms River WDS.

Table 7.6 Optimal sensor locations and performance measures for minimizing expected detection time with $M = 20$ in the variant attack cases for the Toms River WDS.

Attack Case	Sensor Nodes	Z_{lik} (%)	Z_{time}* (min)	Z_{vol} (gal)
default	382, 2678, 3740, 4399, 5081, 5593, 6333, 7688, 7710, 7920, 9215, 9660, 9806, 10130, 10412, 10682, 12926, 14202, 14496, 14746	54.5	3,062.9	40,841
sensor detection resolution = 0.3 mg/L	382, 3506, 3740, 5103, 5576, 6333, 6961, 7710, 7920, 9215, 9660, 10130, 10412, 10682, 12051, 12683, 14202, 14536, 14746, 14894	52.8	3,176.5	56,236
contaminant injection duration = 10 h	382, 2265, 3506, 3740, 4340, 5511, 6321, 7688, 7710, 7920, 8259, 9215, 9660, 9806, 10130, 10412, 10682, 13114, 14490, 14746	55.9	2,987.2	71,375
2 injection nodes at 1 time	382, 414, 2673, 3740, 5223, 5593, 6376, 7710, 7748, 7920, 8259, 9215, 9660, 9806, 10130, 10412, 12271, 12683, 14490, 14746	78.1	1,886.5	87,327
2 injection nodes at 2 times	382, 2001, 3740, 3991, 5196, 5566, 5917, 6321, 7710, 7755, 7920, 8469, 9215, 9660, 9806, 10130, 10412, 10677, 14496, 14746	79.1	1,982.0	63,540

* Performance measure for protection goal of interest.

Figure 7.11 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of a detection resolution of 0.3 mg/L for the Toms River WDS.

Figure 7.12 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of a 10-hour injection duration for the Toms River WDS.

Figure 7.13 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of 2 nodes injected at 1 time for the Toms River WDS.

Figure 7.14 Optimal sensor locations with $M = 20$ for minimizing expected detection time in the case of 2 nodes injected at 2 times for the Toms River WDS.

7.2.2 EVALUATION OF NODAL IMPORTANCE

The sensor nodes solutions above were found in the same way solutions for minimizing Z_{time} were found for the other two study systems: by conducting ISSM aided by the rankings resulting from the nodal importance function in (4.23). Tables 7.7 and 7.8 provide the ranks of the lowest ranked nodes for solutions found for minimizing expected detection time in the applicable attack cases. The effectiveness of nodal importance in narrowing the search domain for this protection goal for the Toms River WDS is mostly comparable to that of BWSN Network 2; the optimal solution is contained to proportions in the 10 to 30% neighborhood of WDS nodes for all relevant attack cases. The lowest ranks for $M = 5$ and $M = 20$ in the default attack case given in Table 7.7 are equal to each other since solutions for those two cases share the same lowest-ranked sensor node.

The phenomenon of broader proportions of WDS nodes being required for the case of a 10-hour injection duration and the two injection node cases to contain the optimal solution observed in optimization sensor placement for BWSN Network 2 is seen for this system as well. It is suspected that this phenomenon occurred for the reasons discussed in Chapter 6. In short, nodes become less distinguishable from each other in terms of importance with increased opportunities for WDS nodes to experience contamination.

Convergence to very good solutions occurs rapidly for this system for all cases according to the plots in Figures 7.15 and 7.16, which is in contrast to the two injection node cases for BWSN Network 2 that are associated with much more gradual convergence rates.

Overall, though, the nodal importance concepts for minimizing expected detection time seem to be as effective in narrowing the search domain to find optimal sensor node solutions for the Toms River WDS as they are for BWSN Network 2.

Table 7.7 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for minimizing expected detection time in the default attack case and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	1,966	13.1%
20	1,966	13.1%

Table 7.8 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for minimizing expected detection time with $M = 20$ in the variant attack cases and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	1,966	13.1%
sensor detection resolution = 0.3 mg/L	2,814	18.8%
contaminant injection duration = 10 h	3,267	21.8%
2 injection nodes at 1 time	3,762	25.1%
2 injection nodes at 2 times	3,874	25.9%

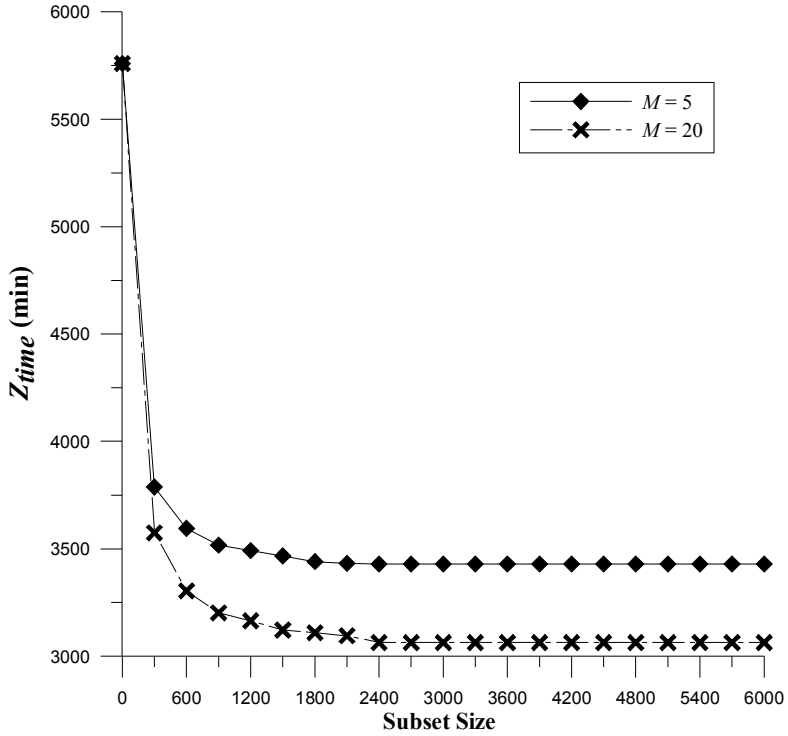


Figure 7.15 Solution convergence as ISSM is carried out for minimizing expected detection time in the default attack case for the Toms River WDS. (Only solution values for subset sizes of 6,000 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 6,000.)

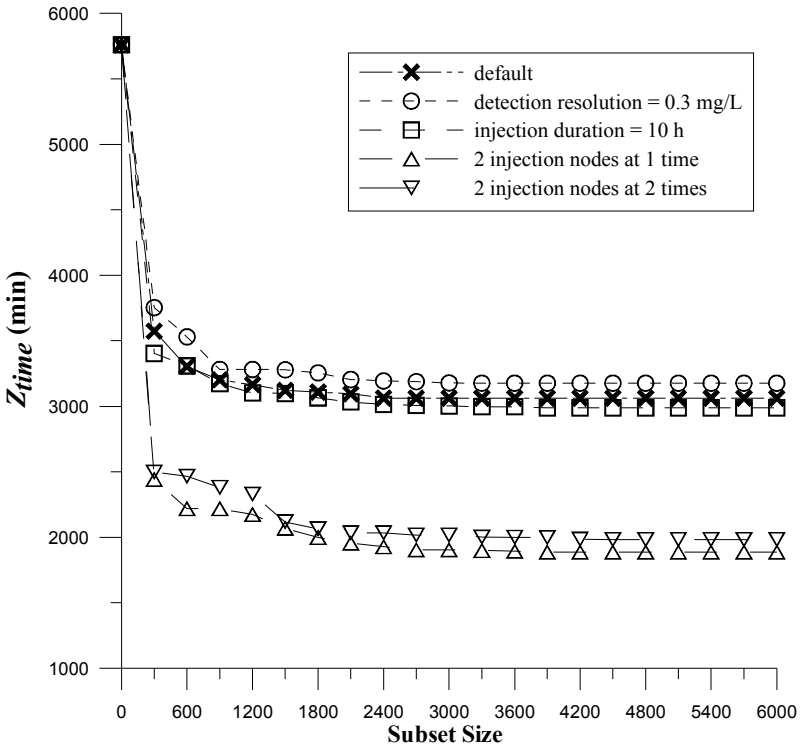


Figure 7.16 Solution convergence as ISSM is carried out for minimizing expected detection time with $M = 20$ in the default and variant attack cases for the Toms River WDS. (Only solution values for subset sizes of 6,000 or less are plotted as solutions for all cases reached best performance by subset size = 6,000.)

7.3 SINGLE-OBJECTIVE PROBLEM: MINIMIZING Z_{vol}

The results of sensor placement with regard to minimizing expected contaminated demand volume are presented and discussed here.

7.3.1 SENSOR NODE SOLUTIONS

Table 7.9 gives the sensor placement results for minimizing expected contaminated demand volume in the default attack case for the Toms River WDS. Considering the Z_{vol} values in Table 7.9 versus those in Tables 7.1 and 7.5 above, it seems that the optimization method for this protection goal reduced Z_{vol} significantly. In fact, the Z_{vol} value for the 5-sensor case is substantially less than its counterpart for BWSN Network 2, supporting that the sensor node solution for this system for the 5-sensor case is indeed good. For the 20-sensor case, the Z_{vol} value is just marginally more than its counterpart for BWSN Network 2; this difference is not surprising given the very large amount of volume that can potentially be contaminated in this system as indicated by Z_{vol} values in Table 7.1. Again, the results for minimizing expected contaminated demand volume for BWSN Network 2 for this work were shown to be slightly suboptimal in Chapter 6, so it is possible the results in Table 7.9 are suboptimal. However, it is likely that the results for this system are very well-performing given that the results for BWSN Network 2 were very good though not truly optimal.

Figures 7.17 and 7.18 depict the locations of sensor nodes in the system for minimizing expected contaminated demand volume in the default attack case. As observed in the

sensor placement optimization for the other two study systems, a noticeable shift of sensor node locations toward the center of the system (i.e., more upstream) occurs for this protection goal relative to locations used for maximizing Z_{lik} and minimizing Z_{time}. Many of the sensor nodes fall on major transmission lines, which was also the case for BWSN Network 2. Due to the need to detect contamination sooner to minimize Z_{vol}, it is not surprising that some sensor nodes are close in proximity to or even match (e.g., Junction 10130) those selected for minimizing expected detection time.

Table 7.9 Optimal sensor locations and performance measures for minimizing expected contaminated demand volume in the default attack case for the Toms River WDS.

M	**Sensor Nodes**	Z_{lik} **(%)**	Z_{time} **(min)**	Z_{vol}* **(gal)**
5	1796, 4819, 5822, 7287, 9864	41.9	3,653.9	44,258
20	571, 953, 1796, 2075, 4275, 4741, 5132, 5207, 5825, 5875, 6439, 6584, 6833, 7375, 8375, 8833, 9174, 9425, 9495, 10130	45.9	3,301.9	14,525

* Performance measure for protection goal of interest.

Figure 7.17 Optimal sensor locations with $M = 5$ for minimizing expected contaminated demand volume in the default attack case for the Toms River WDS.

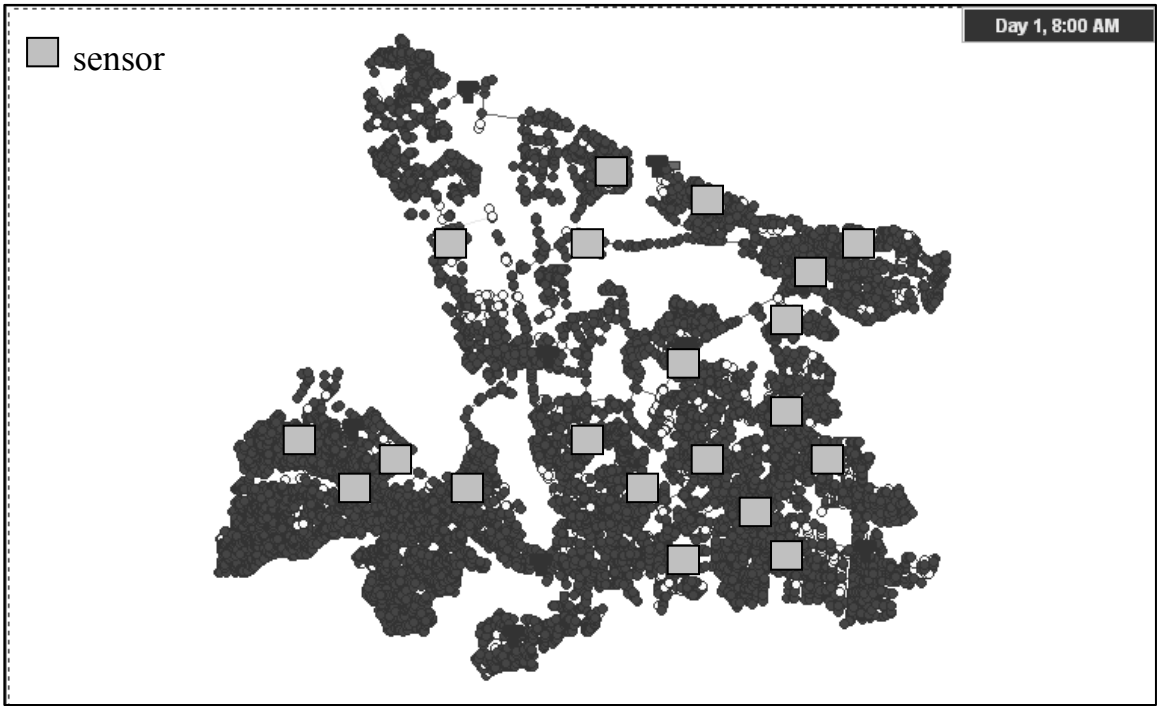


Figure 7.18 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the default attack case for the Toms River WDS.

Table 7.10 provides the sensor nodes selected for the variant attack cases, and the locations of the nodes are indicated in Figures 7.19 through 7.23. Sensor node locations are the same for the two injection nodes cases, so only sensor placement in the case of 2 nodes injected at 1 time is illustrated (Figure 7.23). Z_{vol} values vary relative to the default attack case in the manners observed in sensor placement optimization for BWSN Network 2 for all variant attack cases. As documented repeatedly for other protection goals and study systems, sensor node locations do not change to a large degree from attack case to attack case.

Table 7.10 Optimal sensor locations and performance measures for minimizing expected contaminated demand volume with $M = 20$ in the variant attack cases for the Toms River WDS.

Attack Case	Sensor Nodes	Z_{lik} (%)	Z_{time} (min)	Z_{vol}* (gal)
default	571, 953, 1796, 2075, 4275, 4741, 5132, 5207, 5825, 5875, 6439, 6584, 6833, 7375, 8375, 8833, 9174, 9425, 9495, 10130	45.9	3,301.7	14,525
sensor detection resolution = 0.3 mg/L	571, 953, 1796, 2075, 4275, 4741, 5132, 5199, 5825, 5875, 6439, 6584, 7064, 7375, 8375, 8833, 9174, 9425, 9495, 10130	43.9	3,394.4	16,522
contaminant injection duration = 10 h	571, 953, 2075, 3713, 4275, 4741, 5132, 5207, 5825, 5875, 6439, 6584, 6833, 7375, 8375, 8833, 9174, 9425, 9495, 10130	46.3	3,281.3	23,053
contaminant hazard threshold = 0.0 mg/L	571, 953, 1796, 2075, 4275, 4741, 5132, 5207, 5825, 5875, 6439, 6584, 6905, 7375, 8375, 8833, 9219, 9425, 9495, 10130	46.1	3,289.8	32,056
response delay = 3 h	571, 953, 1796, 2075, 4284, 4741, 5132, 5207, 5825, 5875, 6458, 6584, 6833, 7375, 8375, 8833, 9219, 9425, 9495, 10130	45.9	3,296.7	22,484
2 injection nodes at 1 time	571, 953, 1796, 2075, 4275, 4741, 5132, 5199, 5825, 5875, 6439, 6584, 6833, 7375, 8375, 8833, 9174, 9425, 9495, 10130	70.0	1,977.0	16,979
2 injection nodes at 2 times	571, 953, 1796, 2075, 4275, 4741, 5132, 5199, 5825, 5875, 6439, 6584, 6833, 7375, 8375, 8833, 9174, 9425, 9495, 10130	69.8	2,146.2	19,635

* Performance measure for protection goal of interest.

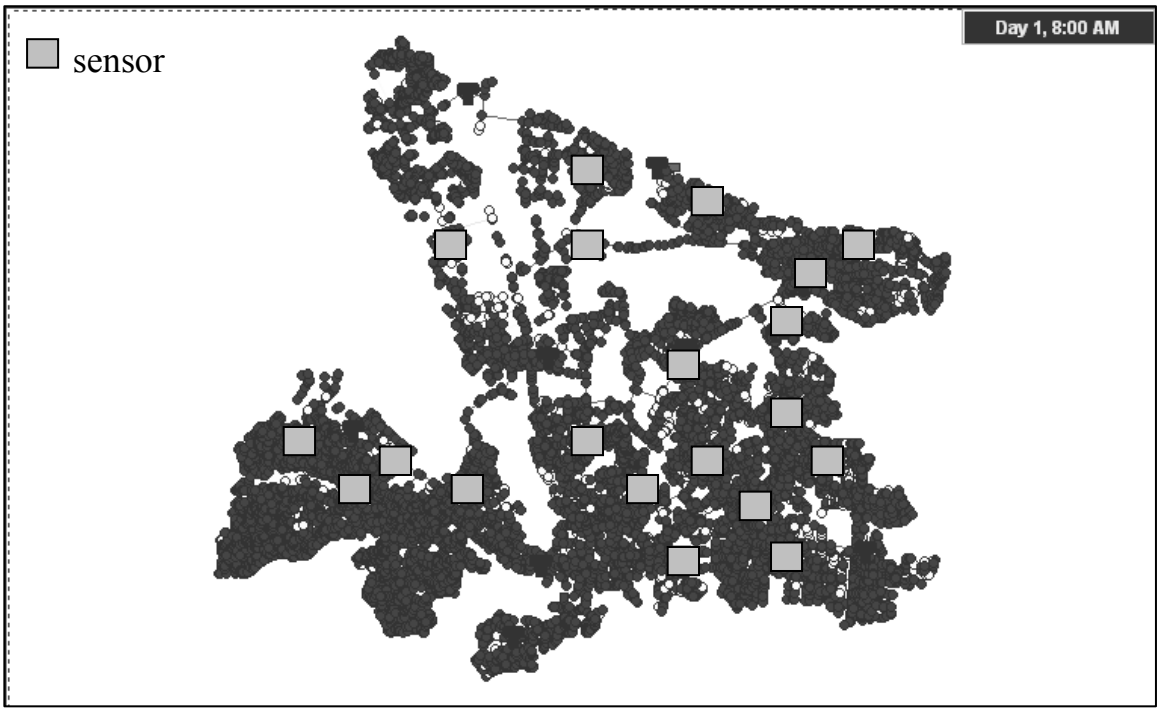


Figure 7.19 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a detection resolution of 0.3 mg/L for the Toms River WDS.

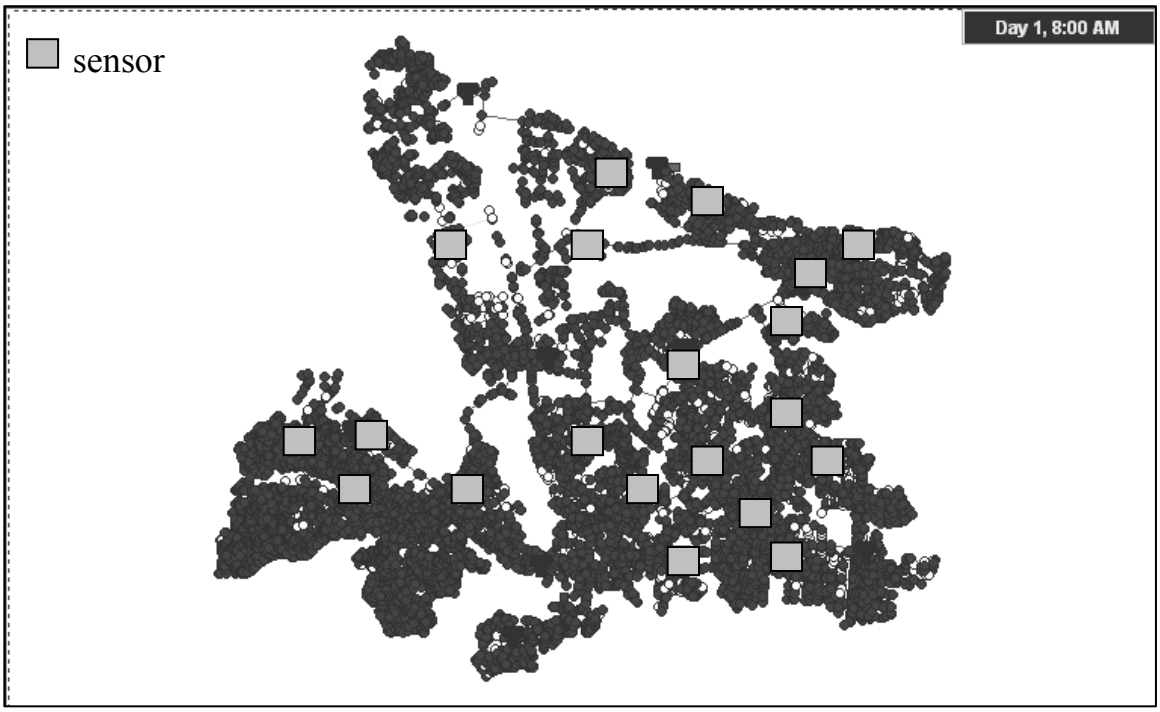


Figure 7.20 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a 10-hour injection duration for the Toms River WDS.

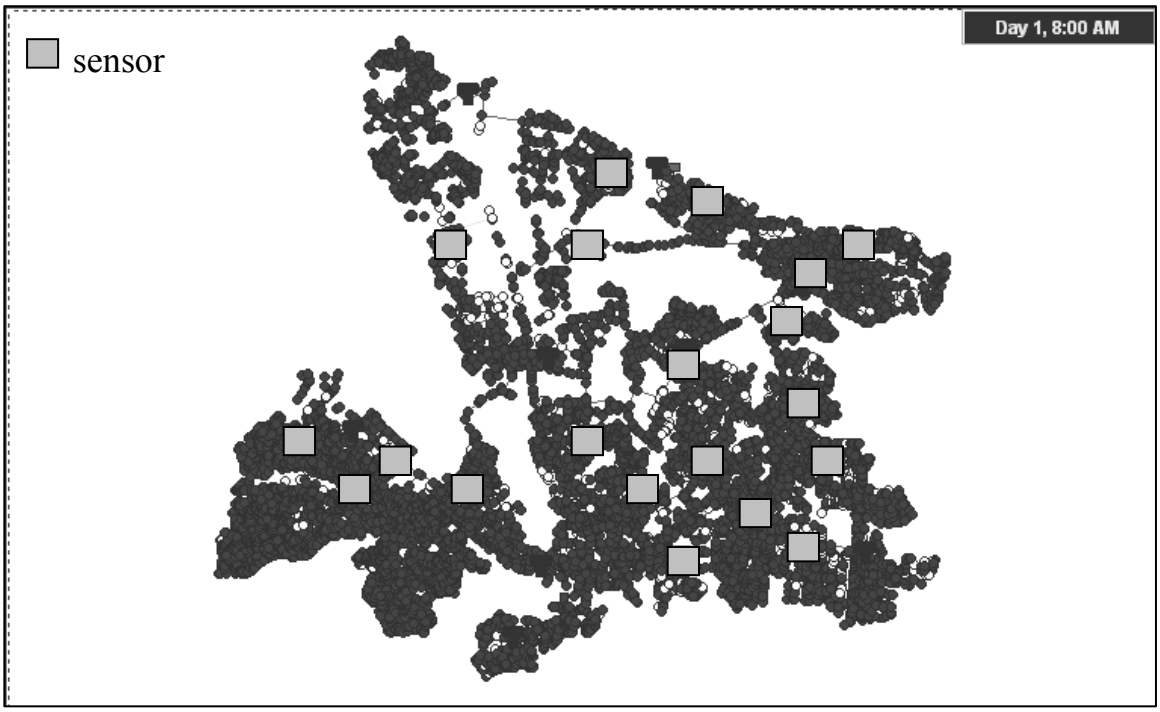


Figure 7.21 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a hazard threshold of 0.0 mg/L for the Toms River WDS.

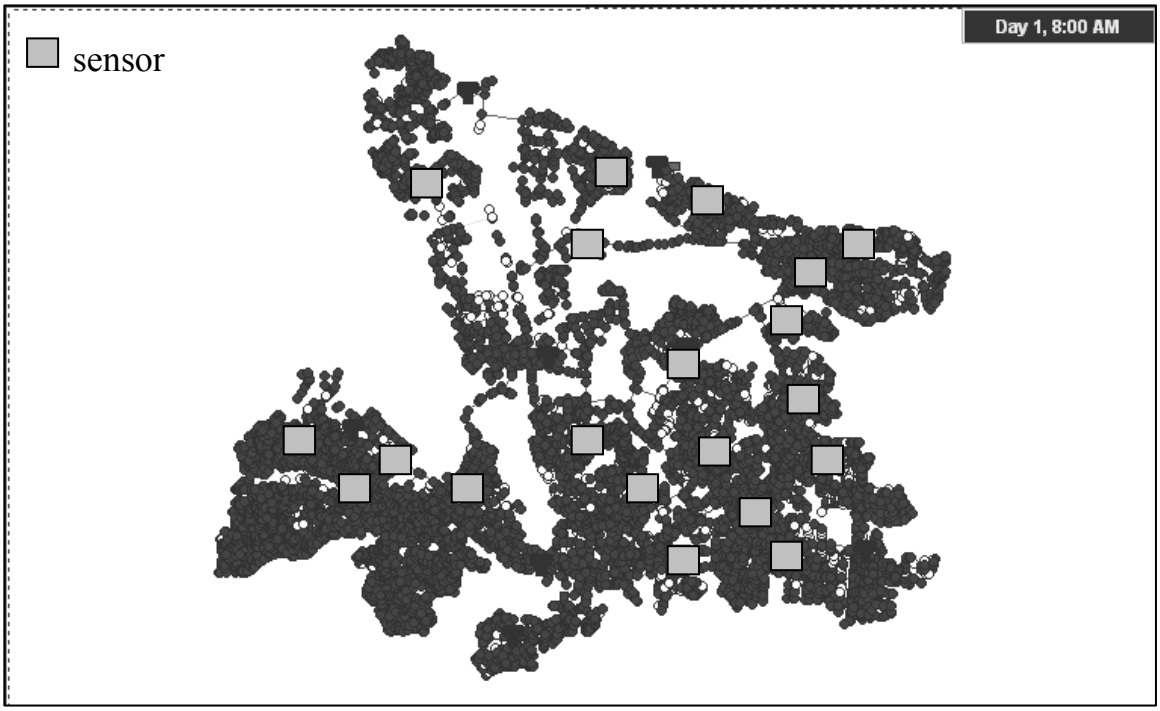


Figure 7.22 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of a 3-hour response delay for the Toms River WDS.

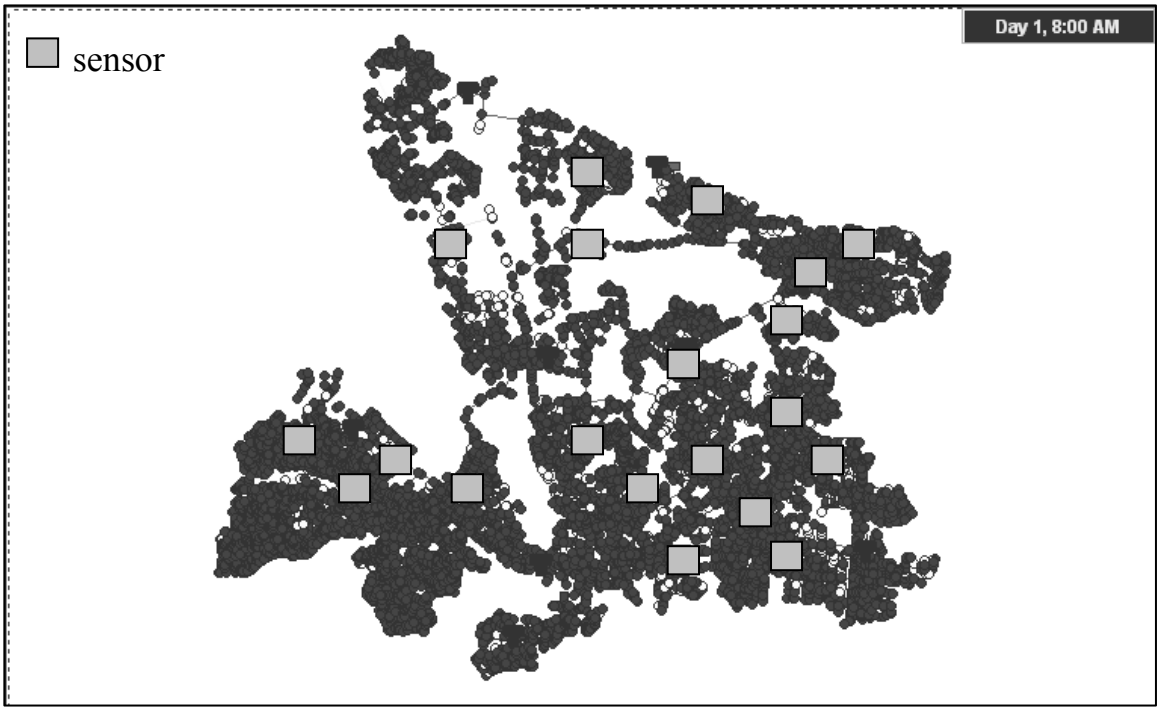


Figure 7.23 Optimal sensor locations with $M = 20$ for minimizing expected contaminated demand volume in the case of 2 nodes injected at 1 time for the Toms River WDS.

7.3.2 EVALUATION OF NODAL IMPORTANCE

Solutions above were found by conducting ISSM using nodal importance values found with the large-system nodal importance function in (4.27). Given the data in Tables 7.11 and 7.12, nodal importance seems to allow for significant narrowing of the search domain to find those solutions. When 5 sensors are employed in the default attack case, the search domain was effectively narrowed to less than 8% of WDS nodes, while for all cases with 20 sensors employed the search was narrowed to roughly between 11 and 18%.

Solution convergence plots in Figures 7.24 and 7.25 show that ISSM can converge to near-optimal solutions with only few iterations. The behavior of the curves in Figure 7.24 is particularly interesting when compared to the counterpart curves in Figure 6.23 for BWSN Network 2. In Figure 6.23, the curves for the different numbers of sensors employed are quite separated from each other, whereas the curves in Figure 7.24 are close to each other when subset size is small. Nodal importance rankings for the Toms River WDS seem to be more grouped according to regions of the system, likely due to the heterogeneity of the system in contrast to the largely homogenous configuration of BWSN Network 2. Small subsets of highly-ranked nodes can provide relatively good solutions when 5 sensors are employed; these 5 sensors play the major roles in reducing Z_{vol}. But, placing 15 additional sensors with these same small subsets would only provide limited additional protection to a few local areas in the system due to the grouping of nodal importance rankings. Therefore, several iterations of ISSM are needed to account

for all areas of the system in order to place 20 sensors in an evenly distributed manner. The consequence is the closeness of the curves in Figure 7.24 for the first few ISSM subsets. Despite the ranking scheme being correlated to particular system areas, the best solution for the Toms River WDS can still be isolated to relatively small proportions of WDS nodes as indicated in Table 7.11.

Table 7.11 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for minimizing expected contaminated demand volume in the default attack case and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	1,135	7.6%
20	1,675	11.2%

Table 7.12 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for minimizing expected contaminated demand volume with $M = 20$ in the variant attack cases and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

Attack Case	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
default	1,675	11.2%
sensor detection resolution = 0.3 mg/L	2,363	15.8%
contaminant injection duration = 10 h	2,685	17.9%
contaminant hazard threshold = 0.0 mg/L	2,248	15.0%
response delay = 3 h	1,675	11.2%
2 injection nodes at 1 time	2,219	14.8%
2 injection nodes at 2 times	2,150	14.4%

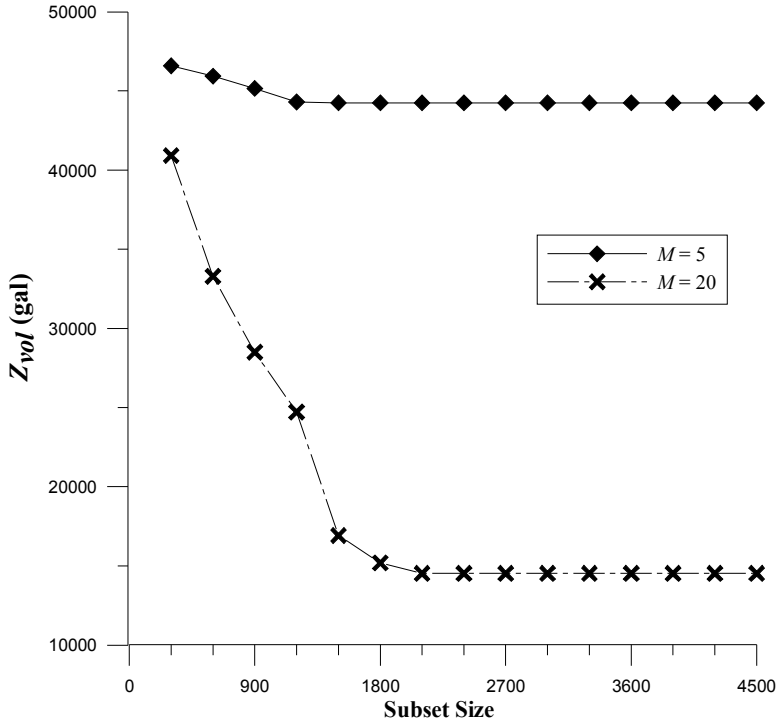


Figure 7.24 Solution convergence as ISSM is carried out for minimizing expected contaminated demand volume in the default attack case for the Toms River WDS. (Only solution values for subset sizes of 4,500 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 4,500.)

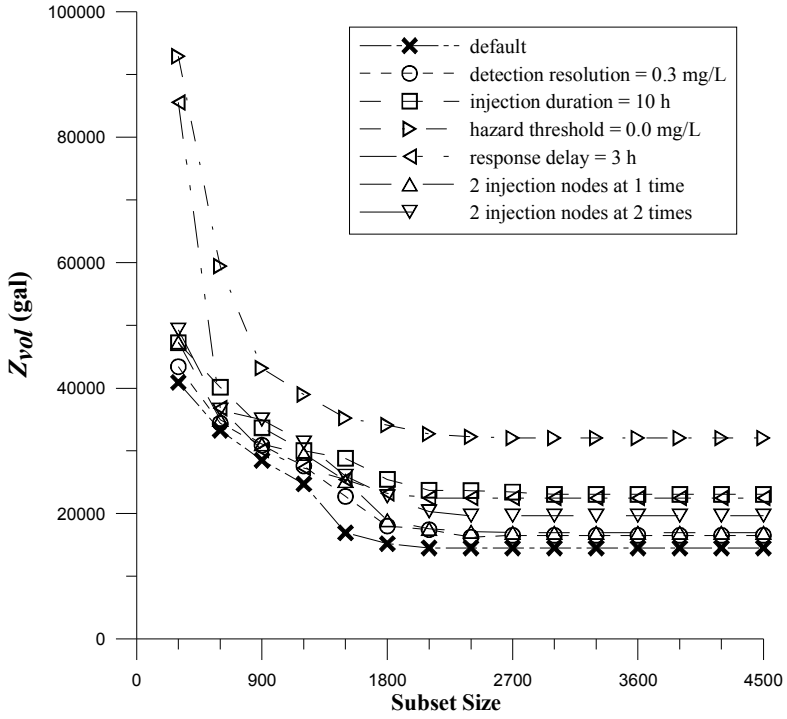


Figure 7.25 Solution convergence as ISSM is carried out for minimizing expected contaminated demand volume with $M = 20$ in the default and variant attack cases for the Toms River WDS. (Only solution values for subset sizes of 4,500 or less are plotted as solutions for all cases reached best performance by subset size = 4,500.)

7.4 MULTIOBJECTIVE PROBLEM: MAXIMIZING Z_{all}

The results of sensor placement with regard to the three-prong multiobjective problem are presented and discussed in this section.

7.4.1 SENSOR NODE SOLUTIONS

The sensor placement optimization results for the three-prong multiobjective problem for the Toms River WDS are presented in this section. Table 7.13 gives the sensor nodes selected for the default attack case, and Figures 7.26 and 7.27 indicate the locations of these nodes in the system. The pattern of solutions for maximizing Z_{all} resembling to some degree those for minimizing Z_{time} is apparent for this system as well, as evidenced by the Z_{time} values in Table 7.13 and sensor node layouts in Figures 7.26 and 7.27. This trend is also seen in examining sensor placement for the variant attack cases, as expected. Table 7.14 gives the sensor nodes for those cases, and Figures 7.28 through 7.33 depict their locations in the system.

Also as expected, sensor node locations do not vary greatly from attack case to attack case. One major difference between the sensor node solutions for this system and those of the other two systems is that the optimal solutions preserve much protection in terms of both minimizing Z_{time} and minimizing Z_{vol}, whereas for the other two systems a greater amount of volume-related protection was sacrificed for the multiobjective solution. Even the sensor node locations in the figures below suggest the protection preservation for both

of those individual protection goals; the locations visually appear to be a compromise between the placements for minimizing Z_{time} and minimizing Z_{vol}. This difference is explored further below.

Table 7.13 Optimal sensor locations and performance measures for the three-prong multiobjective problem in the default attack case for the Toms River WDS.

M	Sensor Nodes	Z_{all}	Z_{lik} (%)	Z_{time} (min)	Z_{vol} (gal)
5	382, 437, 7125, 8792, 10130	2.79	50.1	3,431.2	61,793
20	382, 437, 1796, 3678, 4657, 5183, 5741, 6667, 7061, 7125, 7428, 7920, 8254, 8375, 8792, 9331, 9660, 10130, 10236, 10412	2.49	53.1	3,064.9	21,923

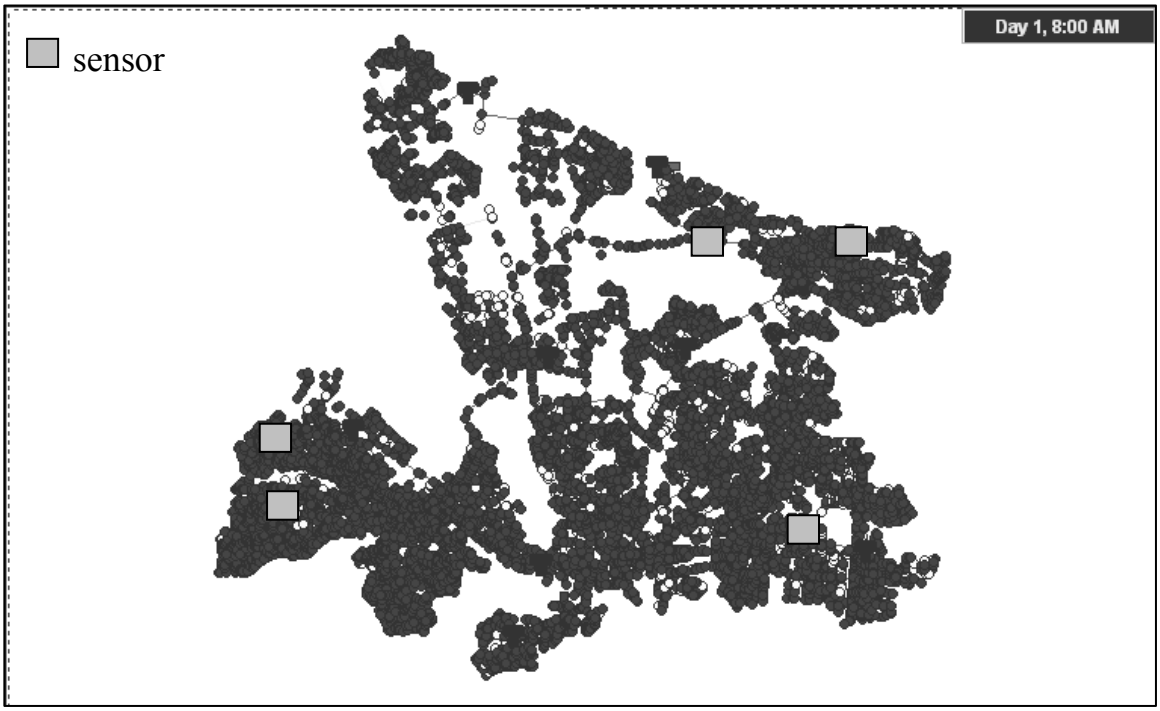


Figure 7.26 Optimal sensor locations with $M = 5$ for the three-prong multiobjective problem in the default attack case for the Toms River WDS.

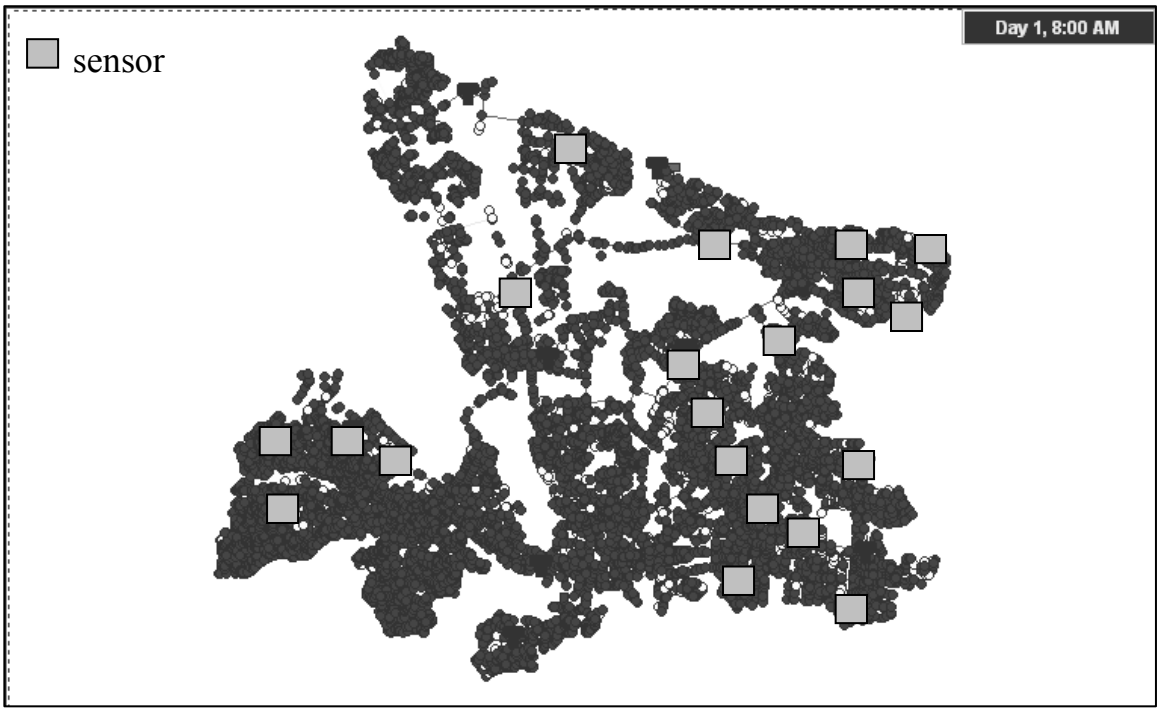


Figure 7.27 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the default attack case for the Toms River WDS.

Table 7.14 Optimal sensor locations and performance measures for the three-prong multiobjective problem with $M = 20$ in the variant attack cases for the Toms River WDS.

Attack Case	Sensor Nodes	Z_{all}	Z_{lik} (%)	Z_{time} (min)	Z_{vol} (gal)
default	382, 437, 1796, 3678, 4657, 5183, 5741, 6667, 7061, 7125, 7428, 7920, 8254, 8375, 8792, 9331, 9660, 10130, 10236, 10412	2.49	53.1	3,064.9	21,923
sensor detection resolution = 0.3 mg/L	382, 437, 1796, 2286, 3678, 3832, 5183, 6667, 7061, 7125, 7428, 7920, 8254, 8375, 8792, 9331, 9660, 10130, 10236, 10412	2.51	51.7	3,176.8	24,874
contaminant injection duration = 10 h	382, 434, 1796, 3678, 3836, 5183, 5741, 6667, 7061, 7125, 7428, 8254, 8375, 8792, 9331, 9516, 9660, 10130, 10236, 10412	2.47	54.2	3,002.1	28,623
contaminant hazard threshold = 0.0 mg/L	437, 1796, 3547, 3678, 4657, 5183, 5741, 6667, 7061, 7125, 7428, 7920, 8254, 8375, 8792, 9331, 9660, 10130, 10236, 10412	2.53	53.2	3,063.4	42,678
response delay = 3 h	382, 434, 1796, 3678, 4657, 5183, 5743, 6667, 7061, 7125, 7428, 7920, 8254, 8375, 8792, 9331, 9660, 10130, 10236, 10412	2.50	53.1	3,064.2	30,769
2 injection nodes at 1 time	382, 437, 3678, 3757, 4657, 5183, 5741, 6667, 7061, 7125, 7428, 8254, 8375, 8792, 9331, 9516, 9660, 10130, 10236, 10412	2.52	77.6	1,902.5	24,712
2 injection nodes at 2 times	437, 3547, 3678, 3757, 4657, 5183, 5741, 6667, 7061, 7125, 7428, 8254, 8375, 8792, 9331, 9516, 9660, 10130, 10236, 10412	2.52	77.5	1,989.2	26,203

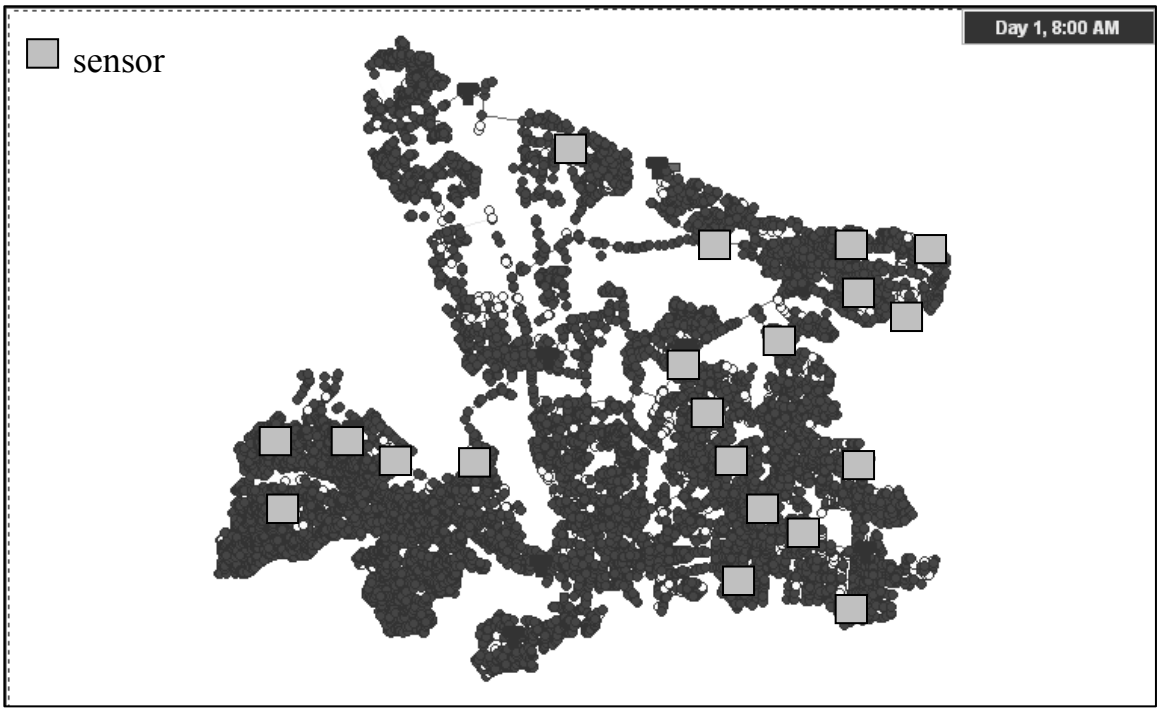


Figure 7.28 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a detection resolution of 0.3 mg/L for the Toms River WDS.

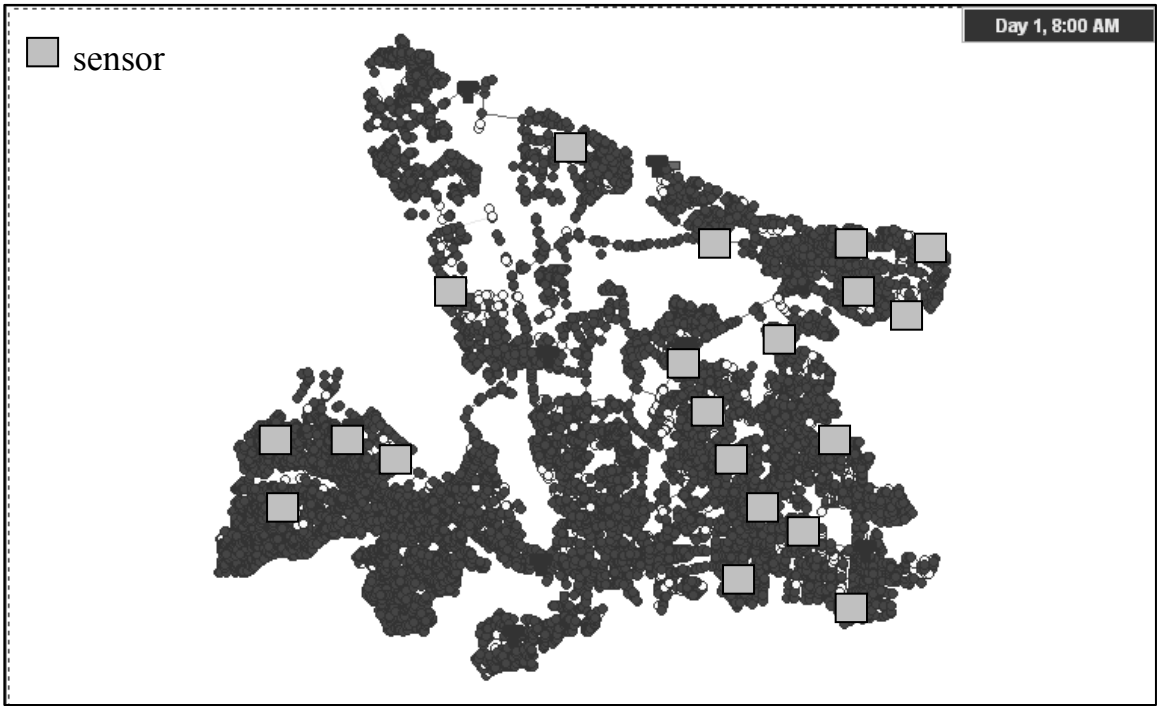


Figure 7.29 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a 10-hour injection duration for the Toms River WDS.

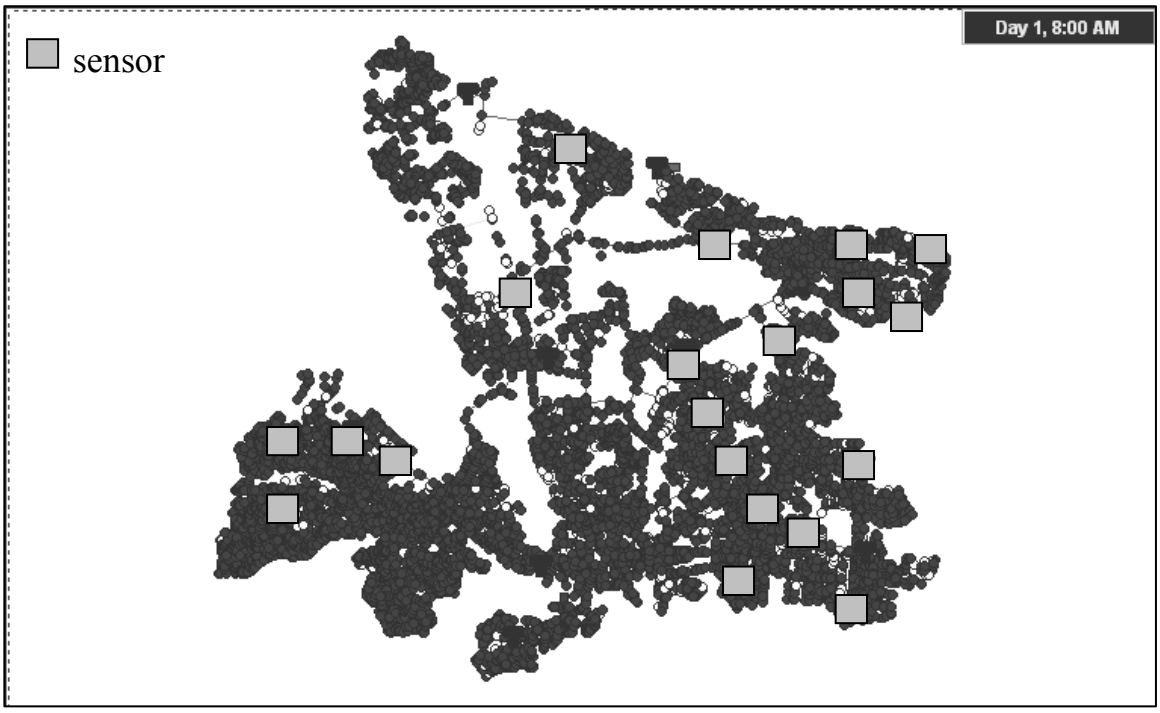


Figure 7.30 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a hazard threshold of 0.0 mg/L for the Toms River WDS.

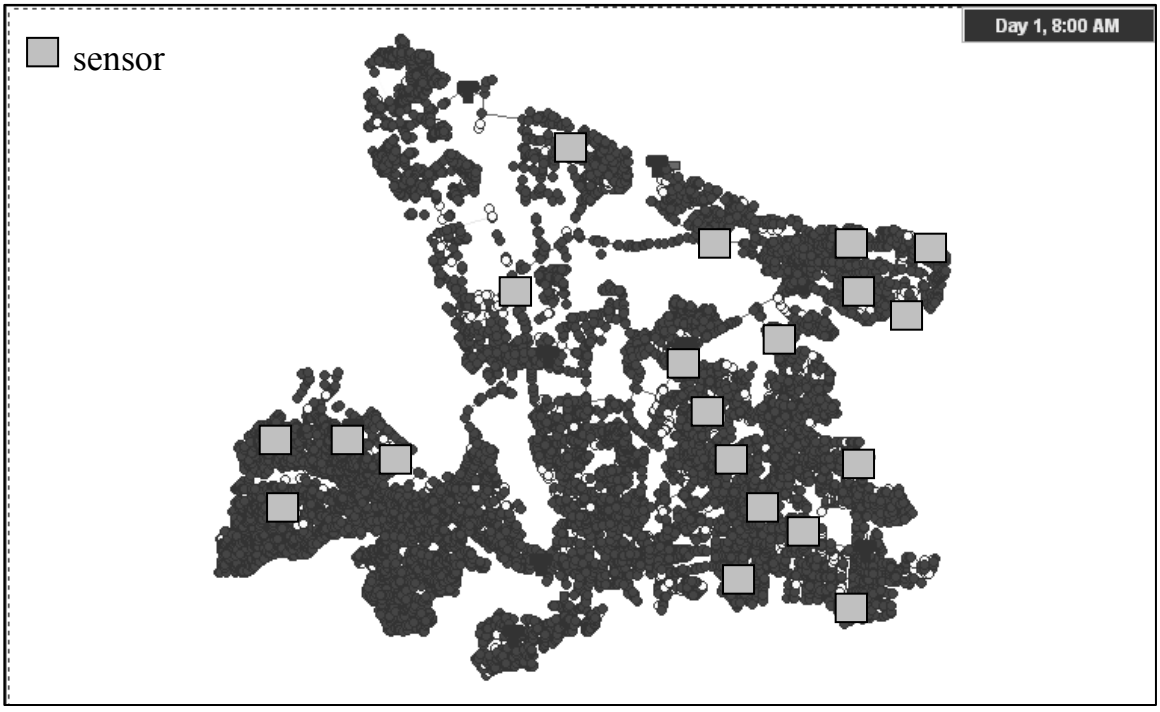


Figure 7.31 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of a 3-hour response delay for the Toms River WDS.

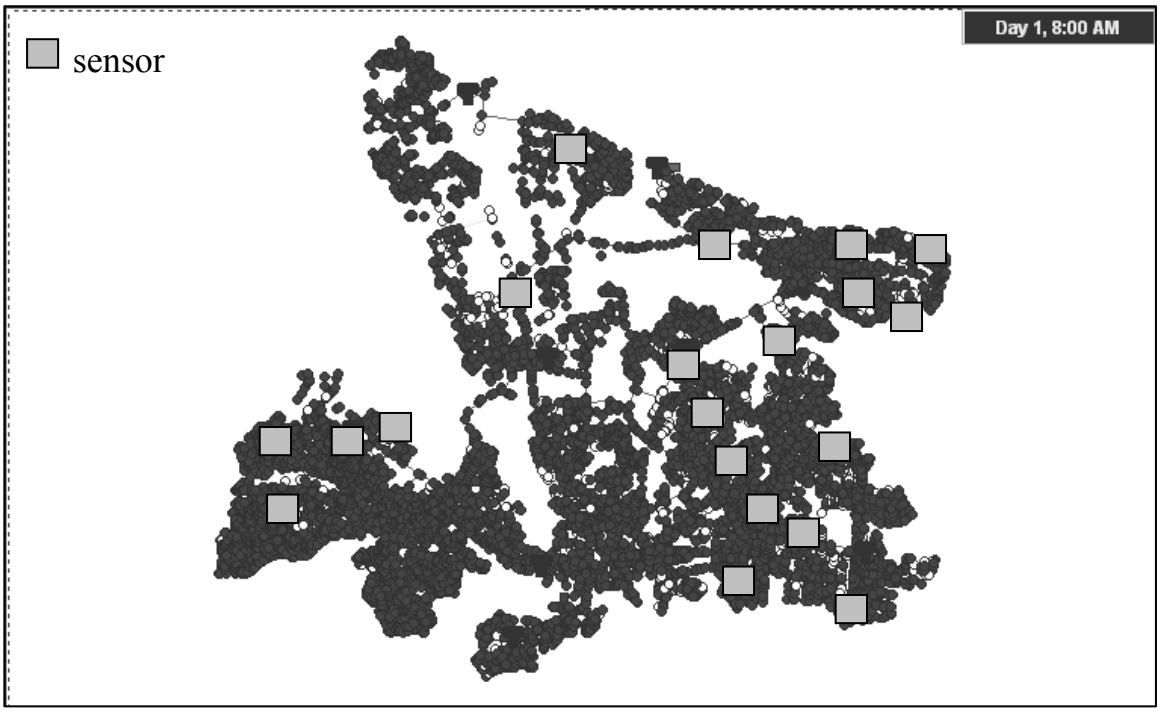


Figure 7.32 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of 2 nodes injected at 1 time for the Toms River WDS.

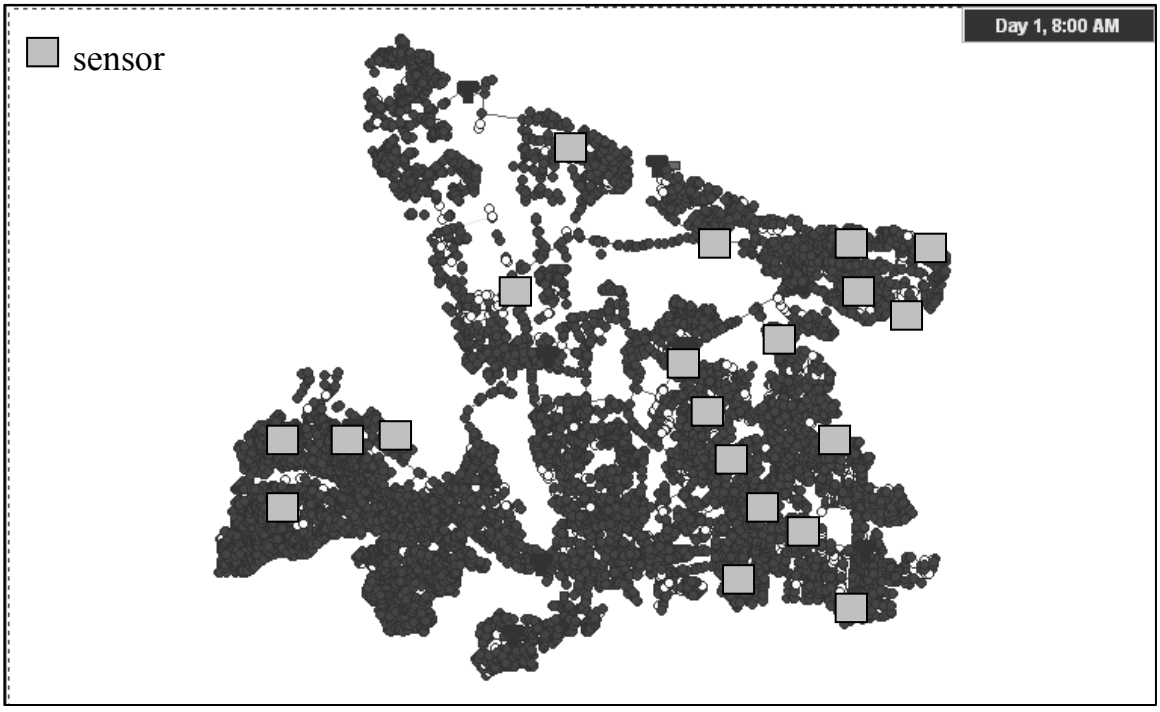


Figure 7.33 Optimal sensor locations with $M = 20$ for the three-prong multiobjective problem in the case of 2 nodes injected at 2 times for the Toms River WDS.

7.4.2 ANALYSIS OF PERFORMANCE TRADEOFFS

The compromises in protection performance made in selecting multiobjective sensor node solutions when 20 sensors are employed are quantified and discussed below. Table 7.15 provides these quantified costs for the default attack case, and Table 7.16 provides the costs for the variant cases. As mentioned above, multiobjective solutions for this system preserve a great deal of protection with regard to both minimizing Z_{time} and minimizing Z_{vol}, and this protection preservation is evident in examining the performance costs in Tables 7.15 and 7.16. The costs in terms of maximizing Z_{lik} are significantly more than they were for multiobjective solutions for the other two study systems.

The differences in performance tradeoffs between this system and the other two systems are reasonable, though. The solutions to maximize Z_{all} mirror those to minimize Z_{time} to significant degrees. The performance costs in terms of maximizing Z_{lik} for this system in minimizing Z_{time} were not large, but they are somewhat more than those for the other two systems. Thus, the performance costs in terms of maximizing Z_{lik} in maximizing Z_{all} are also expected to be somewhat larger. In order to compensate for the increase costs, multiobjective solutions are chosen to preserve greater protection in terms of minimizing Z_{vol}.

Table 7.15 Performance tradeoffs for individual protection goals resulting from maximizing Z_{all} with $M = 20$ in the default attack case for the Toms River WDS.

Protection Goal Performance Measure	Pareto Front Space Minimum Value	Pareto Front Space Maximum Value	Value for Maximizing Z_{all}	Protection Performance Cost
Z_{lik}	45.9%	59.4%*	53.1%	46.7%
Z_{time}	3063.0 min*	3367.0 min	3064.9 min	0.6%
Z_{vol}	14,525 gal*	227,484 gal	21,923 gal	3.5%

* Value represents maximum protection with respect to the particular protection goal in question.

Table 7.16 Performance tradeoffs for individual protection goals resulting from maximizing Z_{all} with $M = 20$ in the variant attack cases for the Toms River WDS.

Attack Case	Protection Performance Cost, Z_{lik}	Protection Performance Cost, Z_{time}	Protection Performance Cost, Z_{vol}
default	46.7%	0.6%	3.5%
sensor detection resolution = 0.3 mg/L	43.6%	0.2%	5.2%
contaminant injection duration = 10 h	45.5%	5.1%	2.1%
contaminant hazard threshold = 0.0 mg/L	46.6%	0.1%	0.2%
response delay = 3 h	46.7%	0.4%	3.2%
2 injection nodes at 1 time	39.3%	5.9%	3.2%
2 injection nodes at 2 times	43.4%	2.0%	2.7%

7.4.3 EVALUATION OF NODAL IMPORTANCE

The multiobjective sensor placement results above were found using ISSM that employed nodal importance rankings obtained from performing the Pareto dominance-based approach described in Section 4.5.2.1. The effectiveness of nodal importance in

narrowing the search for those solutions is indicated by Tables 7.17 and 7.18. The ranks of lowest-ranked nodes in the tables suggest that a proportion of approximately 15% of WDS nodes is required to contain the optimal solution with 5 sensors employed, and roughly 25 to 30% of WDS nodes are needed to contain the optimal solution with 20 sensors employed. Solution convergence plots in Figures 7.34 and 7.35 show that convergence to good solutions is not especially rapid, though. ISSM requires a subset of almost 2,400 nodes in size to find a near-optimal solution when 5 sensors are employed in the default attack case, and a subset of about 4,000 nodes is required with 20 sensors employed for all attack cases.

Table 7.17 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for the three-prong multiobjective problem in the default attack case and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

M	**Lowest Nodal Importance Rank**	**Proportion of Ranked WDS Nodes Containing Optimal Solution**
5	2,229	14.9%
20	4,574	30.5%

Table 7.18 Nodal importance ranks for the lowest-ranked Toms River WDS nodes of optimal solutions for the three-prong multiobjective problem with $M = 20$ in the variant attack cases and the proportions of ranked Toms River WDS nodes containing the optimal solutions.

Attack Case	Lowest Nodal Importance Rank	Proportion of Ranked WDS Nodes Containing Optimal Solution
default	4,574	30.5%
sensor detection resolution = 0.3 mg/L	4,109	27.4%
contaminant injection duration = 10 h	3,859	25.8%
contaminant hazard threshold = 0.0 mg/L	4,126	27.6%
response delay = 3 h	4,574	30.5%
2 injection nodes at 1 time	4,454	29.7%
2 injection nodes at 2 times	4,017	26.8%

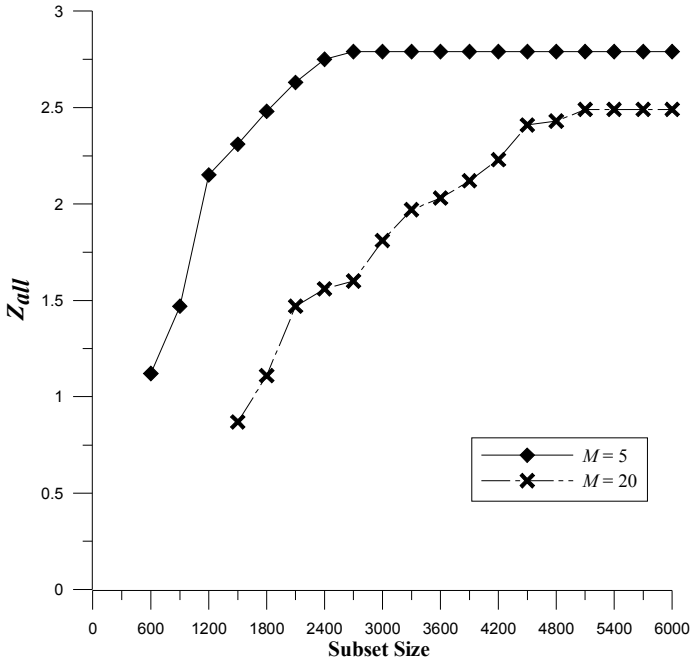


Figure 7.34 Solution convergence as ISSM is carried out for the three-prong multiobjective problem in the default attack case for the Toms River WDS. (Only solution values for subset sizes of 6,000 or less are plotted as solution reached best performance for both $M = 5$ and $M = 20$ by subset size = 6,000; negative Z_{all} values are not plotted.)

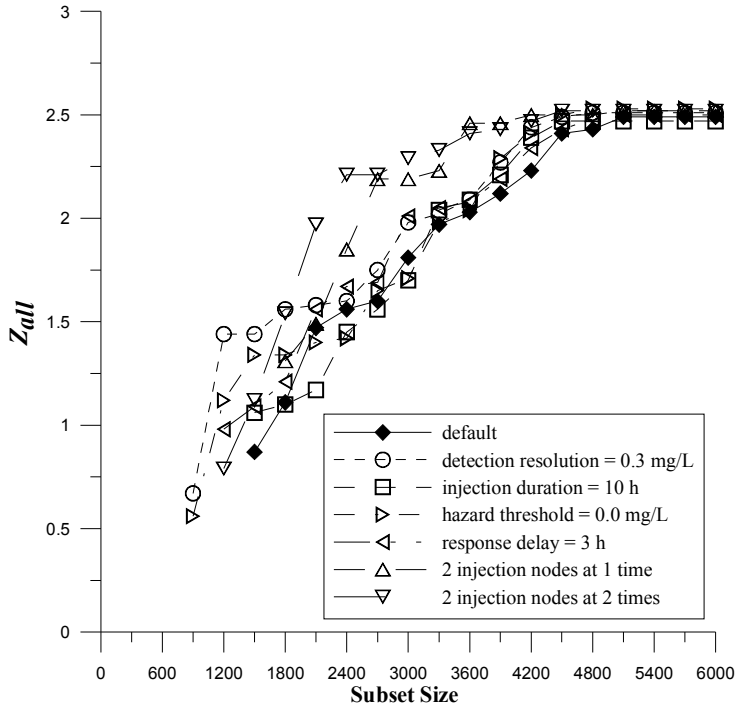


Figure 7.35 Solution convergence as ISSM is carried out for the three-prong multiobjective problem with $M = 20$ in the default and variant attack cases for the Toms River WDS. (Only solution values for subset sizes of 6,000 or less are plotted as solutions for all cases reached best performance by subset size = 6,000; negative Z_{all} values are not plotted.)

7.5 COMPUTATIONAL EXPENSE NOTES

The computational resources needed to conduct sensor placement optimization for the Toms River WDS are comparable to those needed for BWSN Network 2, which is not surprising since the two systems are of comparable size. Scenario node generation took about 16 hours to carry out and produced about 1.1 GB of data for use in the importance and optimization phases. The importance and optimization phases required between 3 and 6 hours for a given multiobjective problem case.

CHAPTER 8

CONCLUSION

In this chapter, the important outcomes of this study are summarized, the observed advantages and drawbacks of the methods developed in this study are discussed, and the implications and possible future applications of this study are suggested. In other words, this chapter will argue

- how well this study answered the research challenges explained in Chapter 2 in bringing efficiency and effectiveness to the optimization of WDS sensor monitoring and
- how this study can contribute to the continued addressing of those challenges in future research and practice.

8.1 OVERALL EVALUATION OF SENSOR PLACEMENT OPTIMIZATION METHODS

The effectiveness of the methodology of this study is evaluated below with respect to four key aspects: (i) sensor placement solution performance, (ii) the role of the multiobjective optimization problem formulation in multiobjective solution performance, (iii) the effectiveness of nodal importance, and (iv) the performance of the Iterative Subset Search Method.

8.1.1 SENSOR PLACEMENT SOLUTION PERFORMANCE

For every protection goal, attack case, and study system examined, the optimization methods developed in this work yielded sensor node solutions that provided very desirable levels of protection with respect to particular protection goals. When sensor placement results for this study were compared to available results in literature, the sensor nodes of this study were at least comparable in performance to those of the other works. In most cases, the sensor nodes of this work were associated with equal or better performance than the performance corresponding to the other documented sensor node schemes. However, for one protection goal and one study system (i.e., minimizing expected contaminated demand volume for BWSN Network 2), the results of this study were of slightly lesser performance than those of another study, although they seemed better than those of the other studies used for comparison purposes.

Another measure of validation for the sensor nodes found in this work is that sensor locations did not vary greatly on average between attack cases for any protection goal or any study system. This phenomenon is seen in the submitted sensor nodes of other works given in Chapters 5 and 6, and it supports the observation of Isovitsch and VanBriesen (2008) regarding the lack of substantial change in sensor locations from attack case to attack case.

It is fair to say, on the whole, that the optimization methods developed in the study performed well in terms of providing protection for a variety of systems, conditions, and protection goals.

8.1.2 MULTIOBJECTIVE OPTIMIZATION PROBLEM FORMULATION

Generally, the multiobjective optimization formulation developed in this study (employing the "lumped" single-objective function to maximize Z_{all}) yielded solutions that performed well in terms of providing desirable levels of protection with regard to the three individual protection goals and minimizing the protection tradeoffs between protection goals. The resulting multiobjective sensor node solutions in all cases provided almost as much--if not as much--protection with respect to minimizing expected detection time as did solving the time-related single-objective problem. Such multiobjective solutions are reasonable given the three particular individual protection goals for this study. As detection likelihood and expected detection time are correlated strongly with each other given the way expected detection time is formulated in this study, reducing expected detection time has the effect on average of increasing detection likelihood. Reducing expected detection time also helps to reduce expected contaminated demand volume by allowing contaminated water to be detected sooner before additional volume can become contaminated. Thus, the protection costs with respect to individual protection goals for the selected multiobjective solutions of this study were minimal as demonstrated in Chapters 5, 6, and 7. Though other works provided multiobjective solutions that were found under different but valid multiobjective problem formulations

and protection goal preferences, comparing the multiobjective solutions for this work with those of the other works at least allowed for verification that other works had not submitted solutions that are better under the multiobjective problem formulation of this work than the solutions found in this work. In sum, the multiobjective sensor placement results found in this study support the legitimacy of the multiobjective formulation to quantify and minimize protection performance tradeoffs.

8.1.3 NODAL IMPORTANCE

The desirable sensor placement solutions found in this study were found largely due to the effectiveness of the nodal importance concepts and functions developed. Nodal importance narrowed the search for a given optimal sensor placement solution considerably in most cases by quantifying the desirability of WDS nodes as sensor node candidates and allowing for the Iterative Subset Search Method to find the optimal sensor nodes—or at least near-optimal sensor nodes—using only a relatively small proportion of WDS nodes.

While all nodal importance functions and approaches developed in this study appear to work very well overall for all protection goals, nodal importance for maximizing detection likelihood appears to work best in narrowing the search for optimal solutions to very small proportions of WDS nodes in most cases. The effectiveness of sorting by end-of-line status (and slow-outflow status for small systems) prior to ranking nodes according to nodal importance value was shown by the results of this study to provide

additional narrowing of the search domain of WDS nodes. In fact, for BWSN Network 1 in the cases employing 5 sensors, the ranking was nearly perfect in identifying the truly important nodes with the aid of the sorting procedure. Even without the sorting procedure implemented, the nodal importance functions for maximizing detection likelihood by themselves provided for a significant amount of narrowing of the search domain, as demonstrated in the cases of a non-zero sensor detection resolution where the underlying assumptions regarding end-of-line and slow-outflow nodes were not applicable. However, the sorting approach may lead to near-optimal if not optimal solutions even if the underlying assumptions are not applicable. For example, all of the solution nodes for maximizing Zlik in the attack case of a 0.3 mg/L sensor detection resolution for BWSN Network 2 were end-of-line nodes. The sensor node results submitted by Berry et al. (2006c) for maximizing detection likelihood for BWSN Networks 1 and 2 provide additional validation that the nodal importance approaches are effective in identifying truly important nodes. Though Berry et al. did not make use of nodal importance concepts such as uniqueness and end-of-line status but instead relied on the measured effects of placing sensors in the system to determine their "nodal impact coefficients", many of the sensor node locations selected in this study for those systems are very close in proximity to if not the same as those submitted by Berry et al. Thus, it seems that the nodal importance approaches developed in this work estimate the "impact" that a chosen node can have as a sensor node well without the need to actually conduct sensor placement to determine this impact.

As was discussed in Chapter 6, the large-system nodal importance function for minimizing expected contaminated demand volume is very effective in providing desirable sensor node solutions to minimize Z_{vol} by narrowing the search domain of WDS nodes significantly, but best solutions found for some of the cases are suboptimal, though only slightly underperforming compared to solutions submitted by Berry et al. (2006c). These slightly suboptimal solutions reiterate that nodal importance concepts developed in this study are not perfect and have room for improvement in future study. It can still be argued, though, that the nodal importance concepts for minimizing Z_{vol} are helpful on average in narrowing the search and finding good solutions to very large degrees.

For the two large systems examined in this study (i.e., BWSN Network 2 and the Toms River WDS), the solutions reported were able to be isolated to proportions of approximately 30% of all WDS nodes or less for all protection goals and attack cases. Often, the proportion was significantly smaller than 30%.

For the small system BWSN Network 1, the search domain was narrowed to small proportions of WDS nodes for all protection goals except in the multiobjective cases when 5 sensors were employed. In solving the multiobjective problem for BWSN Network 1, it was observed that the Pareto-dominance based nodal importance approach did not allow for the entire set of optimal sensor nodes to be captured in a small proportion of WDS nodes in the one injection node attack cases. As discussed in Chapter 5, most of the sensor nodes selected are very highly ranked in terms of multiobjective importance, and near-optimal solutions can be found with small subsets using ISSM. The

problem in these cases, though, is that one optimal sensor node is lowly-ranked as it is very important for both maximizing detection likelihood and minimizing expected detection time but considered unimportant for minimizing expected contaminated demand volume. This issue is a concern for small systems only. For large systems, there are many nodes that are sufficiently important with respect to all 3 individual protection goals and, in turn, would be ranked adequately high. Indeed, this issue does not appear to be a significant problem in optimizing sensor placement for the three-prong multiobjective problem for BWSN Network 2 or the Toms River WDS.

When 20 sensors were employed in the cases of minimizing Z_{time} and Z_{vol} for BWSN Network 1, the search domain to find the optimal sensor node solution was not able to be narrowed much at all for reasons discussed in Chapter 5. To summarize the discussion regarding these reasons, with a relatively high number of sensors employed to WDS node ratio (M/N ratio), some sensor nodes provide the "major" protection with regard to the particular protection goal; these nodes ranked highly when ranked appropriately in terms of nodal importance. If the major protection is provided, then remaining available sensor nodes provide protection by adding to the coverage (i.e., increasing detection likelihood) so that instances of "infinite" detection time or undetected contaminated volumes are not counted in Z_{time} or Z_{vol} computations. The resulting problem is that the importance functions do not account for this "changed" priority for those remaining sensors. Therefore, the corresponding sensor nodes are ranked relatively low. For larger systems of interest, though, the M/N ratio is very low under practical conditions such that this

drawback is not an issue of concern, as indicated by results for BWSN Network 2 and the Toms River WDS.

8.1.4 ITERATIVE SUBSET SEARCH METHOD

In general, the Iterative Subset Search Method compensated well for the error associated with nodal importance-based rankings for all attack cases, protection goals, and study systems. As supported by solution convergence plots in Chapter 5, 6, and 7, the subset sizes for ISSM chosen for this study allowed for reasonable rates of solution improvement as the search domain broadened iteratively to find optimal solution in most cases.

The inclusion of randomly chosen "unranked" nodes in subsets also proved to be a wise decision, especially considering the case of the three-prong multiobjective problem with 20 sensors employed in the default attack case; several optimal sensor nodes were ranked low in this case, but the random inclusion of unranked nodes allowed those optimal nodes to be included in a subset of size 70 and the entire set of optimal solution nodes to be found without searching nearly the entire set of WDS nodes.

The ISSM stopping criteria chosen for this study (i.e., three iterations of no solution improvement with all nodes in the optimal solution being "ranked" nodes) seems to be suitable overall given the ISSM parameters employed for particular study systems in this work. In most cases, the stopping criteria would yield the solutions presented in Chapters

5, 6, and 7, but there are two key exceptions. One exception involves the cases of high *M*/*N* ratio where the solution is steady over several iterations then can increase in performance when nodes increasing system coverage are included in the best solution. The other exception is that of the multiobjective problem cases with 5 sensors employed for BWSN Network 1; one optimal solution node was ranked low in some of these cases as discussed above, so the solution increased slightly in performance when this node was included in the solution. These issues, though, are isolated to small systems only. Employing a larger subset size and/or including a higher proportion of unranked nodes in the ISSM subset may rectify these issues in future study of small systems.

A final concern involving ISSM, though tangentially, was the particular genetic algorithm model and library used in this study to actually conduct the search for the best solution in a given subset. Relatively longer runtimes, stricter convergence criteria, and larger population sizes were needed in order to achieve the results provided in Chapters 5 through 7, even with narrowed search domains. Though these GA-related drawbacks were manageable and did not result in excessive computational expense, another GA model or library or even another optimization algorithm may prove to make ISSM more efficient and the nodal importance concepts more attractive for use in future study.

8.2 POTENTIAL APPLICATIONS FOR THIS WORK

There is an array of distinct advantages in applying the methods and concepts developed in this work in future studies over methods and concepts previously documented in

literature. As stated in Chapter 1, large systems are of key concern for monitoring in the event of a terrorist attack. Monitoring studies of these systems have been either infeasible or compromised by oversimplifying assumptions due to the large computational expense and resources required to conduct such studies, as discussed in Chapter 2. In this study, it was demonstrated that optimization implementing nodal importance concepts can allow for feasible and realistic contaminant sensor placement optimization for large, real-world WDSs. Even systems on the order of millions of nodes, such as metropolis-level systems, may now be reasonably practical to study given the ability of the methods developed in this work to narrow the search for optimal sensor nodes considerably.

The conventional thinking in previous works that attempt to preprocess WDS sensor nodes prior to the actual optimization of sensor placement is that it is necessary to characterize exactly how WDS nodes would perform as sensor nodes. This thinking has led to the prescribing of extensive computational simulation (e.g., Berry et al. 2006c) or oversimplification of WDS behavior (e.g., Xu et al. 2008). One key outcome of this study was showing that such an exact characterization of nodes is not required. Nodal importance provides only an estimate of the potential for a node to perform well as a sensor node, and a search algorithm can be employed to find optimal solutions despite the associated errors in nodal importance. Instead of extensive simulation being conducted to preprocess nodes, only a manageable set of attack scenarios need to be simulated to measure contaminant propagation throughout the WDS in order to perform straightforward nodal importance calculations. Computational expense is reduced

relative to previously documented methods, and the complexity of WDS behavior is adequately modeled. Thus, optimization can be performed for virtually any attack case, unlike in the case of the methods of Berry et al. Also, optimization with the methods of this study can be performed with computational resources (e.g., one desktop computer) and limited time (e.g., within one day) that are practical for most researchers and practitioners. Nodal importance functions and approaches of this study can even be adopted for use with other GA models or other types of optimization algorithms that would further realize the efficiency gained by employing nodal importance.

The relative simplicity of the nodal importance concept could also encourage wider use of nodal importance among researchers and professionals in future study. Nodal importance is based on almost intuitive relationships between WDS behavior at the nodal level and sensor placement desirability that are accessible for those with only a basic understanding WDS behavior, whereas other methods require an in-depth knowledge of underlying theories and/or particular procedures, leading for discouragement for common use. Additionally, given that effective nodal importance approaches were developed in this study for three rather different protection goals, other nodal importance functions and approaches should be able to be developed for protection goals beyond those included in this study by further considering those intuitive relationships. In sum, the potential to extend this work to further address the research challenges corresponding to WDS sensor placement more effectively and efficiently is immense.

REFERENCES

Al-Zahrani, M. A., and Moied, K. "Locating optimum water quality monitoring stations in water distribution system." *World Water and Environmental Resources Congress 2001*, Reston, VA.

Aral, M. M., Guan, J., and Maslia, M. L. "A multi-objective optimization algorithm for sensor placement in water distribution systems." *World Environmental and Water Resources Congress 2008*, Honolulu, HI.

ASCE. (2004). "Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System." Reston, VA.

Berry, J., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., and Watson, J.-P. "On the Placement of Imperfect Sensors in Municipal Water Networks." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA, 129-129.

Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G., and Watson, J.-P. (2006b). "Sensor Placement in Municipal Water Networks with Temporal Integer Programming Models." *Journal of Water Resources Planning and Management*, 132(4), 218-224.

Berry, J. W., Fleischer, L., Hart, W. E., Phillips, C. A., and Watson, J.-P. (2005). "Sensor Placement in Municipal Water Networks." *Journal of Water Resources Planning and Management*, 131(3), 237-243.

Berry, J. W., Hart, W. E., Phillips, C. A., and Watson, J.-P. "A Facility Location Approach to Sensor Placement Optimization." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Carr, R. D., Greenberg, H. J., Hart, W. E., Konjevod, G., Lauer, E., Lin, H., Morrison, T., and Phillips, C. A. (2006). "Robust optimization of contaminant sensor placement for community water systems." *Mathematical Programming*, 107(1/2), 337-356.

Copeland, C., and Cody, B. (2002). "Terrorism and Security Issues Facing the Water Infrastructure Sector." Congressional Research Service.

Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S. T., and Savic, D. "An Efficient Algorithm for Sensor Placement in Water Distribution Systems." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Eliades, D., and Polycarpou, M. "Iterative Deepening of Pareto Solutions in Water Sensor Networks." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Goldberg, D. E. (1989). *Genetic algorithms in search and optimization*, Addison-Wesley, Reading, Mass.

Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. "Optimization Model and Algorithms for Design of Water Sensor Placement in Water Distribution Systems." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Gueli, R. "Predator - Prey Model for Discrete Sensor Placement." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Harmant, P., Nace, A., Kiene, L., and Fotoohi, F. "Optimal supervision of drinking water distribution network." *26th Annual Water Resources Planning and Management Conference 1999*, Reston, VA.

Huang, J. J., McBean, E. A., and James, W. "Multi-Objective Optimization for Monitoring Sensor Placement in Water Distribution Systems." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Isovitsch, S. L., and VanBriesen, J. M. (2008). "Sensor Placement and Optimization Criteria Dependencies in a Water Distribution System." *Journal of Water Resources Planning and Management*, 134(2), 186-196.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting Accidental Contaminations in Municipal Water Networks." *Journal of Water Resources Planning and Management*, 124(4), 192-198.

Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., and Fischbeck, P. "Optimizing Sensor Placements in Water Distribution Systems Using Submodular Function Maximization." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Kumar, A., Kansal, M. L., and Arora, G. (1997). "Identification of Monitoring Stations in Water Distribution System." *Journal of Environmental Engineering*, 123(8), 746-752.

Lee, B. H., and Deininger, R. A. (1992). "Optimal Locations of Monitoring Stations in Water Distribution System." *Journal of Environmental Engineering*, 118(1), 4-16.

Maslia, M. L., Sautner, J. B., and Aral, M. M. (2000). "Analysis of the 1998 water-distribution system serving the Dover Township area, New Jersey: Field-data collection activities and water-distribution system modeling." Agency for Toxic Substances and Disease Registry, Atlanta, GA.

Maslia, M. L., Sautner, J. B., Aral, M. M., Gillig, R. E., Reyes, J. J., and Williams, R. C. (2001). "Historical reconstruction of the water-distribution system serving the Dover Township area, New Jersey: January 1962 - December 1996." Agency for Toxic Substances and Disease Registry, Atlanta, GA.

Ostfeld, A., and Salomons, E. (2004). "Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security." *Journal of Water Resources Planning and Management*, 130(5), 377-385.

Ostfeld, A., and Salomons, E. (2005). "Securing Water Distribution Systems Using Online Contamination Monitoring." *Journal of Water Resources Planning and Management*, 131(5), 402-405.

Ostfeld, A., and Salomons, E. "Sensor Network Design Proposal for the Battle of the Water Sensor Networks (BWSN)." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Ostfeld, A., Uber, J., and Salomons, E. (2006). "Battle of the water sensor networks: A design challenge for engineers and algorithms."

Propato, M. (2006). "Contamination Warning in Water Networks: General Mixed-Integer Linear Models for Sensor Location Design." *Journal of Water Resources Planning and Management*, 132(4), 225-233.

Propato, M., and Piller, O. "Battle of the Water Sensor Networks." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Rogers, S. W., Guan, J., Maslia, M. L., and Aral, M. M. "Nodal Importance Concept for Computational Efficiency in Optimal Sensor Placement in Water Distribution Systems." *World Environmental and Water Resources Congress 2007*, Tampa, Florida, USA, 463-463.

Rossman, L. A. (2000). "EPANET 2 users manual." National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, Ohio.

Shastri, Y., and Diwekar, U. (2006). "Sensor Placement in Water Networks: A Stochastic Programming Approach." *Journal of Water Resources Planning and Management*, 132(3), 192-203.

U.S. Congress. (2002). "Public law 107-188: Public health security and bioterrorism preparedness and response act of 2002." Washington, D.C.

U.S. EPA. (2003). "Instructions to Assist Community Water Systems in Complying with the Public Health Security and Bioterrorism Preparedness and Response Act of 2002." U.S. EPA Office of Water.

Wall, M. (2005). "GAlib: A C++ library of genetic algorithm components."

Wang, J. (2008). "Sensitivity and Uncertainty Analyses of Contaminant Fate and Transport in a Field-Scale Subsurface System," Ph.D., Georgia Institute of Technology Atlanta.

Watson, J.-P., Greenberg, H. J., and Hart, W. E. "A multiple-objective analysis of sensor placement optimization in water networks." *World Water and Environmental Resources Congress 2004*, Reston, VA.

Woo, H.-M., Yoon, J.-H., and Choi, D.-Y. "Optimal monitoring sites based on water quality and quantity in water distribution systems." *World Water and Environmental Resources Congress 2001*, Reston, VA.

Wu, Z. Y., and Walski, T. "Multi Objective Optimization of Sensor Placement in Water Distribution Systems." *Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA.

Xu, J., Fischbeck, P. S., Small, M. J., VanBriesen, J. M., and Casman, E. (2008). "Identifying Sets of Key Nodes for Placing Sensors in Dynamic Water

Distribution Networks." *Journal of Water Resources Planning and Management*, 134(4), 378-385.