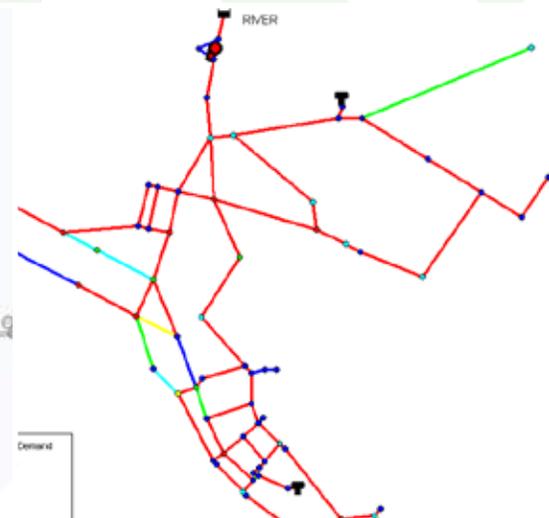
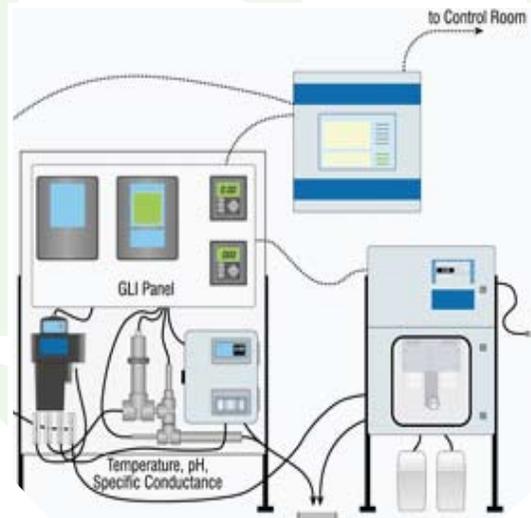


# Sensor Network Design for Drinking Water Contamination Warning Systems

A Compendium of Research Results and Case Studies  
Using the TEVA-SPOT Software





# Sensor Network Design for Drinking Water Contamination Warning Systems

A Compendium of Research Results and Case Studies Using  
the TEVA-SPOT Software

**Regan Murray, Terra Haxton, and Robert Janke**

National Homeland Security Research Center  
Cincinnati, OH 45268

**William E. Hart, Jonathan Berry, and Cynthia Phillips**

Sandia National Laboratories  
Albuquerque, NM 87185

NATIONAL HOMELAND SECURITY RESEARCH CENTER  
OFFICE OF RESEARCH AND DEVELOPMENT  
U.S. ENVIRONMENTAL PROTECTION AGENCY  
CINCINNATI, OH 45268

# Disclaimer

The information in this document has been funded wholly or in part by the U.S. Environmental Protection Agency (EPA). It has been subjected to the Agency's peer and administrative review, and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This work was performed under Interagency Agreement (IA) DW89921928 with Sandia National Laboratories, Contract EP-C-05-056 with Pegasus Technical Services Inc., and IA DW89922555 with Argonne National Laboratory. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U. S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

The TEVA-SPOT software described in this manual is subject to copyright. It is free software that can be redistributed and/or modified under the terms of the GNU Lesser General Public License as published by the Free Software Foundation and to the terms of other third-party software licenses. Specifications of these terms are included with the TEVA-SPOT software distribution.

The authors and the U.S. Environmental Protection Agency are not responsible and assume no liability whatsoever for any results or any use made of the results obtained from this software, nor for any damages or litigation that result from the use of this software for any purpose.

# Foreword

Following the events of September 11, 2001, EPA's mission expanded to address critical needs related to homeland security. Presidential Directives identified EPA as the primary federal agency responsible for safeguarding the nation's water supplies and for decontamination following a chemical, biological, and/or radiological (CBR) attack. To provide scientific and technical support in meeting this expanded mission, EPA's National Homeland Security Research Center (NHSRC) was established. NHSRC is focused on conducting research and delivering products that improve the capability of the Agency to carry out its homeland security responsibilities.

As a part of this mission, NHSRC conducts research and provides technical assistance to support America's drinking water utilities so they can improve their security preparedness, response and recovery. Over the last several years, NHSRC has been developing new methods to help design, implement, and evaluate drinking water contamination warning systems. These new systems integrate a variety of monitoring technologies to rapidly detect contamination. One important question for contamination warning system design is where to most effectively place a limited number of sensors in a water distribution network. This network may be composed of hundreds to thousands of miles of pipe and the contamination warning system must economically safeguard the largest number of people. **This publication summarizes a large body of research addressing sensor placement issues, and provides critical information for water utilities to use when considering where to place sensors in their own distribution networks.**

NHSRC works with many partners to meet its responsibilities. This research was conducted in collaboration with EPA's Office of Water, across the federal government working with the U.S. Department of Energy's Sandia National Laboratories and Argonne National Laboratory, with academia through the University of Cincinnati, and with the American Water Works Association and their member utilities.

This publication provides a comprehensive resource on sensor placement methods and case studies and is intended for a broad audience of water utility staff, policy makers, and researchers. NHSRC has made this publication available to help improve the security and the quality of our nation's drinking water. This research is intended to move EPA one step closer to achieving its homeland security goals and its overall mission of protecting human health and the environment while providing sustainable solutions to our environmental problems.

Cynthia Sonich-Mullin, Acting Director  
National Homeland Security Research Center

# Acknowledgments

The National Homeland Security Research Center would like to acknowledge the following organizations and individuals for their support in the development of this report, providing a peer review, or testing the TEVA-SPOT software.

**American Water**

Zia Bukhari

**American Water Works Association**

Kevin Morley

**Argonne National Laboratory**

Mike Davis

Thomas Taxon

**EPA Office of Research and Development**

Terra Haxton

Robert Janke

Regan Murray

Lewis Rossman

**EPA Office of Water**

Steve Allgeier

Mike Henrie

**Sandia National Laboratories**

William E. Hart

Jonathan Berry

Lee Ann Fisk

Cynthia Phillips

Jean-Paul Watson

**Technion, Israel Institute of Technology**

Avi Ostfeld

**University of Cincinnati**

James Uber

Walter Grayman, Consultant

# Table of Contents

Disclaimer .....	iv
Foreword .....	v
Acknowledgments.....	vi
List of Acronyms and Abbreviations .....	viii
1. Background and Purpose.....	1
2. TEVA Decision Framework: Modeling.....	9
3. TEVA Decision Framework: Decision Process.....	15
4. Real-World Applications .....	27
5. Challenges for Real-World Applications.....	35
6. Impact Assessment Methodology.....	45
7. Optimization Methodology .....	53
Appendix A. Literature Review .....	61
Appendix B. Battle of the Water Sensor Networks.....	67
Appendix C. Quality Assurance.....	71
References.....	73

# List of Acronyms and Abbreviations

AMSA	Association of Metropolitan Sewerage Agencies
ASCE	American Society of Civil Engineers
ASME-ITI	American Society of Mechanical Engineers-Innovative Technologies Institute
ATUS	American Time Use Survey
AWWA	American Water Works Association
AwwaRF	American Water Works Association Research Foundation (now the Water Research Foundation)
BLS	Bureau of Labor Statistics
BTACT	Bioterrorism Act (Public Health Security and Bioterrorism Preparedness and Response Act of 2002)
BWSN	Battle of the Water Sensor Networks
CUD	Compromise Utility Design
CVaR	Conditional Value at Risk
CWS	Contamination Warning System
DHS	U.S. Department of Homeland Security
EC	Extent of Contamination
EMPACT	Environmental Monitoring for Public Access and Community Tracking
EPA	U.S. Environmental Protection Agency
EPANET-MSX	EPANET Multi-Species Extension
EPS	Extended Period Simulation
FIFO	First In-First Out
FN	False Negative
FP	False Positive
GAO	U.S. General Accounting Office (now the U.S. Government Accountability Office)
GB	Gigabyte
GCWW	Greater Cincinnati Water Works
GIS	Geographic Information Systems
GRASP	Greedy Randomized Adaptive Search Procedure
HVAC	Heating, Ventilating, and Air Conditioning
IBWA	International Bottled Water Association
ID	Number of Incidents Detected
IDSE	Initial Distribution System Evaluation
LAG	Lagrangian
LD <sub>50</sub>	Lethal dose at which half the exposed population would die
LIFO	Last In-First Out
LP	Linear Program
MB	Megabytes
MC	Mass of contaminant Consumed
MGD	Million Gallons per Day
MIP	Mixed-Integer Program
NFD	Number of Failed Detections
NJAW	New Jersey American Water
NRWA	National Rural Water Association
ORP	Oxidation Reduction Potential

# List of Acronyms and Abbreviations

PE	People Exposed
PH	Public Health
PICO	Parallel Integer and Combinatorial Optimization
RAM	Random Access Memory
RAM-W	Risk Assessment Methodology for Water
RAMCAP	Risk Assessment Model for Critical Asset Protection
ROC	Receiver Operating Characteristic
SCADA	Supervisory Control and Data Acquisition
SEMS	Security Emergency Management System
SP	Sensor Placement
TD	Time of Detection
TEVA	Threat Ensemble Vulnerability Assessment
TEVA-SPOT	Threat Ensemble Vulnerability Assessment Sensor Placement Optimization Tool
TOC	Total Organic Carbon
UD	Utility Design
U.S.	United States of America
USGS	United States Geological Survey
UV	Ultraviolet
VA	Vulnerability Assessment
VaR	Value at Risk
VC	Volume of Contaminant Consumed
VOC	Volatile Organic Compound
VSAT	Vulnerability Self-Assessment Tool
VSL	Value of Statistical Life
waSP	witness aggregation Sensor Placement
WQ	Water Quality
WS	Water Security



# Background and Purpose

Protecting our nation's critical infrastructure from terrorist attacks has become a federal and local priority over the last several years. Under Homeland Security Presidential Directive 7, the United States Environmental Protection Agency (EPA) is the lead federal agency for protecting the water infrastructure in the United States. In this capacity, EPA has worked with public and private water utilities, federal, state and local agencies, and the public health community to develop assistance and research programs to improve the safety and security of drinking water systems. Water associations, community water systems, academia, private industry, and others have focused attention and research on developing new methods, policies, and procedures to secure drinking water and wastewater systems.

The Public Health Security and Bioterrorism Preparedness and Response Act of 2002 required drinking water systems serving more than 3,300 people to conduct vulnerability assessments and prepare or update emergency response plans that address a range of potential terrorist threats (BTACT 2002). In 2006, a report on the fourteen features of an active and effective security program informed the water community about the most important organizational, operational, infrastructure, and external features of resilient and secure systems (U.S. EPA 2006a). Many representatives of the water sector have joined together to prepare a sector-specific plan that coordinates activities across organizations (U.S. DHS et al. 2007). These activities have reduced water sector vulnerabilities through increasing awareness, hardening of critical assets, improved physical security, and more comprehensive response plans.

Recently, water security research efforts have focused on the advancement of methods for mitigating contamination threats to drinking water systems (see for example, Ostfeld 2006; AWWA 2005; Murray 2004). A promising approach for the mitigation of both accidental and intentional contamination is a *Contamination Warning System (CWS)*, a system to deploy and operate online sensors, other surveillance systems, rapid communication technologies, and data analysis methods to provide an early indication of contamination (U.S. EPA 2005c). CWSs with multiple approaches to monitoring — like water quality sensors located throughout the distribution system, public health surveillance systems, and customer complaint monitoring programs — are theoretically capable of detecting a wide range of contaminants in water systems. However, CWSs are expensive to purchase, install, and maintain. To make them a viable option, there is a clear need to minimize the investment required by individual drinking water systems.

The purpose of this report is to provide documentation on strategies and tools needed to assist in the design of an online sensor network for a CWS. A key aspect of CWS design is

the strategic placement of sensors throughout the distribution network. There has been a large volume of research on this topic in the last several years, including a “Battle of the Water Sensor Networks” (Ostfeld et al. 2008) that compared 15 different approaches to solving this problem. This report focuses on the sensor placement methodologies that have been developed by EPA's Threat Ensemble Vulnerability Assessment (TEVA) Research Team, which is composed of researchers from EPA, Sandia National Laboratories, the University of Cincinnati, and Argonne National Laboratory. This team has developed TEVA-SPOT — the Threat Ensemble Vulnerability Assessment Sensor Placement Optimization Tool — a collection of software tools that can help utilities design sensor networks (Berry et al. 2008b; U.S. EPA 2009).

This report is organized as follows. Chapters 1–5 are intended for a broad audience of water utility staff, policy makers, and researchers. This chapter provides background information and an overview of the research on sensor placement methods. Chapter 2 discusses the data required as input to sensor placement methods, highlighting the important design decisions a utility would need to make. Chapter 3 describes the iterative decision-making process a utility would follow when implementing optimization software. Chapter 4 provides several real-world case studies, and Chapter 5 discusses several common challenges that a user might face when applying sensor placement software to real water systems.

Chapters 6 and the rest of this report are intended for researchers and others who want to understand the modeling and optimization methods in greater detail. Chapter 6 is focused on the methodology for estimating the impacts of drinking water contamination, including methods for estimating dose and public health response. Chapter 7 describes the optimization problem for locating sensors. Appendix A includes a full literature review, and Appendix B provides a summary of the Battle of the Water Sensor Networks (Ostfeld et al. 2008).

## Vulnerability of Drinking Water Distribution Systems

The heightened risk of terrorist attacks on our nation's critical infrastructure has placed the security of the water supply in the same league as the security of our nation's treasured monuments. There is a long history of threats to water systems and a shorter list of actual incidents at water systems (AwwaRF 2003; Kunze 1997; Staudinger et al. 2006). However, public awareness of the threat has increased dramatically since the 9/11 attacks partly due to media coverage of two international terrorist plots against drinking water supplies; one premised on the introduction of a cyanide

compound into water pipes near a U.S. Embassy in Italy (Henneberger 2002), and the other a direct threat to American water supplies from an Al-Qaeda operative (Cameron 2002).

Although the threat of terrorist attacks might not be a daily worry for water utilities, terrorist threats are of significant concern because of their potentially large public health and economic impacts. Conceivable terrorist threats to drinking water systems include the physical destruction of facilities or equipment, airborne release of hazardous chemicals stored onsite, sabotage of Supervisory Control and Data Acquisition (SCADA) and other computer systems, and the introduction of chemical, biological, or radiological contaminants into the water supply (ASCE 2004). Explosive and flammable agents that could cause physical destruction of facilities might be threats to drinking water systems because of the ease of obtaining the necessary equipment, the past use of these agents as terrorists' weapons of choice, and the general ease of access to water facilities, such as storage tanks and pumping stations. However, contamination hazards might pose a more significant threat because they could result in major public health and economic impacts and long-lasting psychological impacts.

### **Drinking Water Vulnerability Assessments**

The Bioterrorism Act of 2002 requires all community water systems serving more than 3,300 customers to “conduct an assessment of the vulnerability of its system to a terrorist attack” and to submit a copy of the assessment to EPA. The law directs vulnerability assessments to include “a review of pipes and constructed conveyances, physical barriers, water collection, pretreatment, treatment, storage and distribution facilities, electronic, computer, or other automated systems which are utilized by the public water system, the use, storage, or handling of various chemicals, and the operation and maintenance of such system.”

Based on its particular facilities, treatment methods, water sources, regional topology, and service community, each water utility faces unique vulnerabilities to terrorist threats. Several risk assessment tools and methodologies have been developed to aid drinking water systems in determining these vulnerabilities. RAM-W, the Risk Assessment Methodology for Water developed by Sandia National Laboratories in 2000–01 with funding from the American Water Works Association Research Foundation (AwwaRF) and EPA, was based on a risk assessment approach for nuclear facilities and was later expanded to apply to buildings, federal dams, prisons, nuclear power plants and now water utilities (AwwaRF et al. 2002). Other methodologies include VSAT, the Vulnerability Self-Assessment Tool developed by the Association of Metropolitan Sewerage Agencies for wastewater and drinking water systems (AMSA 2003), and SEMS, the Security Emergency Management System developed by the National Rural Water Association (NRWA 2003).

Staudinger et al. (2006) provide a review of vulnerability assessment (VA) methods for small systems, and they suggest that standards and minimum requirements should be

developed. Along these lines, the Department of Homeland Security (DHS) has developed the Risk Assessment Model for Critical Asset Protection (RAMCAP). RAMCAP allows the risk of a specific asset to be compared to the risk of assets from different critical infrastructure sectors, e.g., communications or energy. The goal of the process is to identify national assets that deserve more thorough assessment of risk. The water sector is working with DHS to ensure that water vulnerability assessment tools are “RAMCAP compliant,” meaning that the results can be used in RAMCAP rankings (U.S. DHS et al. 2007).

Most VA tools are based on the following six common elements: (1) characterization of the water system's mission, objectives, facilities, and operations; (2) identification of potential adverse consequences and prioritization of the water quality, public health, and economic impacts; (3) determination of critical assets; (4) assessment in partnership with law enforcement of the likelihood of malevolent acts; (5) evaluation of existing countermeasures; and (6) analysis of risk and development of a risk reduction plan (U.S. EPA 2002b). In general, a utility selects a team composed of employees, law enforcement and community officials, and consultants who share their expertise in order to identify collectively the most likely malevolent acts against the utility, its most vulnerable assets, and the actions that will optimally reduce the risk associated with these assets. The RAMCAP framework is a seven-step approach that includes all of the above steps with additional threat assessment performed by DHS (ASME-ITI 2005).

### **Need for Distribution System Vulnerability Framework**

Drinking water distribution systems are large networks of storage tanks, valves, pumps, and pipes that transport finished water to customers over vast areas; typically hundreds to thousands of miles of pipe. A General Accounting Office (GAO) report found that 75% of the water experts interviewed believe distribution systems are the *most vulnerable* component of drinking water systems (U.S. GAO 2003). Moreover, EPA's Office of Inspector General found that “neither EPA nor the different [VA] methodologies adequately emphasized distribution system threats as the most susceptible components of water systems to include in vulnerability assessments,” (U.S. EPA 2003). Thousands of drinking water systems across the country have completed vulnerability assessments and are using the results to plan security improvements to their facilities, but the existing VA methodologies lack a thorough analysis of distribution systems.

In particular, none of the VA methodologies adequately reflect the vulnerabilities of distribution systems to contamination. Contamination of distribution systems might occur through intentional terrorist or criminal acts, but could also occur accidentally. Many warfare agents have been noted as potential drinking water contamination threats (Burrows et al. 1999). Accidental human contamination of distribution systems with pesticides, toxic industrial chemicals, and other materials has been documented (Watts

2009). Distribution systems can also be contaminated during the course of normal operations; for example, metals, organic contaminants, and asbestos in pipe materials and linings can leach into the system, and soil and ground water contaminants can permeate plastic pipes, (U.S. EPA 2002a). In addition, persistent or transient pressure loss can result in pesticides, insecticides, or other chemicals entering the system through accidental backflow incidents, and contaminated soil water entering through pipe breaks or leaking joints.

An adequate distribution system VA methodology should take into account the unique features of distribution systems: complicated networks of pipes, pumps, valves, tanks, and other physical components, dynamic and complex flows, the randomness of demand, and population mobility (Clark et al. 2001). Moreover, because of the uncertainties involved in predicting the characteristics of a contamination event and its consequences, a VA methodology should allow for a probabilistic assessment of potential public health and

economic consequences. All these characteristics require the dynamic and probabilistic modeling of the vulnerability of distribution systems.

### The Threat Ensemble Vulnerability Assessment Framework

To meet this need, EPA and its collaborators at Sandia National Laboratories, Argonne National Laboratory, and the University of Cincinnati developed a probabilistic framework for analyzing the vulnerability of drinking water distribution systems called Threat Ensemble Vulnerability Assessment (TEVA). **Figure 1-1** outlines the major modules of the framework: the simulation of contamination incidents, the assessment of potential consequences of those incidents, and the design and evaluation of threat mitigation strategies. Together, these modules allow one to develop an integrated view of the vulnerability of a distribution system to a wide variety of contamination threats, and the potential to decrease this vulnerability through a set of mitigation strategies.

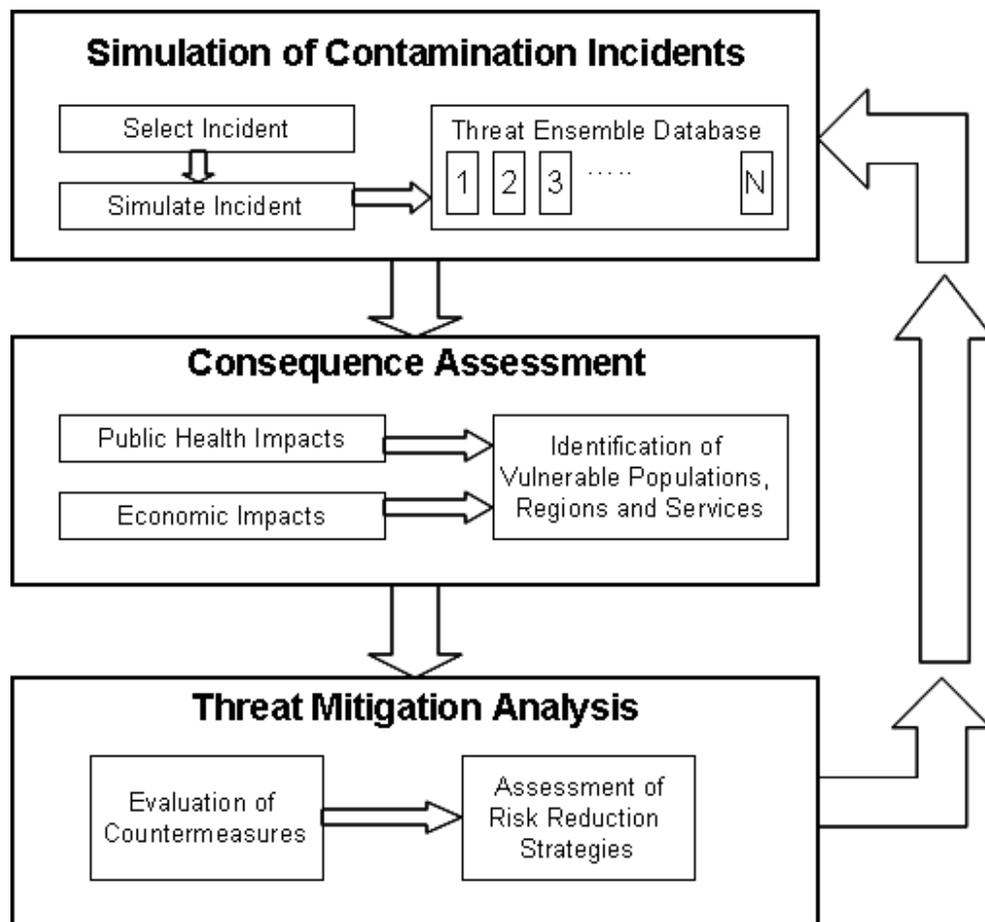


Figure 1-1. Threat Ensemble Vulnerability Assessment (TEVA) framework.

**Table 1-1.** How TEVA supports the six basic vulnerability assessment elements.

VA Basic Element	TEVA Element
Characterize water system	Simulation of Incidents (development of EPANET network model)
Identify and prioritize adverse impacts	Simulation of Incidents Consequence Assessment
Identify critical assets	Consequence Assessment
Assess likelihood of adverse impacts	Simulation of Incidents
Evaluation of existing countermeasures	Threat Mitigation Analysis
Develop risk reduction plan or actions	Threat Mitigation Analysis

Without specific intelligence information, one cannot predict exactly how terrorist groups might sabotage a water system. Therefore, the TEVA framework is based on a probabilistic analysis of a large number of likely contamination incidents. Although the number of possible variations on terrorist attacks is nearly infinite, by selecting a “large enough” set of likely incidents, the expected impacts of contamination incidents can be assessed. A single contamination incident can be defined by the type of contaminant, the amount and concentration of the contaminant, the location of the injection into the distribution system, and the start and stop time of the injection. A threat ensemble, then, is a large collection of distinct incidents. In the TEVA framework (as well as in previous work by Ostfeld et al. 2004), the vulnerability of a water system is based on an assessment of the entire threat ensemble. TEVA fits into the general VA structure as shown in **Table 1-1**.

## Drinking Water Contamination Warning Systems

Research on methods to mitigate the impacts of contamination incidents have converged over the last several years on the concept of a contamination warning system (CWS).

CWSs have been proposed as a promising approach for the early detection and management of contamination incidents in drinking water distribution systems (ASCE 2004; AWWA 2005; U.S. EPA 2005a). EPA is piloting CWSs through the Office of Water’s Water Security (WS) Initiative, formerly called WaterSentinel, at a series of drinking water utilities.

The key to an effective response to a water contamination incident is minimizing the time between detection of a contamination incident and implementation of effective response actions that mitigate further consequences. Implementation of a robust CWS can achieve this objective by providing an earlier indication of a potential contamination incident than would be possible in the absence of a CWS. A CWS is a proactive approach that uses advanced monitoring technologies and enhanced surveillance activities to collect, integrate, analyze, and communicate information that provides a timely warning of potential contamination incidents.

The WS Initiative promotes a comprehensive CWS that is theoretically capable of detecting a wide range of contaminants, covering a large spatial area of the distribution

system, and providing early detection in time to mitigate impacts (U.S. EPA 2005c). Components of the WS Initiative include:

- **Online water quality monitoring.** Continuous online monitors for water quality parameters, such as chlorine residual, total organic carbon, electrical conductivity, pH, temperature, oxidation reduction potential, and turbidity help to establish expected baselines for these parameters in a given distribution system. Event detection systems, such as CANARY (Hart et al. 2007), can be used to detect anomalous changes from the baseline to provide an indication of potential contamination. Other monitoring technologies can be used as well, such as contaminant-specific monitors, although the goal is to detect a wide range of possible contaminants.
- **Consumer complaint surveillance.** Consumer complaints regarding unusual taste, odor, or appearance of the water are often reported to water utilities, which track the reports as well as steps taken by the utility to address these water quality problems. The WS Initiative is developing a process to automate the compilation and tracking of information provided by consumers. Unusual trends that might be indicative of a contamination incident can be rapidly identified using this approach.
- **Public health surveillance.** Syndromic surveillance conducted by the public health sector, including information such as unusual trends in over-the-counter sales of medication, as well as reports from emergency medical service logs, 911 call centers, and poison control hotlines might serve as a warning of a potential drinking water contamination incident. Information from these sources can be integrated into a CWS by developing a reliable and automated link between the public health sector and drinking water utilities.
- **Enhanced security monitoring.** Security breaches, witness accounts, and notifications by perpetrators, news media, or law enforcement can be monitored and documented through enhanced security practices. This component has the potential to detect a tampering event in progress, potentially preventing the introduction of a harmful contaminant into the drinking water system.

- **Routine sampling and analysis.** Water samples can be collected at a predetermined frequency and analyzed to establish a baseline of contaminants of concern. This will provide a baseline for comparison during the response to detection of a contamination incident. In addition, this component requires continual testing of the laboratory staff and procedures so that everyone is ready to respond to an actual incident.

A CWS is not merely a collection of monitors and equipment placed throughout a water system to alert of intrusion or contamination. Fundamentally, it is information acquisition and management. Different information streams must be captured, managed, analyzed, and interpreted in time to recognize potential contamination incidents and mitigate the impacts. Each of these information streams can independently provide some value in terms of timely initial detection. However, when these streams are integrated and used to evaluate a potential contamination incident, the credibility of the incident can be established more quickly and reliably than if any of the information streams were used alone. The primary purpose of a CWS is to detect contamination incidents, and implementation of a CWS is expected to result in dual-use benefits that will help to ensure its sustainability within a utility.

Many utilities are currently implementing some monitoring and surveillance activities, yet these activities are either lacking critical components or have not been integrated in a manner sufficient to meet the primary objectives of a CWS — timely detection of a contamination incident. For example, although many utilities currently track consumer complaint calls, a CWS requires a robust spatially-based system that, when integrated with data from public health surveillance, online water quality monitoring, and enhanced security monitoring, will provide specific, reliable, and timely information for decision makers to establish credibility and respond in an effective manner. Beyond each individual component of the CWS, coordination between the utility, the public health agency, local officials, law enforcement, and emergency responders, among others, is needed to develop an effective consequence management plan that ensures appropriate actions will occur in response to detection by different components. Critical to timely response is an advanced and integrated laboratory infrastructure to support baseline monitoring and analysis of samples collected in response to initial detections. In the absence of a reliable and sustainable CWS, a utility’s ability to respond to contamination incidents in a timely and appropriate manner is limited. Still, the challenge in applying a CWS is to reliably integrate the multiple streams of data in order to decide if a contamination incident has occurred.

## Sensor Network Design Research and Application

The overall goal of a CWS is to detect contamination incidents in time to reduce potential public health and economic consequences. The locations of online sensors can be optimized to help achieve these goals as well as other

objectives — for example, minimizing public exposure to contaminants, the spatial extent of contamination, detection time, or costs. These objectives are often at odds with each other, making it difficult to identify a single best sensor network design. In addition, there are many practical constraints and costs faced by water utilities. Consequently, designing a CWS is not a matter of performing a single optimization analysis. Instead, the design process is truly a multi-objective problem that requires informed decision making, using optimization tools to identify possible sensor network designs that work well under different assumptions and for different objectives. Water utilities must weigh the costs and benefits of different designs and understand the significant public health and cost tradeoffs.

There has been a large volume of research on techniques for sensor placement in the last several years, including a Battle of the Water Sensor Networks that compared 15 different approaches to this problem (Ostfeld et al. 2008). For a review of the large body of sensor placement research for water security, see Appendix A. Sensor placement strategies can be broadly characterized by the technical approach and the type of computational model used. The following categories reflect important differences in proposed sensor placement strategies:

- **Expert Opinion:** Although expertise with water distribution systems is always needed to design an effective CWS, here we refer to approaches that are solely guided by expert judgment. For example, Berry et al. (2005a) and Trachtman (2006) consider sensor placements developed by experts with significant knowledge of water distribution systems. These experts did not use computational models to carefully analyze network dynamics. Instead, they used their experience to identify locations whose water quality is representative of water throughout the network.
- **Ranking Methods:** A related approach is to use preference information to rank network locations (Bahadur et al. 2003; Ghimire et al. 2006). In this approach, a user provides preference values for the properties of a “desirable” sensor location, such as proximity to critical facilities. These preferences can then be used to rank the desirability of sensor locations throughout the network. Further, spatial information can be integrated to ensure good coverage of the network.
- **Optimization:** Sensor placement can be automated with optimization methods that computationally search for a sensor configuration that minimizes contamination risks. Optimization methods use a computational model to estimate the performance of a sensor configuration. For example, a model might compute the expected impact of an ensemble of contamination incidents, given sensors placed at strategic locations. See Appendix A for further discussion on sensor placement optimization literature.

This report focuses on the use of optimization to select sensor locations for a CWS. However, designing a CWS is not a matter of performing a single sensor placement analysis;

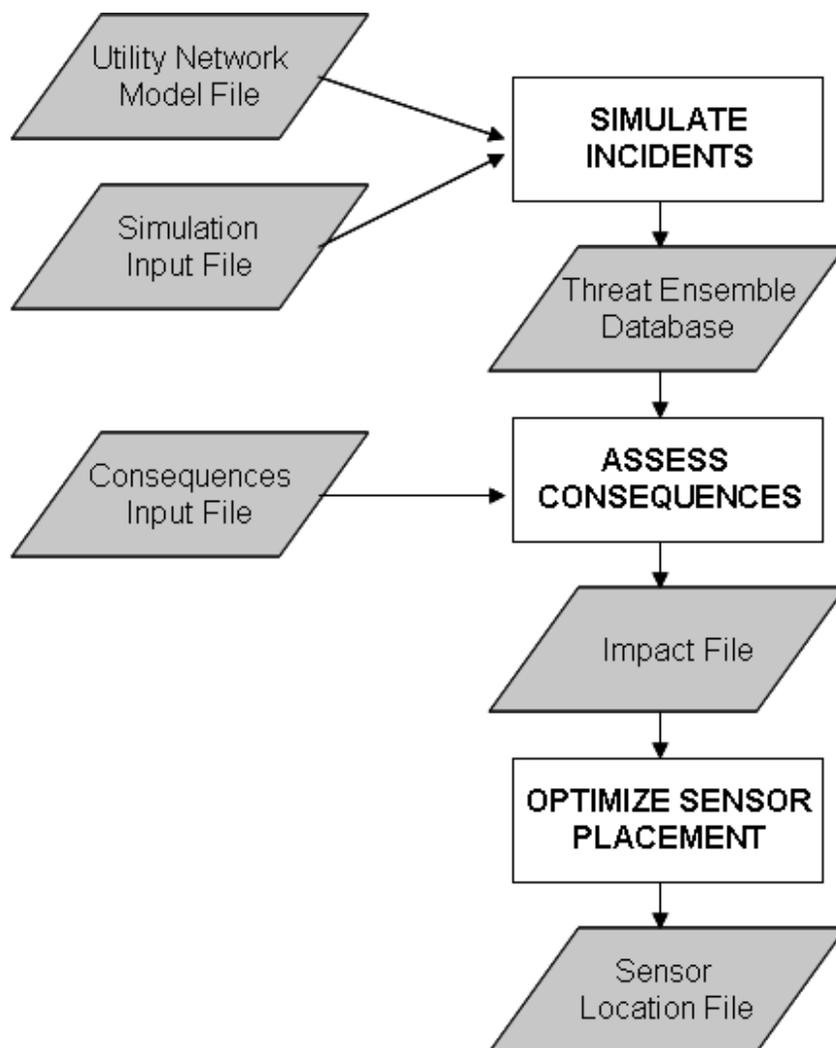
there are many factors that need to be considered when performing sensor placement, including utility response, the relevant design objectives, sensor behavior, practical constraints and costs, and expert knowledge of the water distribution system. In many cases, these factors can be at odds with one another (e.g., competing performance objectives), which makes it difficult to identify a single best sensor network design.

The TEVA Research Team has developed a decision-making process for CWS design that is composed of a modeling process and a decision-making process that employs optimization (Murray et al. 2008b). This modeling process includes creating or utilizing an existing network model for hydraulic and water quality analysis, describing sensor characteristics, defining the contamination threats, selecting performance measures, estimating utility response times following detection of contamination incidents, and identifying a set of potential sensor locations. The decision-making process involves applying an optimization method

and evaluating sensor placements. The process is informed by analyzing tradeoffs and comparing a series of designs to account for modeling and data uncertainties. The subsequent chapters of this report discuss this process in detail and illustrate sensor placement optimization using the TEVA-SPOT Toolkit (Berry et al. 2008b).

### The TEVA-SPOT Software

The TEVA-SPOT software is an application of the TEVA framework. The software consists of three main software modules that follow the diagram that was shown in **Figure 1-1**, and more specifically, in **Figure 1-2**. The first software module simulates the set of incidents in the threat ensemble. The second software module calculates the potential consequences of the contamination incidents contained in the threat ensemble. The third software module optimizes for sensor placement. The software is described in more detail in Chapters 6 and 7 of this report, and briefly summarized here



**Figure 1-2.** Data flow diagram for the TEVA-SPOT software.

**Consequence assessment.** Given a utility network model, and the set of parameter values determined in the modeling process, TEVA-SPOT calculates the consequences of each contamination incident in the design basis threat. The design basis threat is the set of incidents that the sensor network is designed to detect. The consequences are estimated in terms of one or more of the performance objectives, such as the number of people made ill or the length of pipe contaminated. Typically, TEVA-SPOT considers contamination incidents that occur at every node in the network model. TEVA-SPOT calculates consequences using EPANET for hydraulic and water quality calculations (Rossman 2000) and models for estimation of exposure and disease progression (Murray et al. 2006b).

**Optimization.** For most utility applications, TEVA-SPOT has been used to place sensors in such a way as to minimize the mean consequences for a given objective (averaged over the ensemble of contamination incidents). Minimizing the mean value is equivalent to assuming that each contamination incident is equally likely, and therefore all are important to consider when selecting a sensor network design. TEVA-SPOT does allow for user-specified weights that can be used to put more weight on locations with a higher likelihood of contamination; practically, this information is unlikely to be available with any certainty. If the user is most interested in protecting against a few catastrophic contamination incidents, TEVA-SPOT can also minimize the max-case impacts (Watson et al. 2004).

**Multi-objective analysis.** There are many competing CWS design objectives, e.g., the number of people made ill, the length of pipe contaminated, or the time to detection. TEVA-SPOT can only optimize over one objective at a time, but it does allow the user to explore tradeoffs between different sensor network designs and to find designs that perform well for more than one objective with the use of side-constraints (see Chapter 7).

**Fast, flexible solvers.** To allow for the comparison of designs based on multiple performance objectives and model parameters, TEVA-SPOT needs to be fast and flexible. Fast heuristic methods, integer programming heuristics and exact solvers are included in the software tool. This enables users to choose faster methods while at the same time understanding the confidence bounds on the sensor placement selected by the method. For most networks, designs can be found in seconds to minutes.

**Solver scalability.** A variety of strategies have been developed to ensure that TEVA-SPOT works on large networks with tens of thousands of pipes and junctions: aggregation of problem constraints, aggregation of contamination incidents, and/or specification of a limited set of feasible junctions for sensor placement (Hart et al. 2008b). Further, several of the TEVA-SPOT solvers have been modified to limit the memory required on standard 32-bit workstations. For example, the heuristic solver includes options that explicitly tradeoff memory and run-time.

**Application.** TEVA-SPOT has been used to design sensor networks for several medium and large U.S. water distribution systems, (Morley et al. 2007). The tool has been shown to outperform utility experts in selecting good sensor locations, see for example Berry et al. (2005a) and Ostfeld et al. (2008).

**Availability.** The authors have developed two versions of TEVA-SPOT: the TEVA-SPOT toolkit, which contains a library of functions and command line executables; and the TEVA-SPOT User Interface, which includes a graphical users' interface. For more information, see EPA's website (<http://www.epa.gov/nhsrc/>).



# TEVA Decision Framework: Modeling

Designing a CWS is not as simple as performing a single optimization analysis. Instead, the design process requires informed decision making, using optimization tools to identify possible network designs that work well under different assumptions and for different objectives. Water utilities must weigh the costs and benefits of different designs and understand the significant public health and cost tradeoffs.

Chapters 2 and 3 of this report describe a decision framework for CWS design. This framework uses optimization to generate sensor placements that allow water utilities to understand the significant public health and cost tradeoffs. The first step is to develop a conceptual model of the sensor network that identifies all the important characteristics of the planned sensor network. To create the conceptual model, one needs to know the layout of the distribution system and the current operating rules (as given by a utility network model), a description of the sensor characteristics, a clearly defined design basis threat for the CWS, appropriate performance measures for the CWS, an understanding of the planned utility response to detection of contamination incidents, and the locations where sensor can be located feasibly.

The goal of the modeling process is to accurately describe and model the characteristics of the planned CWS. This chapter focuses on the data required to complete the sensor network design and the decisions a utility will have to make prior to the optimization process. **Table 2-1** summarizes the data and information required; each component is described in more detail in the text. By gathering this data and making

these decisions up front, simulation tools can be used to measure how well such a sensor network would perform, and optimization methods can be used to find the best sensor network design.

## Utility Network Model

In order to determine system-specific sensor network designs, one needs a utility network model as input to a hydraulic and water quality modeling software package (e.g., EPANET). Sensor designs are based on minimizing the impacts of contamination incidents, which must be calculated using a utility network model. Therefore, an acceptable network model of the distribution system is needed in order to effectively design the sensor system. The following subsections describe the various issues/characteristics of an acceptable network model for use in sensor placement optimization, and more generally, for most water security modeling applications.

## Water Distribution System Models

Currently, most sensor placement optimization tools (e.g., TEVA-SPOT and PipelineNet) utilize EPANET to simulate flow and quality in water distribution systems. EPANET is a public domain water distribution system software package (Rossman 2000). Although sensor placement optimization tools are not dependent on features of EPANET, currently, its use requires the conversion of existing utility network models to EPANET input files.

**Table 2-1.** Information and data needed to perform sensor placement optimization.

Information and Data Needed for Sensor Placement	Description
Utility Network Model	The model (e.g., EPANET input file) should be up-to-date, capable of simulating operations for a 3-10 day period, and calibrated with field data
Sensor Characteristics	Type of sensors or sampling program, detection limits, and (if applicable) event detection system
Design Basis Threat	Data describing type of event that the utility would like to be able to detect: specific contaminants, behavior of adversary (quantity of contaminant, injection locations and durations), and customer behavior (temporal pattern of water consumption)
Performance Measures	Utility specific critical performance criteria, such as time to detection, number of illnesses, etc.
Utility Response	Plan for response to a positive sensor reading, including total time required for the utility to limit further public exposure
Potential Sensor Locations	List of all feasible locations for placing sensors, including associated model node/junction

Most commercial software packages utilize the basic EPANET calculation engine and contain a conversion tool for creating an EPANET input file from the files native to the commercial package. The user might encounter two potential types of problems when they attempt to make the conversion: (1) some commercial packages support component representations that are not directly compatible with EPANET such as representation of variable speed pumps, thus, the representation of these components might need to be modified in order to operate properly under EPANET; (2) conversion from the commercial software packages might also introduce some unintended representations within EPANET that could require manual correction. Following conversion, the output from the original model should be compared with the EPANET model output to ensure that the model results are the same or acceptably close (see section below on Model Testing).

An alternative to conversion is to use the commercial software to simulate contamination incidents and store the output in a properly formatted database. For example, as shown in **Figure 1-2**, TEVA-SPOT stores the EPANET output in the Threat Ensemble Database, which is then used independently by the sensor placement optimizer. Thus, it is possible to adapt output from a commercial tool into this format (for more details, see the TEVA-SPOT User Manual, Berry et al. 2008b).

### Extended Period Simulation

In order to support modeling of contamination incidents, the network model must be capable of extended period simulation (EPS) that represents the system operation over a period of several days. Typically, a network model that uses rules to control operations (e.g., turn pump A on when the water level in tank B drops to a specified level) are more resilient and amenable to long duration runs than are those that use controls based solely on time clocks. Simulations should be performed over a long duration to ensure that tank water levels are not systematically increasing or decreasing over the course of the run, since that will lead to situations that are not sustainable in the real-world.

The required length of simulation depends on the size and operation of the specific water system. However, in general, the length of the simulation should be as long as the longest travel times from sources to customer nodes. This can be calculated by simulating water age. In determining the required simulation length, small dead-ends (especially those with no-demand nodes) can be ignored. Typically a run length of 7 to 10 days is required for contamination simulations, though shorter periods could suffice for smaller systems and longer run times might be required for larger or more complex systems.

### Seasonal Models

In most cases, water security incidents can take place at any time of the day or any season of the year. As a result, sensor systems should be designed to operate during more than one representative time period in the water system.

It should be noted that this differs significantly from the normal design criteria for a water system where pipes are sized to accommodate water usage during peak seasons or during unusual events such as fires. In many cases, the only available network models are representative of these extreme cases. Generally, modifications should be made to reflect a broader time period prior to sensor placement optimization. Suggestions for model selection are provided below:

- **Optimal situation:** The utility has multiple network models representing common operating conditions throughout the year, such as a typical high demand case (e.g., average summer day) and a typical low demand case (e.g., average winter day).
- **Minimal situation:** The utility has a single network model representing relatively “average” conditions throughout the year.
- **Situations to avoid:** The utility has a single network model representing an extreme case (e.g., maximum day model).
- **Exceptions:** (1) If a sensor system is being designed to primarily monitor a water system during a specific event such as a major annual festival, then one of the models should reflect conditions during that event; and (2) if the water system experiences little variation in water demand and water system operation over the course of the year, then a single representative network model would suffice.

### Network Model Detail

A sufficient amount of detail should be represented in the network model to allow for the effective characterization of contaminant flow. This does not mean that an all-pipes network model is required nor does it mean that a network model with only transmission lines would suffice. At a minimum, all parts of the water system that are considered critical from a security standpoint should be included in the model, even if they are on the periphery of the system. The following guidance drawn from the Initial Distribution System Evaluation (IDSE) Guidance Manual of the Final Stage 2 Disinfectants and Disinfection Byproducts Rule provides a reasonable **lower limit** for the level of detail required (U.S. EPA 2006b).

Most distribution system models do not include every pipe in a distribution system. Typically, small pipes near the periphery of the system and other pipes that affect relatively few customers are excluded to a greater or lesser extent depending on the intended use of the model. This process is called skeletonization. Models including only transmission networks (e.g., pipes larger than 12 inches in diameter only) are highly skeletonized; models including smaller diameter distribution mains (e.g., 4 to 6 inches in diameter) are less skeletonized. In general, water moves quickly through larger transmission piping and slower through the smaller distribution mains. Therefore, the simulation of water age or water quality requires that the smaller mains be included in the model to fully capture the residence time and potential

water quality degradation between the treatment plant and the customer. Minimum requirements for physical system modeling data for the IDSE process are listed below.

- At least 50 percent of total pipe length in the distribution system.
- At least 75 percent of the pipe volume in the distribution system.
- All 12-inch diameter and larger pipes.
- All 8-inch diameter and larger pipes that connect pressure zones, mixing zones from different sources, storage facilities, major demand areas, pumps, and control valves, or are known or expected to be significant conveyors of water.
- All 6-inch diameter and larger pipes that connect remote areas of a distribution system to the main portion of the system or are known or expected to be significant conveyors of water.
- All storage facilities, with controls or settings applied to govern the open/closed status of the facility that reflects standard operations.
- All active pump stations, with realistic controls or settings applied to govern their on/off status that reflects standard operations.
- All active control valves or other system features that could significantly affect the flow of water through the distribution system (e.g., interconnections with other systems, pressure reducing valves between pressure zones).

### Network Model Demand Patterns

The movement of water through a distribution system is largely driven by water demands (consumption) throughout the system. During higher demand periods, flows and velocities generally increase and vice versa. Demands are usually represented in a network model by daily averaged or typical demands at most nodes with (a) global or regional demand multipliers applied to all nodes to represent periods of higher or lower demand, and (b) temporal demand patterns to define how the demands vary over the course of a day.

Ideally, the demand at each node would be calculated based on recent billing data. However, in some network models, demands across a large area have been aggregated and assigned to a central node. When building a model, each demand should be assigned to the node that is nearest to the actual point of use, rather than aggregating the demands and assigning them to only a few nodes. Both EPANET and most commercial software products allow the user to assign multiple demands to a node with different demands assigned to different diurnal patterns. For example, part of the demand at a node could represent residential demand and utilize a pattern representative of residential demand. Another portion of the demand at the same node could represent commercial usage and be assigned to a representative commercial diurnal water use pattern.

### Network Model Calibration/Validation

Calibration is the process of adjusting network model parameters so that simulated outputs generally reflect the true behavior of the system. Validation is the next step after calibration, in which the calibrated model is compared to independent data sets (i.e., data that was not used in the calibration phase) in order to ensure that the same model is valid over a wide range of conditions. There are no formal standards in the water industry governing how closely the simulated results need to match field results, nor is there formal guidance on the amount of field data that must be collected. Calibration methods that are frequently used include roughness (c-factor) tests, hydrant flow tests, and tracer tests. Simulation results for pressure, flow and tank water levels are compared to field data collected from SCADA systems or special purpose data collection efforts.

The IDSE Guidance Manual stipulates the following minimum criteria in order to demonstrate calibration: “The model must be calibrated in extended period simulation for at least a 24-hour period. Because storage facilities have such a significant impact upon water age and reliability of water age predictions throughout the distribution system, you must compare and evaluate the model predictions versus the actual water levels of all storage facilities in the system to meet calibration requirements.” Thus, the water utility should calibrate the network model so that it is confident that the network model adequately reflects the actual behavior of the water system. Some general guidelines for calibration/validation are shown below:

- If the model has been actively in operation for several years and has been applied successfully in a variety of extended period simulation situations, then further substantial calibration might not be necessary. However, even in this case, it is prudent to demonstrate the validity of the model by comparing simulations to field measurements such as time-varying tank water levels and/or field pressure measurements.
- If the model has been used primarily for steady state applications, then further calibration/validation emphasizing extended period simulation is needed.
- If the model has been recently developed and not undergone significant application, then a formal calibration/validation process is needed.

### Network Model Tanks

Most water distribution system models use a “complete mixing” tank representation that assumes that tanks are completely and instantaneously mixed. EPANET (and most commercial modeling software models) allow for alternative mixing models such as last in-first out (LIFO), first in-first out (FIFO), and compartment models. If a utility has not previously performed water quality modeling, they might not have determined the most appropriate tank mixing model for each tank. Since the tank mixing model can affect simulations of the fate and transport of contaminants, and thus the sensor placement decisions, tank mixing models should be specified correctly in the network model.

## Network Model Testing

The final step in preparing the model is to put it through a series of candidate tests. Following is a list of potential tests that should be considered.

If the model was developed and applied using a software package other than EPANET, then following its conversion to EPANET, the original network model and the new EPANET network model should be run in parallel under EPS and the results compared. Both simulations should give virtually the same or similar results. Comparisons should include tank water levels and flows in major pipes, pumps and valves over the entire time period of the simulation. If there are significant differences between the results, then the EPANET network model should be modified to better reflect the original network model or differences should be explained and justified.

The EPANET network model should be run over an extended period (typically 1 to 2 weeks) to test for sustainability. In a sustainable model, tank water levels cycle over a reasonable range and do not display any systematic drops or increases. Thus, the range of calculated minimum and maximum water levels in all tanks should be approximately the same in the last few days of the simulation as they were in the first few days. Typically, a sustainable model will display results that are in a dynamic equilibrium in which temporal tank water level and flow patterns will repeat themselves on a periodic basis.

If the water system has multiple sources, then the source tracing feature in EPANET should be used to test the movement of water from each source. In most multiple source systems, operators generally have a good idea as to how far the water from each source travels. The simulation results should be shown to the knowledgeable operators to ensure that the model is operating in a manner that is compatible with their understanding of the system.

In order to determine travel times, the network model should be run for a period of 1 to 2 weeks using the water age option in EPANET. Since the water age in tanks is not usually known before modeling, a best guess (not zero) should be used to set an initial water age for each tank. Then after the long simulation, a graph of calculated water age should be examined for each tank to ensure that it has reached a dynamic equilibrium and is still not increasing or decreasing. If the water age is still systematically increasing or decreasing, then the plot of age for each tank should be visually extrapolated to estimate an approximate final age and that value should be reinserted in the model as an initial age, and the model rerun for the extended period. Water age should be investigated for reasonableness. For example, are there areas where water age seems unreasonably high? This exercise will also help to define a reasonable upper limit for the simulation duration.

Following these test runs, any identified modifications should be made in the network model to ensure that it runs properly. Many utilities will not be able to make all of the above

described modifications to their network model. In that case, sensor placement optimization can still be applied; however the overall accuracy of the results will be questionable and should only be considered applicable to the system as described by the network model.

## Sensor Characteristics

In addition to a network model, other input data are needed to run sensor placement optimization tools. Characterization of sensor behavior is required to predict the performance of a CWS; in particular, the sensor type, detection limit, and accuracy need to be specified. For example, the analysis can specify a contaminant-specific detection limit that reflects the ability of the water quality sensors to detect the contaminant. Alternatively, the analysis can assume perfect sensors that are capable of detecting all non-zero concentrations of contaminants with 100% reliability. The latter assumption, though not realistic, provides an upper bound on realistic sensor performance. A slightly more realistic modeling assumption is to assume a detection limit for sensors: the sensor is 100% reliable above a specified concentration, but, below that concentration the sensor always fails to detect the contaminant.

In order to quantify detection limits for water quality sensors, one must indicate the type of water quality sensor being used, as well as the disinfection method used in the system. Generally, water quality sensors are more sensitive to contaminants introduced into water disinfected with chlorine than chloramines. As a result, contaminant detection limits might need to be increased in the design of a sensor network for a chloraminated system; and, in particular, chlorine residual might not be an effective parameter for chloraminated systems.

Ongoing pilot studies for EPA's Water Security Initiative use a platform of water quality sensors, including free chlorine residual, total organic carbon (TOC), pH, conductivity, oxidation reduction potential (ORP), and turbidity (U.S. EPA 2005c). The correlation between contaminant concentration and the change in these water quality parameters can be estimated from experimental data, such as pipe loop studies (Hall et al. 2007; U.S. EPA 2005b). Of these parameters, chlorine residual and TOC seem to be most likely to respond to a wide range of contaminants.

Detection limits for water quality sensors can be defined in terms of the concentration which would change one or more water quality parameters enough to be detected by a water utility operator or an event detection system (e.g., Cook et al. 2005; McKenna et al. 2006; McKenna et al. 2008). A utility operator might be able to recognize a possible contamination incident if a change in water quality is significant and rapid. For example, if the chlorine residual decreased by 1 mg/L, the conductivity increased by 150  $\mu\text{Sm/cm}$ , or TOC increased by 1 mg/L.

It is possible to represent the accuracy of sensors in terms of the likelihood of sensor failure. For example, Berry et al. (2009) explored sensor placement for sensors with known

false negative and false positive rates. These rates might also be parameterized by concentration level. However, such assumptions make the sensor placement problem significantly harder to solve on desktop computers.

## The Design Basis Threat

A design basis threat identifies the type of threat that a water utility seeks to protect against when designing a CWS. In general, a CWS is designed to protect against contamination threats; however, there are a large number of potentially harmful contaminants and a myriad of ways in which a contaminant can be introduced into a distribution system. Some water systems might wish to design a system that can detect not only high impact incidents, but also low impact incidents that might be caused by accidental backflow or cross-connections. It is critical to agree upon the most appropriate design basis threat before completing the sensor network design.

Contamination incidents are specified by a specific contaminant(s), the quantity of contaminant, the location(s) at which the contaminant is introduced into the water distribution system, the time of day of the introduction, and the duration of the contaminant introduction. Given that it is difficult to predict the behavior of adversaries, it is unlikely that anyone will know, with any reasonable level of certainty, the specific contamination threats one might face. Most of these parameter values cannot be known precisely prior to an incident; therefore, the modeling process must take this uncertainty into account.

For example, probabilities can be assigned to each location in a distribution system indicating the likelihood that the contaminant would be introduced at that location. The default assumption is that each location is equally likely to be selected by an adversary (each has the same probability assigned to it). A large number of contamination incidents (an ensemble of incidents) are then simulated and sensor network designs are selected based on how well they perform for the entire set of incidents.

## Performance Measures for CWS

A sensor network design can be selected that best minimizes one of the following performance objectives, as estimated through modeling and simulation:

- the number of people who become ill from exposure to a contaminant
- the percentage of incidents detected
- the time to detection
- the length of pipe contaminated

Other objectives such as the costs of a CWS or the economic impacts to a water system could be considered as well. In order to quantify these objectives, a set of contamination incidents (an ensemble defined by the design basis threat) must be simulated.

Public health and economic impacts are contaminant-specific. Contaminants behave differently in water distribution systems: some can be modeled as tracers, but other contaminants might react with disinfectant residuals, attach to biofilms, or adsorb to pipe walls. These cases require more sophisticated models (Shang et al. 2008). Human health impacts are also contaminant-specific, and require assumptions about human consumption patterns: for example, estimates of the spatial and temporal distribution of the people that have been exposed; calculations of the number of people that might become ill according to contaminant-specific dose-response curves; and predictions of the time evolution of health impacts (Murray et al. 2006b).

It is also possible to consider multiple objectives in a sensor network design analysis. If one has several priorities in the area of performance measures, these can be accounted for by assigning the relative importance (weight) to each measure. In addition, one might have non-security related objectives that could also be considered. For example, one might wish to co-locate sensors with current monitoring stations that are in place to meet regulatory requirements.

## Utility Response to Detection of Contamination Incidents

In designing the WS Initiative, EPA has said that “the key to an effective response to a water contamination threat is minimizing the time between indication of a contamination incident and implementation of effective response actions to minimize further consequences,” (U.S. EPA 2005a). Modeling the human response to the detection of a contamination incident is difficult because of the site-specific logistics of response and because of uncertainty in the confidence attributed to detection of contamination incidents.

The following response activities are likely following detection of potential contamination incidents (Bristow et al. 2006; U.S. EPA 2004):

- Credibility determination: Integrating data to improve confidence in detection; for example, by looking for confirmation from other sensor stations, or detection by a different monitoring strategy, and checking sensor maintenance records.
- Verification of contaminant presence: Collection of water samples in the field, field tests and/or laboratory analysis to screen for potential contaminants.
- Public warning: Communication of public health notices to prevent further exposure to contaminated water.

- **Utility mitigation:** Implementing appropriate utility actions to reduce the likelihood of further exposure, such as isolation of contaminated water in the distribution system or other hydraulic control options.
- **Medical mitigation:** Public health actions to reduce the impacts of exposure, such as providing medical treatment and/or vaccination.

Computational models of CWS performance typically make the assumption that there is a response time after which contaminants are no longer consumed or propagated through the network. Response time is the time between initial detection of an incident and effective warning of the population. The response time, then, is the sum of the time required to implement various activities, and is typically considered to be between 0 and 48 hours. A zero-hour response time is obviously infeasible but can be considered the best-case scenario, which reflects the upper bound on sensor network performance. Water utilities should assess their own emergency response procedures and their acceptable risk tolerance in terms of false negative and false positive responses in order to define a range of response times to be used in the network design analysis.

### Potential Sensor Locations

The primary physical requirements for locating sensors at a particular location are accessibility, electricity, physical security, data transmission capability, sewage drains, and temperatures within the manufacturer specified range for the instrumentation (ASCE 2004). Accessibility is the amount of space required for installation and maintenance of the sensor stations. Electricity is necessary to power sensors, automated sampling devices, and computerized equipment. Physical security protects the sensors from natural elements and vandalism or theft. Data transmission sends sensor signals to a centralized SCADA database via wireless cellular, radio, land-line, or fiber-optic cables. Sewage drains are

required to dispose of water and reagents from some sensors. Temperature controls might be needed to avoid freezing or heat damage.

Most drinking water utilities can identify many locations that satisfy the above requirements, such as pumping stations, tanks, valve stations, or other utility-owned infrastructure. Many additional locations might meet the above requirements for sensor locations or could be easily and inexpensively adapted. Other utility services, such as sewage systems, own sites that likely meet most of the requirements for sensor locations (e.g., collection stations, wastewater treatment facilities, etc.). In addition, many publicly-owned sites could be easily adapted, such as fire and police stations, schools, city and/or county buildings, etc. Finally, many consumer service connections would also meet many of the requirements for sensor placement, although there could be difficulties in securing access to private homes or businesses. Nevertheless, the benefit of using these locations might be worth the added cost. Compliance monitoring locations could also be feasible sites.

The longer the list of feasible sensor sites, the more likely one is to design a high-performing CWS. With that said, the authors' experience with water utilities suggests that for various reasons, some locations truly are infeasible. Therefore, the authors typically restrict the sensor placement analysis to three sets of feasible locations: all locations (represented by the nodes in a network model), all public-owned facilities, and all utility-owned facilities. These lists can be further refined by field verification of sites to ensure that sites meet all of the requirements discussed here. Finally, it is important to note that field verification is needed after selection of sites in order to verify that the hydraulics at the site match the hydraulics simulated in the network model. Some models might not be detailed enough to show service connections and thus field verification is needed to show that the sensor can be installed on the correct line.

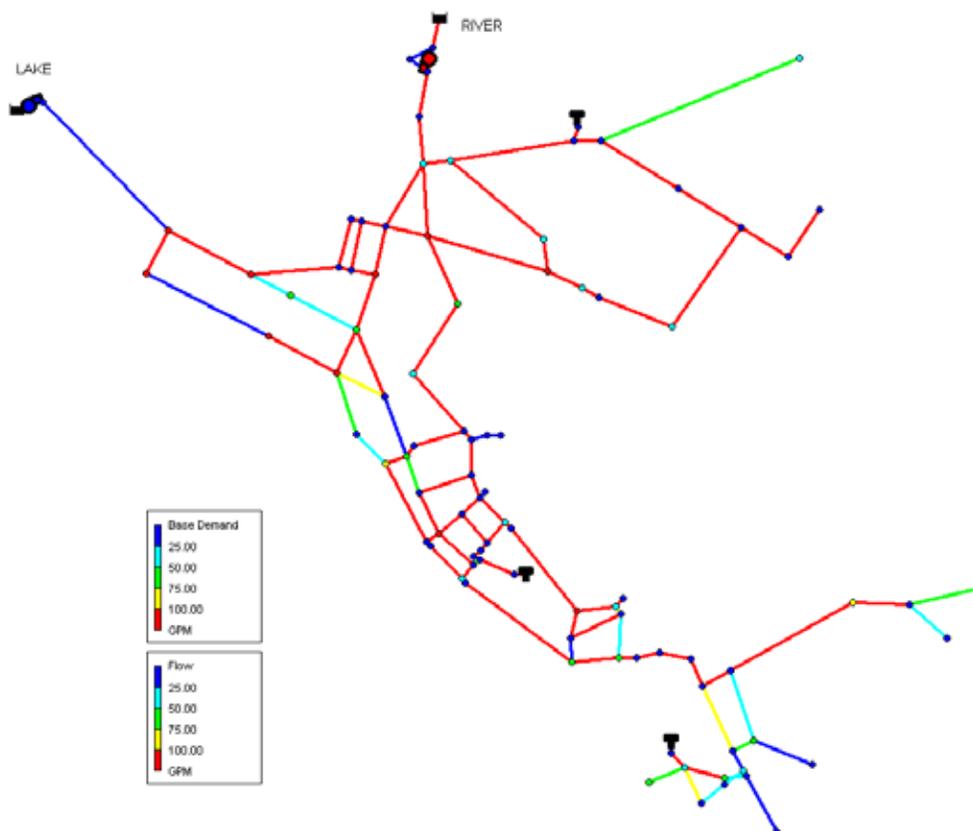
# TEVA Decision Framework: Decision Process

This chapter describes the second part of a decision framework for CWS design (Murray et al. 2008b). This decision framework is composed of a modeling process and a decision-making process. The modeling process is described in Chapter 2, and its goal is to accurately describe in a conceptual model the characteristics of the planned CWS. The decision process is an incremental approach for applying optimizers in order to generate a sequence of sensor network designs, the merits of which are then compared and contrasted. The ultimate goal is to enable utilities to understand the significant public health and cost tradeoffs between designs, and to ultimately select the one that best meets the goals of the utility.

Optimization methods can be used to determine sensor network designs for water distribution systems. However, there are a series of questions that must be answered prior to the optimization regarding the type of sensors, the design basis threat, and the utility response time. Thus, there is uncertainty associated with these utility decisions and their impact on the final CWS design.

The decision process begins by finding a sensor placement under ideal conditions and simplifying assumptions. The assumptions are then removed one by one in order to make the results more realistic. At each iteration, the performance of the given sensor network design is compared quantitatively and visually with previous designs in order to understand what has been gained or lost with each assumption.

This process is illustrated and discussed in this chapter with an analysis of an example water distribution system shown in **Figure 3-1**: EPANET Example 3. This example network is supplied by two surface water sources — a lake provides water for the first part of the day and a river for the remainder of the day. Example 3 has 3 tanks, 2 pumps, and serves approximately 79,000 people (assuming a usage rate of 200 gallons per person per day). This simple example is used to illustrate the decision process, but this same approach has been applied to larger networks serving up to several million customers (see for example, Murray et al. 2008b).



**Figure 3-1.** Map of the network model used for the sensor placement analysis. The system is served by both a river and a lake. The colors of the nodes indicate the relative base demand and the colors of the pipes indicate the bulk flow rates.

## A Preliminary Sensor Network Design

A preliminary sensor network design is generated using the TEVA-SPOT optimization software to illustrate the steps involved. First modeling decisions must be made.

### Modeling Information

Assume that the following information was collected by the water utility during the modeling process for this CWS design application:

- Utility Network Model. EPANET Example 3 network is used for sensor placement analysis in this chapter. This network has 92 junctions, 2 reservoirs, 3 tanks, 117 pipes, and 2 pumps. The sources include a river and a lake which provides water to the system for 14 hours a day. The average residence time in the network is 13.5 hours, while the maximum residence time is 130 hours. Assuming a typical usage rate of 200 gallons per person per day, the population served is 78,823 people. This model simulates seven days of flow.
- Sensor Characteristics. The sensor stations are multiple parameter water quality sensor stations (modeled with contaminant-specific detection limits that reflect the ability of water quality sensors to detect the chemical contaminant). The sensors are assumed to perform with 100% accuracy (i.e., no failures).
- Design Basis Threat. The design basis threat is the scenario in which a large quantity of a highly toxic chemical contaminant is injected over a 1-hour period starting at midnight with a rate of 17,333 mg/min. The location of the attack is not known, so every location in the model is considered a possible source. Thus, 92 nodes were considered potential points of entry, resulting in a total of 92 contamination incidents in the design basis threat.
- Performance Measures. Public health impacts that might result from a contamination incident are the highest priority and therefore the performance measure selected is the number of people who become ill from exposure to a contaminant (hereafter, referred to as PE).
- Utility Response. It is assumed that it would take two hours for the utility to respond effectively to a positive detection, eliminating further exposures. Note that two hours is quite optimistic and more realistic response times could vary from 6 to 24 hours.

- Potential Sensor Locations. It is assumed that there are 20 locations that are feasible sites for locating sensors, made up of public and utility-owned facilities. These locations are specified by nodes 208, 209, 1, 169, 143, 231, 219, 101, 184, 127, 275, 129, 125, 145, 237, 20, 183, 601, 271, and 189.
- Number of Sensor Locations. Three sensor locations will be selected.

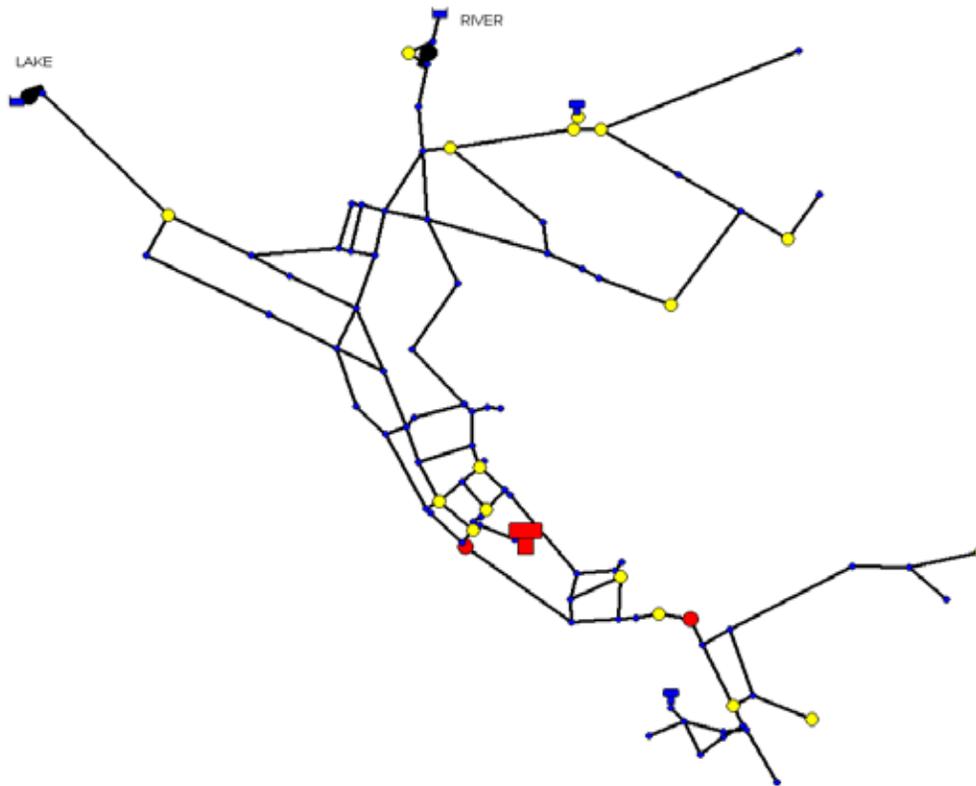
### Quantifying the Potential Consequences

A variety of impact measures are used to compare and contrast sensor network designs in this example. PE is the number of people sickened due to the exposure, EC is the number of pipe feet contaminated, MC and VC are the mass of contaminant and volume of contaminated water removed from the system by consumer demand, TD is the time of detection, and NFD is the number of failed detections (shown here as a percentage).

**Figure 3-2** shows the distribution of public health impacts for the set of chemical incidents. It was assumed that there were no sensors in the system to detect the contaminants and that the public health system or the water utility had taken no actions to reduce the impacts. For each of the 92 incidents that were simulated, the public health impacts were calculated. The majority of contamination incidents result in less than 5% of the population being impacted but there were four incidents that impacted more than 20,000 people. Over all the 92 incidents, on average 6,444 people would become ill (or 8% of the population), with a median of 4,041 people, and a maximum of 21,244 people. Node 203 serves more than 32,000 people; injections at this or at one of the nearby nodes (201, 199, and 173) impacted a large number of people. Similarly, the average length of pipe contaminated was 9.6 miles (50,527 feet), with a median of 6.6 miles, and a maximum of 38 miles (see **Table 3-1**).

The mean values can be interpreted in the following way: if one randomly selected a location from which to introduce the chemical contaminant, one could expect that 6,444 people would become ill and 9.6 miles of pipe would be contaminated.





**Figure 3-3.** Map of the network model with the three selected sensor locations in red (one tank and two nodes) and the remaining 17 potential locations in yellow.

**Table 3-2.** Summary of impact statistics resulting from the chemical incidents with three optimally placed sensors. PE is the number of people sickened after exposure, EC is the number of pipe feet contaminated, and MC and VC are the mass of contaminant and volume of contaminated water removed from the system by consumer demand.

Performance Measure/Statistic	Mean	25 <sup>th</sup>	50 <sup>th</sup> (median)	75 <sup>th</sup>	100 <sup>th</sup> (max)
PE (people)	4,318	741	2,158	4,985	21,244
EC (pipe feet)	39,996	1,960	18,574	61,340	155,250
MC (mass)	917,344	681,524	978,488	1.14E6	2.02E6
VC (gallons)	717,662	19,810	134,453	1.19E6	2.98E6
TD (minutes)	4359	120	180	10,080	10,080
NFD (fraction)	0.42	0	0	1	1

**Table 3-3** lists the sensors selected when optimizing the six performance metrics. Note that the designs for the PE, MC, and VC metrics were very similar and tended to place all three sensors near the center of the network. The locations were not identical, but upon inspection of the map, one would find that they are very close. In contrast, the designs for TD and NFD placed sensors near the end of the flow

paths in the southern and eastern boundaries of the network. Also notably different, the design metric EC minimizes the extent of contamination and placed one sensor near the lake source, one in between the river source and the northeastern tank, and one near the central tank. The selection of the most appropriate performance metric, then, is quite important.

**Table 3-3.** Sensor designs for six performance measures. PE is the number of people sickened after exposure, EC is the number of pipe feet contaminated, and MC and VC are the mass of contaminant and volume of contaminated water removed from the system by consumer demand.

Performance Measure	Selected Optimal Sensor Locations
PE	Nodes 1, 184, 209
EC	Nodes 1, 20, 101
MC	Nodes 1, 237, 209
VC	Nodes 1, 189, 237
TD	Nodes 143, 219, 231
NFD	Nodes 143, 219, 231

### A More In-Depth Investigation of Sensor Network Design

In the last section, all of the input data was well known. Suppose that the utility did not know how many sensor stations to install and wanted to consider anywhere between 1 and 10 sensor stations. In addition, the utility wanted to protect against a large scale biological contamination scenario in addition to the chemical scenario. The utility also wanted to consider the extent of contamination as an optimization objective. Finally, the utility was uncertain about the response time and wanted to examine a range between 0 and 12 hours (0 is analyzed in order to understand the best case scenario). The new range of parameter values to be considered is listed below.

#### Modeling Decisions

- Utility Network Model. The same network is used — EPANET Example 3.

- Sensor Characteristics. The sensor stations are multiple parameter water quality sensor stations (modeled with contaminant-specific detection limits that reflect the ability of water quality sensors to detect the chemical and/or biological contaminant).
- Design Basis Threat. The system is designed for a large quantity of a highly toxic chemical contaminant injected over a 1 hour period and for a large quantity of an infectious biological agent injected over a 24-hour period, both starting at midnight. The location of the attack is not known, and so every location in the model is considered a possible source. Thus, 184 contamination incidents are simulated.
- Performance Measures. In this analysis, both PE and EC are considered.
- Utility Response. It was assumed that it would take between 0 and 12 hours for the utility to respond effectively to a positive detection, eliminating further exposures. The 0 hour response case is considered even though it is not physically possible because it gives an upper bound on performance.
- Potential Sensor Locations. It is assumed that sensor locations can be selected from all 92 nodes or from the set of 20 feasible locations identified in the last section.
- Number of Sensor Locations: 1–10 sensor locations will be selected.

**Table 3-4** lists all of the sensor network designs that will be generated as part of the investigation in this chapter. Designs are created both for the chemical and biological incidents, four performance measures (PE, NFD, TD, and EC), and for four different response times (0, 2, 6, and 12 hours). For the chemical the accurate detection limit is assumed to be 0.001 mg/L; for the biological, 1,000 organisms/L. The list of potential sensor locations was first allowed to be all possible locations, and later restricted to the 20 locations determined by the utility.

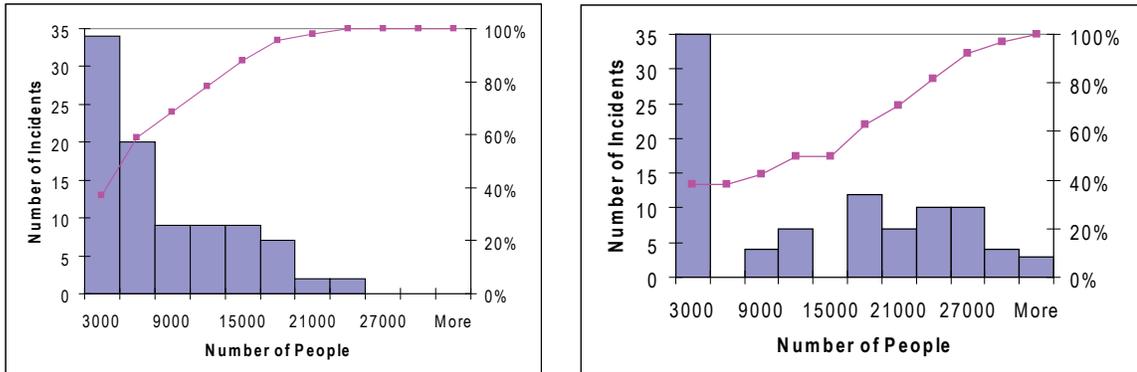
**Table 3-4.** List of sensor designs and associated parameter values analyzed in TEVA-SPOT decision-making application. PE is the number of people sickened after exposure, EC is the number of pipe feet contaminated, and MC and VC are the mass of contaminant and volume of contaminated water removed from the system by consumer demand.

Sensor Design	Design Basis Threat	Performance Objective	Response Time (hours)	Detection Limit (org/L)	Potential Locations
1	Chemical	PE	2	0.001	ALL
2	Biological	PE	2	1,000	ALL
3	Biological	NFD	2	1,000	ALL
4	Biological	TD	2	1,000	ALL
5	Biological	EC	2	1,000	ALL
6	Biological	PE	0	1,000	ALL
7	Biological	PE	6	1,000	ALL
8	Biological	PE	12	1,000	ALL
9	Biological	PE	2	1,000	20 Locs

### Quantifying the Potential Consequences

**Figure 3-4** shows the predicted distribution of impacts for the chemical and biological incidents when there are no sensors in the system to detect the contaminants and where the public health system and/or the water utility have taken no actions to reduce the magnitude of impacts. For each contaminant, 92 incidents were simulated, and the public health impacts were calculated. **Table 3-5** lists the statistics for the number of people sickened, the extent of contamination, the mass of contamination consumed, and the volume of contaminated water consumed.

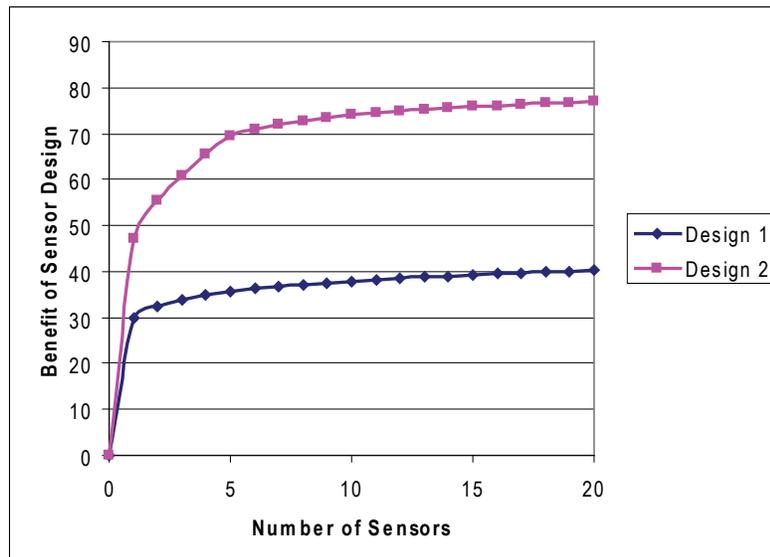
Note the difference in impacts between the chemical and biological scenarios. The average number of people made ill from the chemical threat ensemble was 6,444, and the average was 12,383 for the biological threat ensemble. The max case incident impacted 21,244 people for the chemical threat ensemble and 31,788 for the biological threat ensemble.



**Figure 3-4.** Histograms of public health impacts resulting from the chemical (left) and biological (right) incidents

**Table 3-5.** Summary of impact statistics resulting from the chemical and biological incidents in the absence of a CWS. For each impact measure, this table shows the mean impact, as well as various percentiles of the distribution of impacts for all simulated incidents. PE is the number of people sickened after exposure, EC is the number of pipe feet contaminated, and MC and VC are the mass of contaminant and volume of contaminated water removed from the system by consumer demand.

Chem Performance Measure/Statistics	Mean	25 <sup>th</sup>	50 <sup>th</sup> (median)	75 <sup>th</sup>	100 <sup>th</sup> (max)
PE (people)	6,444	1,460	4,041	10,335	21,244
EC (pipe feet)	50,527	5,100	34,629	82,390	200,280
MC (mass)	1.12E6	9.91E5	1.09E6	1.28E6	2.02E6
VC (gallons)	3.14E7	42,817	3.69E6	7.25E7	9.55E7
Bio Performance Measure/ Statistics	Mean	25 <sup>th</sup>	50 <sup>th</sup> (median)	75 <sup>th</sup>	100 <sup>th</sup> (max)
PE (people)	12,833	1,720	10,887	22,811	31,778
EC (pipe feet)	55,395	12,444	39,603	82,390	200,281
MC (mass)	2.04E13	2.04E13	2.08E13	2.10E13	2.27E13
VC (gallons)	1.93E7	642,438	7.79E6	2.29E7	9.20E7



**Figure 3-5.** Performance curve showing the tradeoff between the number of sensors and the benefit provided (in terms of the percent reduction in illnesses for a given number of sensors). Design 1 is based on the chemical threat ensemble. Design 2 is based on the biological threat ensemble.

### Comparison of Design Basis Threats

Ideally, a sensor network design would be based on a very large threat ensemble (set of contamination incidents). However, in practice, computer memory limits the number of incidents that can be included. In this case, the chemical and biological incidents were separated into two threat ensembles and two different sensor network designs were generated. In this section, the biological and chemical designs are compared to one another. The TEVA-SPOT toolkit software version 2.2 was used to generate two sensor designs, Designs 1 and 2 listed in **Table 3-4**. The first design was based on the chemical threat ensemble, and the second design was based on the biological threat ensemble. **Figure 3-5** illustrates the tradeoff between the number of sensors and the benefit provided by each of the two designs. The performance of each sensor network design is measured in the percentage reduction in the number of illnesses relative to the baseline case with no sensors. As the number of sensors increased, the benefit of the sensor network increased with diminishing marginal returns. Note that for a given performance level, fewer sensors were needed to detect the 24-hour biological contamination incidents than were needed for the 1-hour chemical incidents (i.e., to achieve a 40% reduction, 19 sensors were needed for the chemical incidents and only one sensor was needed for the biological incidents).

Why are these two curves different? The differences are not due to hydraulics or operations since they were the same for chemical and biological incidents. The contaminants were

both modeled as tracers, so the differences in the curves are not due to reaction with disinfectant residual or other materials. The differences in **Figure 3-5** result from three factors: the difference in the injection times (1 hour for the chemical versus 24 hours for the biological), the different detection limits for each contaminant (0.001 mg/L for the chemical and 1,000 organisms/L for the biological), and the toxicity characteristics of the contaminants (specifically, the potency of each attack measured for instance by the number of lethal doses introduced to the system and/or the slope of the dose-response curve for each contaminant). It is much more difficult to detect a quick pulse of highly toxic chemicals at low concentrations than a long slow pulse of less toxic biological organisms at higher concentrations.

**Figure 3-5** can be used to make an initial decision on the number of sensors. This decision can be refined later in the process after considering the effect of the various parameters on sensor network design performance. From looking at **Figure 3-5**, one can see that the greatest gains for both threat ensembles occurred with only a few sensors. With six sensors, Design 1 reduced the number of people sickened by 36% and Design 2 by 71%. Focusing on raw numbers rather than percentages, if the utility decided that the sensor network should be designed so that the mean number of people sickened for both threat ensembles should be less than 4,000 people, then five sensors would be needed for the biological threat ensemble and 11 sensors for the chemical threat ensemble.

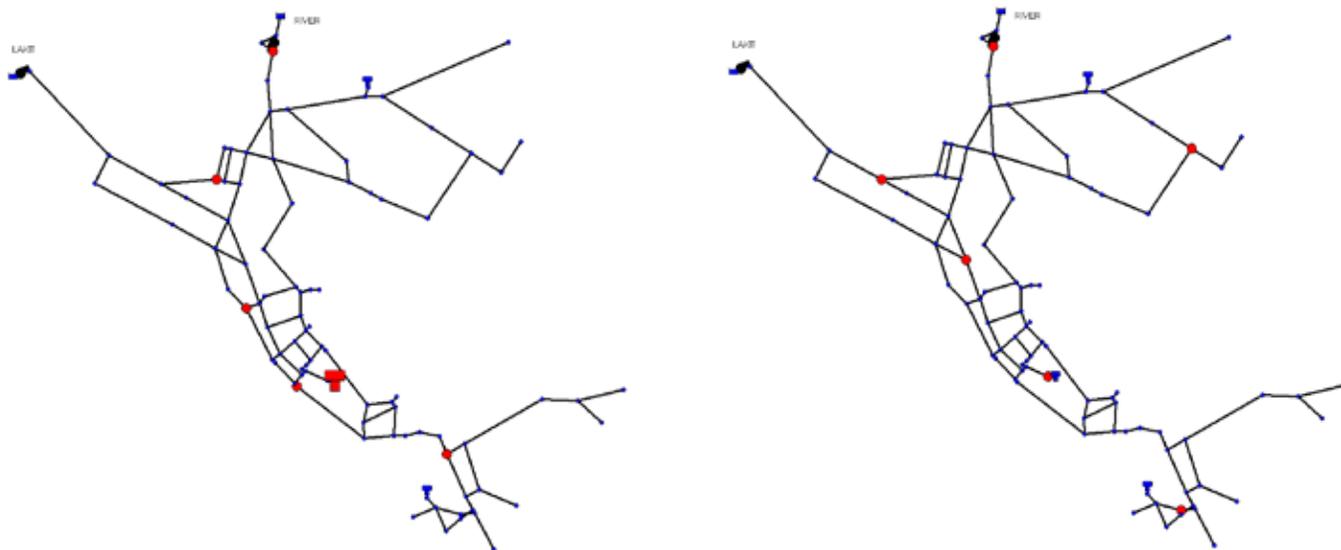
If six sensors were selected, Design 1 would include Nodes 61, 184, 191, 211, 263, and Tank 1. Design 2 would include Nodes 40, 61, 105, 113, 141, and 247. **Figure 3-6** shows the two sensor designs with 6 sensors each on spatial plots. One can see that the designs are different but there are some similarities. Node 61 is common to both designs (just below the river). Among the six selected nodes for Design 1, Node 211 (near the bottom of the map) provides the most benefit, but Node 40 is the most effective sensor location for Design 2 (near the central tank). It is challenging to look at locations on the map and try to determine how well they will protect the population. Therefore, a “regret analysis” was completed to answer the question: “If the chemical and biological incidents are equally likely to occur, which sensor design is preferable?” It is called a regret analysis because it reveals how much one might regret selecting the sensor design for the chemical threat ensemble when the biological threat ensemble actually occurs (or vice versa).

For this (and subsequent regret analyses), it was assumed that six sensors (or sensor stations containing multiple water quality parameters) were utilized as part of the CWS. The

results of the regret analysis are given in **Table 3-6**. For example, if the biological incident occurred, then Design 2 performed best. It reduced the number of people sickened by 71%, while Design 1 reduced that number by 66%. The error measure, or measure of regret, was calculated by summing the square of the differences between the performance measure of the given sensor design and the performance measure of the optimal sensor design. **Table 3-6** shows that both designs performed fairly well for both threat ensembles, yet Design 2 had a slightly lower regret measure.

**Table 3-6.** Regret analysis comparing Designs 1 and 2. Higher percentages reflect better performance.

Threat Ensemble/Sensor Design	Design 1	Design 2
Chemical	36%	34%
Biological	66%	71%
Error measure (regret)	.05	.02



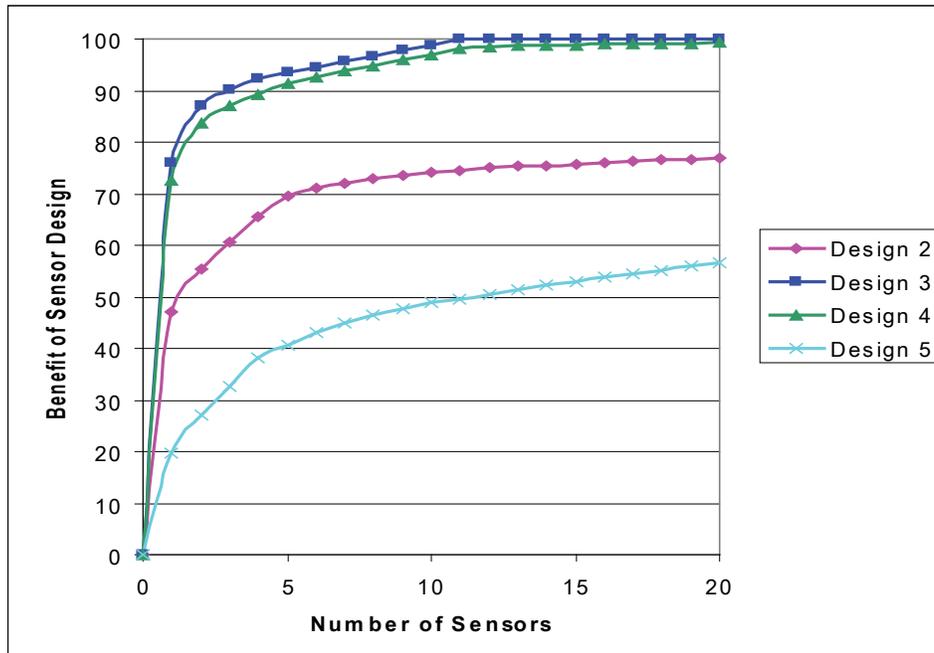
**Figure 3-6.** Design 1 with six sensors is shown on the left based on the chemical threat ensemble, while Design 2 with six sensors is shown on the right and is based on the biological threat ensemble.

### Comparison of Performance Objectives

Sensor network designs can also be developed based on other objectives, such as the number or percent of incidents not detected (NFD), the detection time (TD), and EC. Sensor Designs 2–5 (from **Table 3-4**) were optimized over the biological incidents, assuming a response time of 2 hours, and results are shown in **Figure 3-7** and in **Table 3-7**.

The tradeoff curves in **Figure 3-7** show that for this example network, it was much more difficult to reduce the number of illnesses (Design 2) or the length of contaminated pipe (Design 5) than it was the detection time (Design 4) or the number of failed detections (Design 3). The flow paths in the network were connected enough so that only 11 sensors were needed to detect all incidents. The flow was fast enough that detection times were fairly short. However, with a response time of two hours, the number of illnesses and the extent of contamination could not be reduced to zero regardless of the number of sensors.

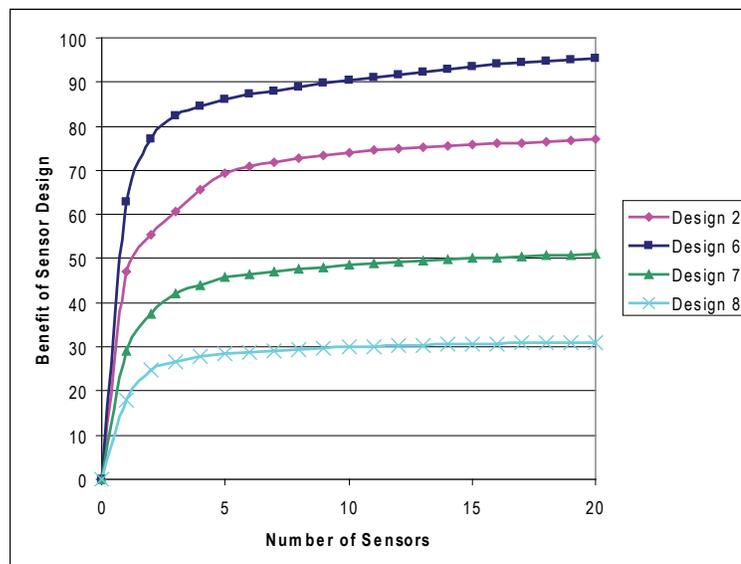
The regret analysis results are given in **Table 3-7**, helping to answer the question “If the chemical and biological incidents are equally likely to occur, which objective for sensor design is preferable?” The second column of **Table 3-7** shows the performance of Design 2 which minimized the number of illnesses over all incidents. For the biological threat ensemble, that design was able to reduce the average number of illnesses by 71%, the average number of failed detects by 84%, the average detection time by 83%, and the length of contaminated pipe by 43%. Design 2 (based on minimizing illnesses) and Design 5 (based on minimizing extent of contamination) had the lowest regret measures, performing well across all incidents. It is important to note that the result might not be the same for different utility networks. Subsequent analyses in this paper use the minimization of illnesses as the main objective for optimization.



**Figure 3-7.** Performance of four sensor network designs that minimize different performance objectives. Design 2 minimized the number of illnesses, Design 3 minimized the number of failed detections, Design 4 minimized the time of detection, and Design 5 minimized the length of contaminated pipe. All designs were optimized over the biological incidents.

**Table 3-7.** Regret analysis for four sensor network designs that minimize different performance objectives: number of illnesses, number of failed detections, time of detection, and the length of contaminated pipe (in terms of percent reduction). Higher percentages reflect better performance. A lower regret measure is better. PE is the number of people sickened after exposure. NFD is the number or percentage of incidents not detected. TD is the time of detection. EC is the number of pipe feet contaminated.

Performance Measure/Sensor Design	Design 2	Design 3	Design 4	Design 5
PE (bio)	71%	53%	55%	71%
NFD (bio)	84%	95%	95%	84%
TD (bio)	83%	92%	93%	83%
EC (bio)	43%	23%	27%	43%
PE (chem.)	34%	28%	29%	34%
NFD (chem)	76%	88%	88%	76%
TD (chem)	75%	85%	86%	75%
EC (chem)	38%	23%	27%	38%
Error measure (regret)	.24	.33	.27	.24



**Figure 3-8.** Performance of sensor network designs with four different response times: 0 hours (Design 6), 2 hours (Design 2), 6 hours (Design 7), and 12 hours (Design 8). All designs minimized the number of illnesses over all biological incidents. Designs defined in Table 3-4.

### Comparison of Response Times

Thus far, the analysis has been based on a utility response time of two hours. In this section, response times of 6 and 12 hours were added to the detection time. For comparison purposes and to establish the upper bound on sensor network performance, a response time of zero hours was also considered. These response times represent the time between detection by a sensor and an effective public order that halts further consumption of water, as described in Chapter 2. TEVA-SPOT was used to select the sensor locations that optimally minimize the mean number of illnesses, given one of four response times.

**Figure 3-8** demonstrates the tradeoff between the number of sensors and the likely benefits provided by Designs 2 and 6-8. Again, as the number of sensors increased, the benefit of the sensor network increased, and the benefit provided

by the first few sensors was significant. It is clear that as the response time increased, the overall performance of the sensor network decreased dramatically. With a residence time of only 13 hours in the network, the time available to reduce exposures was relatively short. Note that just by adding additional sensors (given the upper bound of 20), the benefits achieved at a given response time could never equal the benefits of a smaller response time. This figure shows the importance of reducing utility response time.

Although a utility will attempt to minimize its response time, the exact response time cannot be predicted precisely prior to an incident and could vary between 0 and 24 hours or more. How then should the response time parameter for sensor network design be selected? To answer this question, a regret analysis was performed as shown in **Table 3-8**. The regret analysis answered the question, “If the response times of 2,

6, and 12 hours were equally likely to occur, which sensor network design would be preferable in all cases?” Note that the zero case was eliminated as it would be impossible to achieve. **Table 3-8** shows that Design 2 had the lowest regret over all incidents, and therefore, a 2-hour response time was used for all further analyses.

### Sensors Restricted to Subsets of Locations

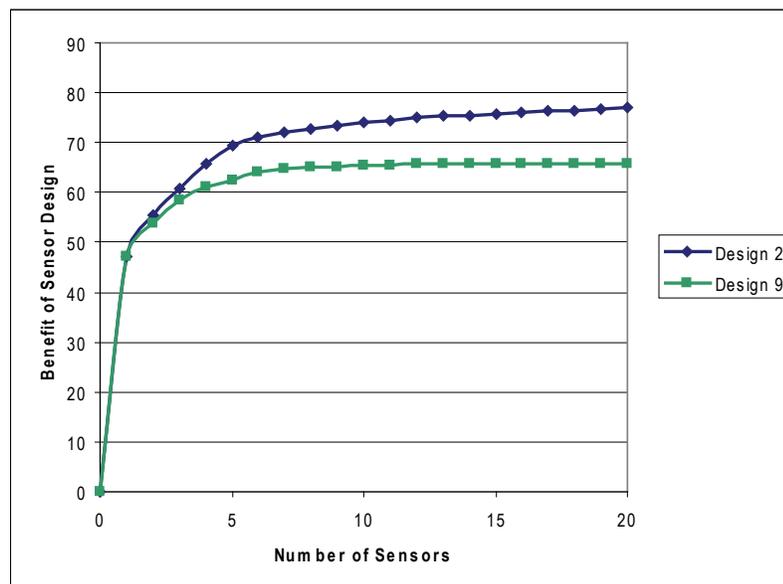
In this section, the set of possible sensor locations was restricted to the set of 20 locations: Nodes 208, 209, 1, 169, 143, 231, 219, 101, 184, 127, 275, 129, 125, 145, 237, 20, 183, 601, 271 and 189. These locations were randomly selected from the list of 92 total locations. In practice, however, a utility usually selects locations that are publicly owned and accessible to water utility staff 24 hours a day. For example, police and fire stations, public buildings, and utility facilities are good locations to consider. The effect

of restricting the potential locations to a smaller subset of locations on sensor placement is demonstrated below in **Figure 3-9** and the regret analysis is shown in **Table 3-9**. TEVA-SPOT was used to select the sensor locations that optimally minimized the mean number of illnesses, given the restricted locations. The regret analysis shows the performance lost by restricting the potential sensor locations.

The TEVA-SPOT software can also be used to rank the 20 locations in order of the benefit they provide to the sensor design. In this case, the locations were ranked from best to worst nodes as follows: 208, 189, 127, Tank 1, 101, 601, 143, 237, 271, 184, 169, 275, 209, 125, 145, 183, 219, 129, 20, and 231. If the utility decided to install 20 sensors, but could only install five this year because of budget constraints, then they could start with sensors at nodes 208, 189, 127, 101 and Tank 1, and install the rest later.

**Table 3-8.** Regret analysis for four sensor network designs based on minimizing illnesses for different response times (0, 2, 6, and 12 hours). Higher percentages reflect better performance.

Threat Ensemble/Sensor Design	Design 2	Design 6	Design 7	Design 8
Bio – 2 hr response	71%	66%	68%	68%
Bio – 6 hr response	45%	44%	47%	46%
Bio – 12 hr response	26%	26%	29%	29%
Chem – 2 hr response	34%	32%	33%	33%
Chem – 6 hr response	28%	28%	28%	28%
Chem – 12 hr response	26%	26%	26%	26%
Error measure (regret)	.037	.073	.043	.051



**Figure 3-9.** Performance of two sensor network designs. Design 2 selected sensors from all nodes in the model while Design 9 selected sensors from a list of 20 possible locations.

**Table 3-9.** Regret analysis for sensor network designs with potential sensor locations selected from all possible locations and 20 selected locations. Higher percentages reflect better performance.

Threat Ensemble/Sensor Design	Design 2	Design 9
Bio – 2 hr response	71%	64%
Bio – 6 hr response	45%	44%
Bio – 12 hr response	26%	27%
Chem – 6 hr response	34%	33%
Chem – 12 hr response	28%	27%
Chem – 24 hr response	26%	26%
Error measure (regret)	0.037	0.082

### Selecting the Sensor Network Design

Several sensor network designs have been presented based on chemical and biological contamination threat ensembles, different design objectives, and different assumptions about the utility and public health response time to a sensor signal, and the available locations for sensor station installation. The decision process followed above determined that the sensor design that performed best across all incidents considered in this analysis was the design for the biological threat ensemble, with a 2 hour response delay, accurate detection limits, and unrestricted locations, Design 2. In most cases, however, not all locations are feasible for placing sensors, either due to installation costs or operational restrictions. As a result, the designs must be restricted to a smaller set of feasible locations (Design 9).

Without sacrificing significant performance of the CWS design, a sensor network design was selected that meets the many goals of the water utility in designing the CWS. This sensor network design protected against both the chemical and biological incidents, performed well over a range of response times (0–12 hours) and performance objectives, and reduced costs by limiting the sensor locations to a subset of feasible facilities. Of the parameters considered in this report, the major factor in limiting overall sensor network performance was the utility and public health response time. No number of sensors can counteract the need for a fast

response. To a lesser yet still significant degree, restricting the potential sites for locating sensors to a small subset of all locations (e.g., only utility owned locations) also limited the performance of a sensor design.

Once a sensor network design is selected, there are a number of additional factors to consider. In the authors’ experience utilizing this design process, the steps listed above are only the first steps; additional modeling and decision-making iterations will follow. For example, once the locations are selected by the model, field investigations need to take place to ensure that the selected locations: (1) have the same hydraulics as described by the network model (the location is on the correct pipe), (2) allow enough space for locating and maintaining sensor stations, (3) can be accessed 24 hours a day, seven days a week by utility personnel. If not, certain locations might need to be removed from the list of feasible locations, and the optimization procedure re-run.

The framework for determining sensor placement presented here shows that a water utility can meet a variety of objectives by optimizing the CWS design. Specifically, while restricting locations to a set preferred by the water utility, sensor locations can be selected to match the performance characteristics of the utility sensor platforms, protect against a variety of contamination threats, optimize the performance measures important to the utility, and accommodate a range of likely utility response times.

## Real-World Applications

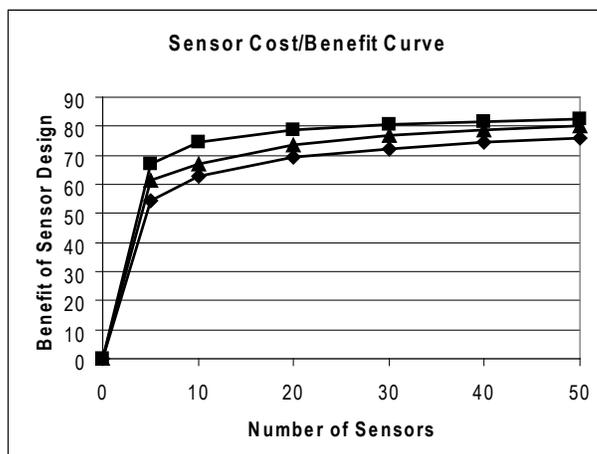
The TEVA-SPOT software was used to help nine partner utilities design sensor networks for Contamination Warning Systems (CWS). The modeling process described in Chapter 2 of this report was utilized: identifying the specific types of sensors to be deployed, the design objectives, and the possible locations of the sensor stations. Some utilities already had sensors in place; in such situations, the objective was to identify a few supplemental locations; however, most utilities had no existing sensors.

The decision-making process described in Chapter 3 was iterative and involved applying the optimization software to select optimal locations, and then verifying the feasibility of those locations with field staff. The software quantified the tradeoffs between the locations selected optimally by the software and the “near-by” locations selected by the utility for practical reasons. In addition, the software was used to develop cost-benefit curves for each utility, see **Figure 4-1**.

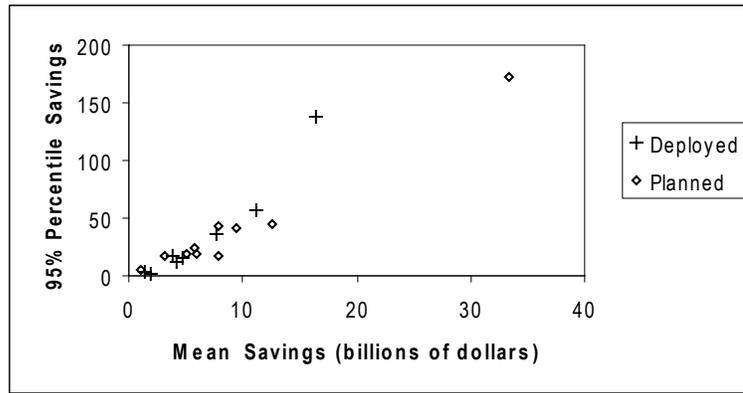
In order to quantify the benefits to each utility, a simulation study was completed (Murray et al. 2009). Two realistic terrorist contamination threat ensembles were considered: contamination with an infectious biological agent introduced over a 24-hour period; and contamination with a toxic

chemical agent introduced over a 1-hour period. **Figure 4-2** illustrates the estimated reduction in economic impacts due to the CWSs deployed or planned for the nine utilities. These economic savings can be attributed to the reduction in the number of fatalities that resulted from early detection and rapid response as part of the CWS.

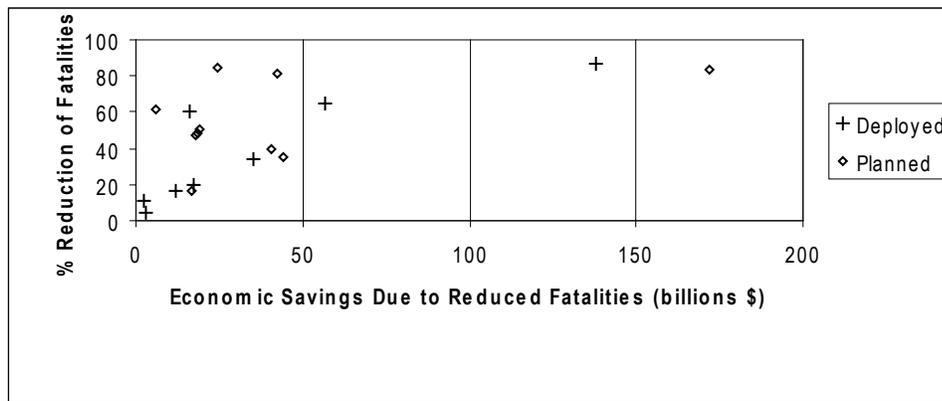
A large set of realistic contamination incidents was considered for these utilities; this plot shows the reduction of the mean and 95<sup>th</sup> percentile of the impact distribution. Fatalities were computed based on contaminant-specific data, after calculating how much contaminant customers at various locations and times throughout the network would consume. Economic impacts as a result of fatalities were computed using a Value of Statistical Life (VSL). VSL is the average value society is willing to pay to prevent a premature death. It does not refer to the value of an identifiable life, but rather the summed value of individual risk reductions across an entire population. In the Groundwater Rule, EPA used a value of \$6.3 million in 1999 dollars and \$7.4 million in 2003 dollars. To be conservative in this analysis, a VSL value of \$6.3 million was used (ATSDR 2001).



**Figure 4-1.** The cost-benefit curves for three utilities show that the benefits of a CWS design increased as the number of sensors (cost) increased.



**Figure 4-2.** The graph shows economic savings because of reduced fatalities (billion \$); it illustrates mean economic savings and corresponding savings for the 95<sup>th</sup> percentile of contamination incidents. Two data points, which represent biological and chemical contamination threat ensembles (not distinguished in the graph), are included for each utility.



**Figure 4-3.** The graph illustrates the relationship between the economic savings from the reduced fatalities and the percentage reduction of fatalities. These data points are for the 95<sup>th</sup> percentile impacts for fatalities and associated economic effects. For each utility, data points are included for a biological and chemical threat ensemble (not distinguished in the graph).

As **Figure 4-2** shows, a CWS can significantly reduce economic consequences of fatalities for biological and chemical incidents. Over the nine utilities, the mean savings are estimated to vary from \$1 billion to \$33.4 billion with a median of \$5.8 billion. However, because an informed terrorist would attempt to maximize the impact of an attack, the mean impacts might not be the best measure. Although the sensor placements were optimized to minimize mean impacts, **Figure 4-2** shows that 95<sup>th</sup> percentile economic savings were also significant: the 95<sup>th</sup> percentile savings range from \$1 billion to \$171.7 billion with a median of \$19 billion.

**Figure 4-3** shows the relationship between the percent reduction of economic impacts because of fatalities and the percentage reduction in fatalities. These are independent, but related measures for CWSs. Points at the top of **Figure 4-3** represent utilities that have a significant reduction in the number of fatalities. However, this percentage reduction is relative to the total number of fatalities without sensors (or, in some cases, with the set of existing sensors). Thus, the

reduction of economic impacts in these utilities could vary dramatically because of differences in the total number of fatalities in these systems.

Finally, economic impacts incurred by the water utility were estimated. Following a contamination incident, contaminants might be difficult to remove from pipe walls and fittings. In the worst cases, utilities might have to relin or replace contaminated pipes. Therefore, the CWS designs were evaluated to determine the fraction of the distribution network that might need to be replaced. A CWS can reduce the cost of replacement by enabling a utility to quickly contain the spread of a contaminant. This study estimated that using CWSs will reduce the expected decontamination and recovery costs for these nine utilities by up to \$340 million. For many utilities, these savings are greater than their annual operating budget.

In the rest of this chapter, case studies are presented for several partner water utilities, including Greater Cincinnati Water Works, New Jersey American Water, Tucson Water, and the City of Ann Arbor.

## Sensor Network Design for Greater Cincinnati Water Works

In 2006, EPA's Office of Water received funding to deploy CWSs at several U.S. utilities as part of the Water Security (WS) Initiative. The WS Initiative promotes a comprehensive CWS that is theoretically capable of detecting a wide range of contaminants, covering a large spatial area of the distribution system, and providing early detection in time to mitigate impacts (U.S. EPA 2005c). Components of the WS Initiative include: online water quality monitoring, consumer complaint surveillance, public health surveillance, enhanced security monitoring, and routine sampling and analysis. Information from these five monitoring strategies is combined to increase contaminant coverage, spatial coverage, timeliness of detection, and reliability of CWS performance.

Greater Cincinnati Water Works (GCWW) was selected as the first WS Initiative pilot city. EPA's report, "Water Security Initiative Cincinnati Pilot Post-Implementation Status," provides extensive detail on the contamination warning system installed at GCWW (U.S. EPA 2008). The sensor network design component of the project is summarized here. In this first WS pilot, EPA had an active and direct role in the design and implementation of the CWS. The online monitoring component for GCWW was designed to expand the existing monitoring capabilities. Prior to the WS Initiative, GCWW had forty chlorine analyzers, three pH meters, and two turbidimeters located at 22 sites in the distribution system that transmitted data over telephone lines to the utility's operations center.

The existing water quality monitoring did not meet all of the objectives of the WS Initiative; for example, spatial coverage, timely detection of contamination incidents, or degree of automation necessary for real-time detection. Therefore, additional sensor stations were installed at locations identified through an analysis using the TEVA-SPOT software. The sensor stations measured multiple water quality parameters including free chlorine, TOC, ORP, conductivity, pH, and turbidity. **Figure 4-4** shows a schematic of the sensor stations installed at GCWW. The new sensor network was designed to minimize average public health consequences over a large set of possible contamination incidents. The sensor network design process involved three steps: validating the utility network model, applying the TEVA-SPOT software, and field investigations to finalize sensor station locations.

To validate the model, a tracer study was performed in the field. A tracer (calcium chloride) was injected at four locations in the distribution system. Each injection consisted of at least six 1-hour pulses over a 24-hour period. Following each injection, conductivity meters were used to measure and record the conductivity signal at approximately

40 locations throughout each study region. The field data was then compared to model predictions in order to assess the accuracy of the model and identify needed improvements.

The validated model was utilized by the TEVA-SPOT software to help identify a set of optimal sensor locations. GCWW identified several hundred potential sensor locations that included all utility owned sites (including office buildings), all police and fire stations in the county, as well as certain schools and hospitals. Using geographic information systems (GIS), these facilities were identified in the utility network model by the nearest node. Initially, the design was based on selecting up to 30 sensors stations, although in the end, 17 stations were deployed. The utility located two stations at its treatment plants and the software was used to help identify the best locations for the remaining 15 stations.

The design and implementation was an iterative process. TEVA-SPOT was used to select a list of optimal locations from the list of several hundred potential locations. The hydraulic connectivity of each location was verified on GIS and/or AutoCAD® (Autodesk, Inc.) maps to ensure that the model representation of the facility was correct. Finally a site visit was conducted to locate the exact installation location within the facility, estimate the hydraulic retention time in the pipes from the distribution system main to the monitoring equipment inside the facility, and address any outstanding concerns with that specific location. Sites were also verified to ensure accessibility, physical security, available sample water and drainage, a reliable power supply, and data communications. If at any point a site was deemed to be unsuitable, it was discarded and the TEVA-SPOT analysis re-run.

The retention time from the distribution system main to the monitoring equipment inside the facility was calculated by taking the quotient of the service pipe volume and the water demand by that facility. The pipe volume is the product of the pipe radius squared, the length of the pipe and the constant  $\pi$ . If the retention time was greater than two hours, then a water bypass to the sensor station had to be installed (which is not always feasible). A service connection of smaller radius was considered favorable, as was choosing a sensor tap-in location near the point where the service connection met the building (to reduce pipe length). A retention time over two hours was not recommended, as it has a negative impact on utility response time, and might require adjustment to the TEVA-SPOT analysis.

All 17 sensor stations were installed at the locations determined by the above process and have been in operation since 2007. In addition, GCWW is testing the performance of event detection systems — automated data analysis tools that convert real-time water quality data into alarms that indicate the likelihood of contamination incidents.

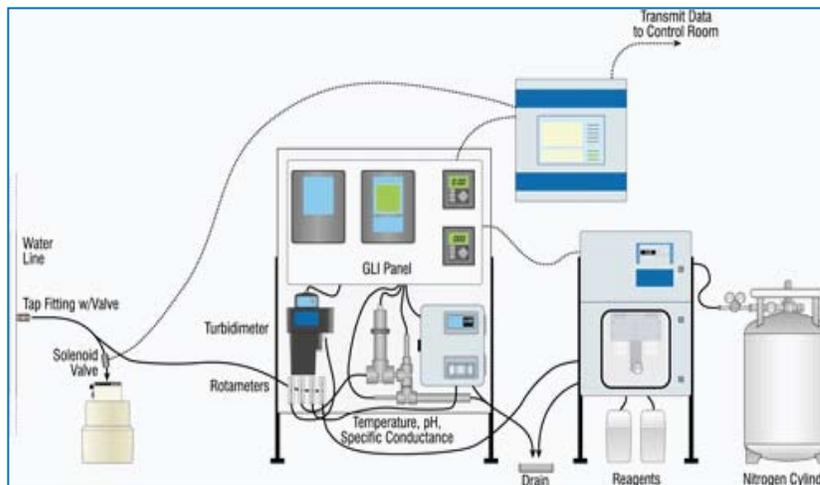


Figure 4-4. Schematic of one type of sensor station installed at the GCWW Pilot.

### Sensor Network Design for New Jersey American Water

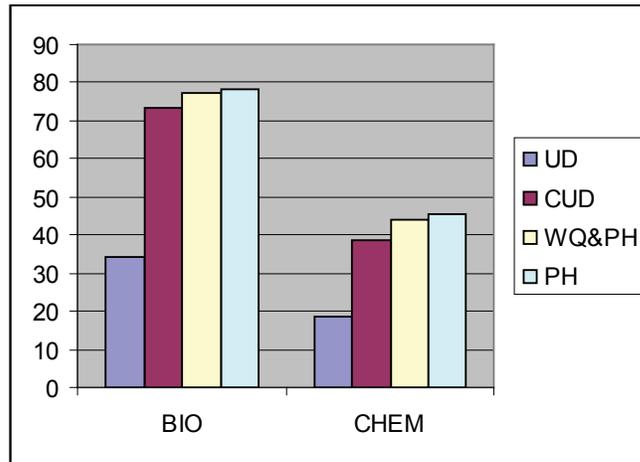
In the Burlington-Camden-Haddon system of New Jersey American Water (NJAW), eleven online monitors were installed as part of a collaborative study between EPA, U.S. Geological Survey (USGS) and NJAW. The purpose of the study was to understand the field performance of water quality monitors and the normal variability in background water quality, as well as to identify calibration requirements arising from fouling and sensor drift. These practical lessons learned would later inform the installation of online water quality sensors to support contamination warning systems. In addition, the data gathered over several years at these locations was used to help develop the CANARY event detection software.

Prior to this study, USGS had already worked with NJAW to install two sensor stations to monitor source water as well as two stations in the distribution system. The TEVA-SPOT software was used to select an additional seven monitoring locations in the distribution system. The three main objectives for sensor design were:

1. To obtain accurate measurements of the true range and variation in water quality in the AW distribution system.
2. To provide protection and early warning of contamination events.
3. To meet the additional needs and interests of NJAW (operational needs, costs, ease of access and maintenance, etc.).

Because one of the goals of this study was to better understand the variability of water quality within the distribution system, EPANET simulations were performed to predict chlorine residuals throughout the distribution system over a 10-day period, and the nodes were separated into three categories of low, medium, or high chlorine variability. Low variability nodes had a standard deviation of chlorine concentration in the lowest 33%; medium variability nodes fell between 33 and 66%; and high variability nodes were in the upper 33% of nodes.

NJAW and USGS provided a list of seven locations where they wanted to locate sensors in the distribution system. EPA used the TEVA-SPOT software to analyze the expected performance of this “utility design” (UD), and to select three additional designs for comparison. One design was optimized solely for public health protection (PH) and selected locations from all possible nodes in the model. Another design optimized for public health protection, but also for water quality (WQ) variability. The WQ&PH design required that two nodes have low variability in residuals, three with medium variability, and two with high variability. Finally the third design was a compromise between the UD and the WQ&PH (compromise utility design, CUD) that selected three locations from the list provided by the utility, and allowed the software to select the additional four locations from all the nodes in the model.



**Figure 4-5.** Benefit of each of the four sensor design for biological and chemical design basis threats. (UD=Utility Design, CUD=Compromise Utility Design, WQ&PH=Design for Water Quality and Public Health Objectives, and PH=Optimal Public Health Design).

**Figure 4-5** shows the performance of each of the four designs as measured against biological and chemical incidents. The UD was estimated to reduce the mean public health impacts associated with biological incidents by 34% and chemical incidents by 19%. In comparison, the optimal designs for public health protection reduced impacts by 48% and 45% respectively. **Table 4-1** shows for each sensor network design the number of sensors in each category of variability. With the information provided by this analysis, the utility was able to make a final decision on locating sensors that met both the objectives of the study to measure water quality but also the needs for public health protection as part of a future contamination warning system.

Following installation of YSI® (YSI Incorporated) multi-parameter sensors at various locations, the USGS was responsible for maintaining the sensor calibrations, manual collection and quality assurance of the data. Data has been collected for several years, and has subsequently been utilized by both EPA and Sandia National Laboratories to develop and refine tools for automated sensor data processing and event detection.

**Table 4-1.** Number of sensor locations in each chlorine variability category.

Sensor Design	Low Chlorine Variability	Medium Chlorine Variability	High Chlorine Variability
UD	2	3	4
PH	3	1	5
WQ&PH	3	3	3
CUD	3	3	3

### Sensor Network Design for Tucson Water

Tucson Water is an innovative and advanced municipal drinking water system that serves nearly 700,000 customers. Through an EPA Environmental Monitoring for Public

Access and Community Tracking (EMPACT) grant, online monitors have been providing near real-time water quality data to the public for several years. Tucson Water is currently considering the expansion of the online monitoring program to meet its security objectives.

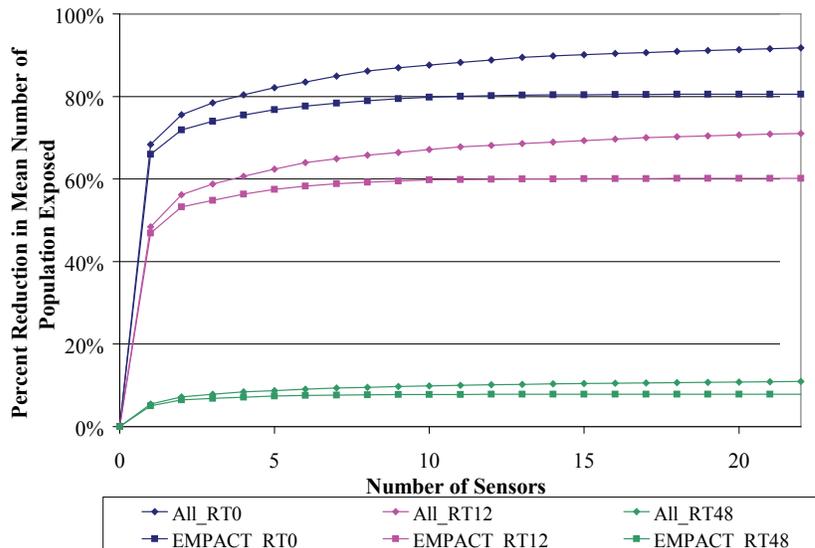
EPA began working with the Tucson Water utility in early 2005. The goal of the Tucson Water TEVA study was to identify new and/or existing EMPACT locations that could be used for monitoring for contamination incidents. The preliminary analysis was performed to answer the following questions:

- What are the best locations for sensor stations in the Tucson Water system?
- What are the best EMPACT locations for sensor stations?
- What are the tradeoffs between the two different sensor designs?

In order to use TEVA-SPOT, Tucson Water’s multiple pressure zone models had to be integrated into a single system-wide model. Separate pressure zone models are sufficient for many utility needs, but water security analyses require a systems-engineering approach, focusing on the entire distribution system as a whole.

Sensor designs were generated assuming that the goal of the monitoring program was to provide public health protection against a long release of a biological agent or a rapid release of a chemical from any service connection in the distribution system. The sensors were assumed to be water quality sensors capable of detecting changes caused by the two contaminants. The sensor designs are sensitive to response time, or the time it takes a utility to effectively respond to a positive detection. Therefore, response time was allowed to vary from 2 to 48 hours.

Optimal locations were selected from all possible locations in the model as well as from the 22 Tucson Water EMPACT monitoring locations based on minimizing mean public



**Figure 4-6.** Tradeoff curves for number of sensors versus percent reduction for all locations (diamond) and EMPACT locations (square) at three different response times.

health impacts from contaminant releases across the network model. All sensor designs were selected using response times of 0, 2, 6, 12, 24, and 48 hours. A total of 48 different sensor network designs were developed and analyzed.

**Figure 4-6** provides sensor tradeoff curves for sensor network designs based on the assumption that sensors can be located anywhere in the distribution system (diamond) or only at EMPACT sites (square). These curves demonstrate the tradeoff between the number of sensors and the performance of the sensor design, as measured by the percent reduction in mean public health impacts. The results for three different response times are also shown in **Figure 4-6**, where the blue line is a response time of 0, the pink line is a response time of 12 hours, and the green line is a response time of 48 hours. The optimal design can reduce the public health impacts from contamination from 10 to 92%. The EMPACT design reduces impacts from 6 to 80%. In addition, it is possible to improve the performance of the EMPACT design by selecting one or two additional locations that are not EMPACT locations.

Tucson Water is evaluating how to effectively use the designs recommended by TEVA-SPOT to create and implement a sustainable and cost-effective contamination warning system. In addition to the number and placement of sensors, Tucson Water is also evaluating vulnerability information provided by EPA researchers to better understand the sensitivity of response time in mitigating public health impacts following a contamination event.

### Sensor Network Design for the City of Ann Arbor

The City of Ann Arbor undertook a study in order to design an online monitoring program that could both minimize public health exposures resulting from a contamination incident and detect water quality degradation due to naturally

occurring processes, such as nitrification, iron corrosion, bacterial re-growth. The results of this study can be found in Skadsen et al. (2008) and are summarized briefly here.

The City of Ann Arbor’s water system provides treated water to about 130,000 customers and encompasses approximately 49 square miles. The average system demand is 15 MGD (million gallons per day) with a range from 7 to 30 MGD depending on the season. The Huron River bisects the City of Ann Arbor. The distribution system is divided into five major pressure districts that have elevated tanks and storage reservoirs to adequately serve the population over a varied topography. The pressure districts are typically operated independently, but include interconnects that are sometimes used to control pressure and flow. The distribution system has an estimated average retention time of two and one-half days, with a maximum of up to 10 days. This long retention time can sometimes result in water quality degradation. Although the distribution system piping consists mainly of cement lined ductile iron, there are areas of unlined cast iron pipe that remain in service. These areas are often heavily tuberculated resulting in problems with rusty water. The utility’s grab sample program addresses distribution system water quality and regulatory concerns.

The study involved four steps: (1) analysis of the distribution system, (2) parameter selection and instrument pilot testing, (3) estimation of costs, and (4) proposal of an online monitoring design. The analysis of the distribution system began with an assessment of the accuracy of the existing network model. Following a series of improvements to the model, it was used in both the TEVA-SPOT software and the PipelineNet software (Pickus et al. 2005) to identify good locations for online monitoring. The analysis from both models was overlaid with staff expertise and practical knowledge to determine the final proposed monitoring locations.

The utility identified a list of 27 potential locations that included water utility facilities (pump stations, reservoirs, tanks, and pressure monitoring locations), other city facilities (fire stations, parking structures), and a limited number of private property sites. Each of these field locations was visited to determine its feasibility as a monitoring location. The sites were ranked based on the ownership of the site, availability of a connection to the distribution system, availability of power, communications, and sanitary sewer. Also, access and existing heating, ventilation, and air condition units (HVAC) were included in the assessment. In addition, the availability of historical water quality data was a factor, and one location was selected because of the abundance of such data.

The TEVA-SPOT software was used to select the best locations for security monitoring from the list of 27 potential locations. The analysis was performed with a variety of response delay times, 0, 4, 12, 24, and 48 hours. Two different contaminants were considered: a fast acting chemical contaminant and a slow acting biological contaminant. The selected locations were spatially distributed throughout the pressure zones ensuring good distribution system coverage.

The TEVA-SPOT analysis found that a small number of monitors provided significant benefits as measured by the percent reduction in public health impact. Four monitors were found to be sufficient — only small incremental benefits were estimated for adding more than four monitors. Given the size of the Ann Arbor distribution system (130,000 people over 49 square miles), this low number of monitors was a surprising outcome. It should be noted, however, that the percentage reduction in health impacts plateaus at about 75% to 80% protection. Therefore, with four monitors, over 20% of the population could still be impacted on average.

PipelineNet was used to evaluate potential sensor locations in order to protect sensitive facilities (schools and hospitals) and high population areas from contaminant attack. This was done by assuming that the contaminant introduction could only occur within a certain distance of critical facilities. Not surprisingly, this resulted in PipelineNet clustering sensor locations around the largest of these facilities. Although this might result in increased protection for these sensitive facilities, the remainder of the potential target population was not protected to the extent provided by the TEVA-SPOT methodology. Additionally, TEVA-SPOT provides the ability to quantitatively evaluate and compare potential sensor locations against different objectives (e.g., minimizing public health impacts, and constraints, e.g., number of sensors), and different threats (e.g., different contaminants and/or release locations).

The PipelineNet software was also used to determine areas with the highest water quality concern based on the criteria established by Ann Arbor staff, and these were matched against the 27 available monitoring locations. The results found that areas of impacted water quality clustered along the edges of the system and

along pressure boundaries, consistent with predictions of high water age areas and previous tracer studies.

The results of the TEVA-SPOT analyses, the PipelineNet results for water quality, and staff knowledge of the system were integrated. Four sites were selected for security monitoring and four locations were selected for water quality monitoring. One of the sites selected for security was the same as a site selected for water quality. This general lack of co-located sites was expected due to the different drivers for security monitoring (protect as much population as possible) versus water quality monitoring (find the areas of high water age usually associated with remote or isolated parts of the distribution system). However, this was considered an important finding by the authors, suggesting that security monitoring locations might not show significant dual benefit in a system where operational concerns are based on water quality effects such as nitrification. The project team was originally interested in the possibility of achieving efficiency in operations and cost savings if the security and quality locations over-lapped. However, this was not a requirement for the project.

A set of parameters were selected for potential monitoring (Hall et al. 2007) using a variety of information, including data from U.S. EPA's Test and Evaluation Facility in Cincinnati, Ohio, other research studies, utility surveys, and a workshop. Chlorine and TOC were the most highly recommended parameters to address water security concerns. TOC was ultimately not selected due to the instrument cost and complexity of operations. Ultraviolet (UV)-254 was selected instead, since it is often used as an alternative for TOC because both measure the organic content of water. Since the utility uses combined chlorine for final disinfection, the utility desired a total chlorine monitor that did not use reagents. Prior experience with analyzers requiring reagents had shown that they worked well at the treatment plant, but routine maintenance in the distribution system proved to be a challenge. Other parameters selected for testing included dissolved oxygen, ammonia, chloride and conductivity. Ammonia was selected as an indicator of water quality, because of nitrification due to the release of free ammonia from the decomposition of chloramines. Chloride was recommended as a general indicator ion of potential contamination. Dissolved oxygen was deemed useful for detection of nitrification, corrosion and contamination. Conductivity was selected as a general parameter for detection of contamination events.

In order to select specific instruments, a variety of criteria were developed to assess instrument performance and acceptability. Of these criteria, accuracy (agreement between lab testing and online instrument results), sensitivity (low level measurement ability), and variability (presumed normal fluctuation in water quality) proved the most important factors. Other factors, such as calibration ease, frequency, and maintenance are also important but the ability of the units to deliver useful data was deemed the most critical function. Based on pilot testing, chloride and ammonia were eliminated

as parameters for monitoring. This analysis concluded that total chlorine and dissolved oxygen were important parameters for measuring water quality degradation, but total chlorine, UV absorbance, conductivity, and dissolved oxygen are important for water security.

In Ann Arbor, the costs for monitor acquisition were estimated at \$25,000 per installation assuming that each location had four instruments: total chlorine, dissolved oxygen, conductivity, and UV-254 absorbance with the selected manufacturers. Installation costs, including infrastructure and communications, were estimated at an average of \$40,000 per location. However, this figure could vary widely depending on the extent of services available. Installation might include building a suitable structure, tapping a water main, installing electrical, sanitary and other support features. Operations and maintenance costs were estimated at \$7,000 per installation per year. This estimate

did not factor in the time needed to provide initial data handling and interpretation to develop response protocols. This consisted primarily of staff time to visit the site and perform routine maintenance and calibration activities. A 10-year life span was assumed for the equipment. Based on these estimated figures, the utility plan included an initial capital investment of approximately \$500,000 for eight sensor locations with an annual operating budget, including replacement costs of \$106,000. These costs do not include initial design and pilot testing work of approximately \$200,000. These figures are important when considering the cost/benefit ratio versus the number of monitors installed. Figures given are for approximate planning purposes only.

Finally, this study resulted in a specific design that was proposed to the City of Ann Arbor. The availability of funding will determine the schedule and implementation of the plan.

# Challenges for Real-World Applications

There are many outstanding research questions in the sensor placement field (Hart et al. in review). Current application of sensor network design optimization, then, can be challenging and sometimes requires imagination in addition to technical skill. In this chapter, several common questions that could arise when applying sensor network design optimization software are addressed. For example,

1. What is the best objective to use for sensor placement?
2. How many sensors are needed?
3. Should a CWS be designed to protect against high impact incidents only?
4. How can I make sensor placement algorithms work on typical desktop computers even for very large utility network models?

Discussions of these questions and suggestions for using the TEVA-SPOT software to help answer these questions are presented; however, in practice, there are no clear-cut answers because these questions involve policy concerns in addition to good science.

For demonstration purposes, these questions are addressed through analysis of a simple example network model: EPANET Example 3 network with 92 junctions, 2 reservoirs, 3 tanks, 117 pipes, and 2 pumps. This network is supplied by two surface water sources — a lake provides water for the first part of the day and a river for the remainder of the day. The average residence time for the network is 14 hours, and the maximum is 130 hours. Based on the average base demands, the total population served by this network is 78,800. The total length of pipe in the system is 215,712 feet.

In these analyses, sensors are assumed to be “perfect” in the sense that they have a zero detection limit and are always accurate and reliable (no false positives or false negatives and no failures). Utility response to detection of contamination incidents is also assumed to be perfect and instantaneous, meaning that following detection, a “Do Not Use” order is issued and made effective immediately, preventing all further consumption. These assumptions are referred to as “perfect sensors and perfect response.” These assumptions are applied in order to provide an upper bound on sensor network performance — the best that is theoretically possible. Even with perfect sensors and perfect response, a sensor network might not detect every event, detect events in a timely manner, prevent all exposure to contaminants, or prevent contamination of pipes. To achieve this perfect performance, in most networks, sensors would need to be placed at almost every junction. This is clearly not feasible in practice; thus, the importance of optimally selecting a small number of sensor locations using optimization software.

## Selecting the Best Objective

The performance objective is one of the most important parameters to select when optimally designing a CWS. For example, should a sensor network be based on minimizing the population exposed or minimizing the detection time? In practice, the authors have found that sensor network designs based on the various objectives can be very different from one another. Thus, it is important to understand the differences between the objectives when designing a CWS. The TEVA-SPOT software can be used to analyze and visualize the tradeoffs between different objective designs.

The following six performance objectives are available in TEVA-SPOT: population exposed (PE), extent of contamination (EC), volume consumed (VC), mass consumed (MC), number of failed detections (NFD), and time of detection (TD). [Note that an additional objective, population dosed (PD), has been added recently.] TEVA-SPOT works by simulating contamination incidents at a set of the nodes in the network specified by the user. For this chapter, contamination incidents were simulated at each of the 59 nodes with positive user demands. TEVA-SPOT calculates the performance objective for each incident, and then finds a single sensor that will best minimize the mean of the performance objective across all of the incidents. Each of the performance objectives is calculated using different equations (see Chapter 6); therefore the sensor that is selected is likely to be different.

**Figure 5-1** displays the sensor locations selected by each of the six objectives for the example network. Note that there is overlap only in two of the six objectives.

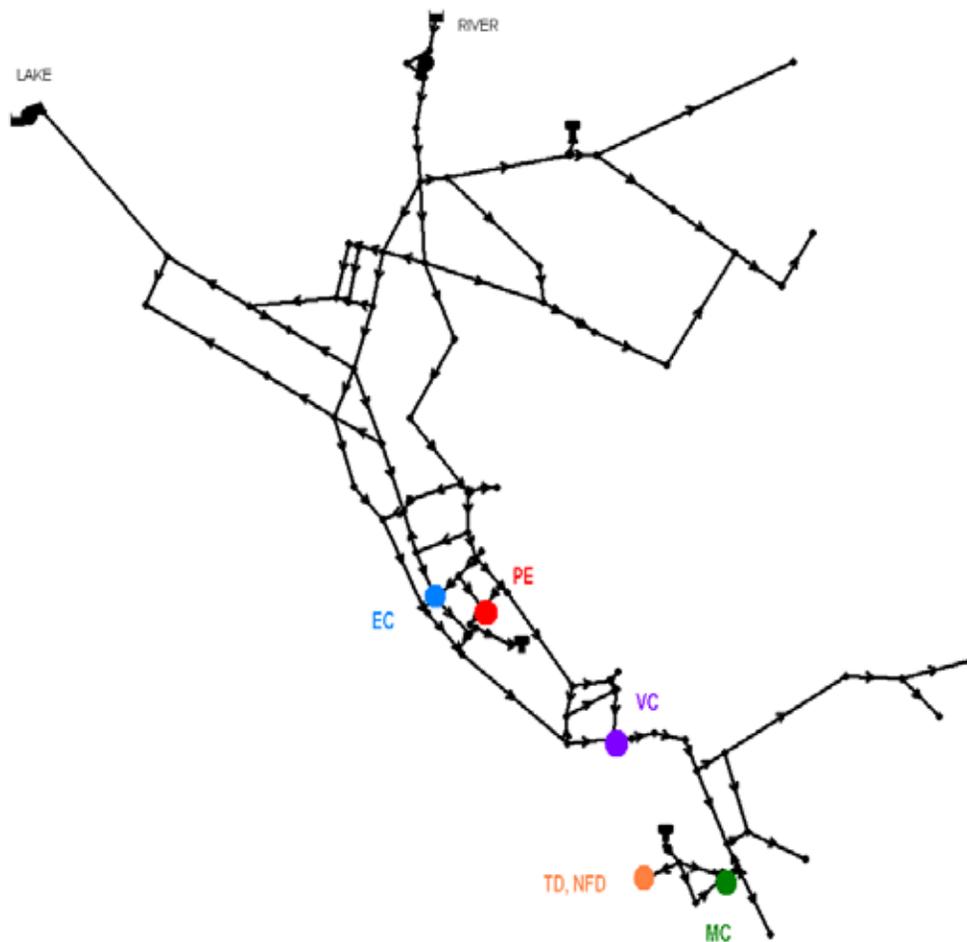
When the optimization method selects a sensor location based on PE, locations are likely to be selected in areas of the network which would detect incidents that impact the largest number of people. In this case, the sensor location that best minimized PE is Junction-271 (red circle in **Figure 5-1**), which is one node upstream of the node with the largest user demand. Thus, the sensor located at Junction-271 would detect all incidents that are along the flow path to the largest demand node. With this single sensor, over all of the 59 contamination incidents, the mean number of people exposed is reduced by 64% from approximately 11,000 people to 4,000 people.

Extent of contamination is another important performance objective for sensor network design, since knowing the length of pipes which become contaminated during an incident is essential in order to effectively decontaminate the system and return it to service. The flow in this network is from the two sources at the top to the bottom and to the right side of the network. The EC sensor location, Junction-189 (blue circle in

**Figure 5-1**), is in the middle of the network, thereby cutting the largest flow paths in half. A sensor at this node would detect incidents that have the potential to contaminate larger portions of the network. For all 59 contamination incidents, this sensor reduces the mean EC by 47% from approximately 46,800 feet to 25,000 feet.

The NFD metric aims to detect as many contamination incidents as possible. In this case, Junction-253 (orange circle in **Figure 5-1**) was selected as the best sensor location for NFD. This location detects the majority of the contamination incidents, since it is at the bottom edge of the network and the majority of the flow paths end here. Thus, at some time in the simulation, water originating from most injection locations will travel to this node. With this single sensor, 39 of the 59 incidents are detected (and 20 are not detected), resulting in a 66% reduction in the number of failed detects.

The TD objective selected the same junction. This might seem counterintuitive since this is near the end of the flow path for many incidents, and the time of detection would be quite large. This problem is due to the way that the TD objective is calculated in TEVA-SPOT. It greatly penalizes sensor network designs for not detecting an incident. If an incident is not detected, the performance metric assigns the detection time to the total length of the simulation. Therefore, a shorter simulation time can be used to generate more realistic designs using TD. In addition, there is an advanced option in TEVA-SPOT that does not penalize a design for the incidents that are not detected (see Berry et al. 2008b). This one sensor shown in **Figure 5-1** reduces average detection times from 192 hours to 69 hours for a 64% reduction in detection times.



**Figure 5-1.** Sensor designs for EPANET Example 3 network based on different performance measures.

If one performance objective is selected for sensor network design, what does that mean about the performance of the design in terms of the other metrics? For example, if a utility decides to design a sensor network based on minimizing public health exposures, what are the detection times for that sensor network? This question can be addressed by evaluating the performance of each design in terms of all the other performance objectives. The results of such an analysis are presented in **Table 5-1**. The first column shows the performance of the PE design in terms of each objective. Although the PE design performs well for both the PE and VC measures (achieving a reduction in impacts greater than 60% for both), it reduces the other impact measures by only 36 to 44%. If all the performance objectives are equally important, the regret score (defined first in Chapter 3) can be used to compare them. The lowest regret score means that the sensor design in that column performed better than the others over all performance objectives. In this case, either the MC or VC sensor designs perform best over all objectives.

Thus, when designing contamination warning systems, it is important to understand the different objectives, since sensor designs based on one objective might not perform well in terms of the other objectives.

### Number of Sensors

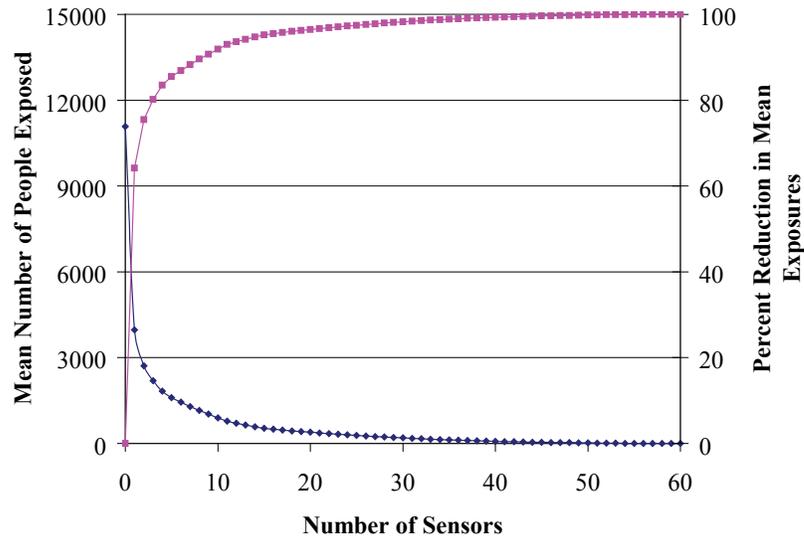
Another essential parameter to the sensor placement optimization problem is the number of sensors. As the capital costs of sensors can be in the tens of thousands of dollars, and the operation and maintenance costs can add up to 15 to 30% of capital costs each year, the number of sensors that

can be installed as part of a CWS is usually limited by utility budget constraints. Sensor placement tools like TEVA-SPOT can be used to develop tradeoff curves that demonstrate the relationship between the number of sensors (cost) and the benefit provided by the sensor network (calculated for a single objective) and can be used to help decision making. However, the question of how many sensors a utility needs in order to reliably reduce the risks of contamination incidents has not been answered definitively, and requires a difficult policy decision.

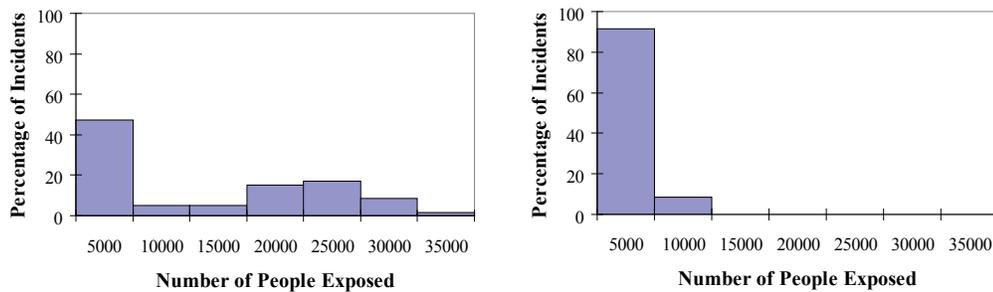
**Figure 5-2** shows such a tradeoff curve based on the PE objective for the EPANET Example 3 network. In the absence of a sensor network, an average of 11,000 people would be exposed to the contaminant out of a total population of approximately 114,200. A single sensor, optimally located, reduces exposures to 4,000 people on average (for a 64% reduction). Thus, the first sensor prevents an average of 7,000 exposures. This can also be stated by saying that the first sensor provides a marginal benefit of 7,000. Two sensors, optimally located, reduce exposures to 2,700 people. The second sensor, then, provides a marginal benefit of 1,300. The third sensor provides a marginal benefit of 500 people. After 10 sensors have been placed, the average number of exposures is reduced to 900 people, but it takes 59 sensors to reduce the average exposure to zero. Note that this would mean placing a sensor at every possible injection location (the 59 nodes with user demands). Each additional sensor yields less and less benefit, reflecting the diminishing marginal returns of sensor placement optimization algorithms.

**Table 5-1.** Percent reduction achievable for the sensor designs (in each row) based on each performance objective. The percentages in bold represent the best performance for the sensor design specified in that row. Higher percentages reflect better performance.

Performance Measure	PE	MC	VC	TD	NFD	EC
Mean PE	<b>64</b>	55	62	44	44	64
Mean MC	36	<b>56</b>	44	56	56	35
Mean VC	91	94	<b>95</b>	92	92	88
Mean TD	38	63	48	<b>64</b>	64	37
Mean NFD	39	64	49	66	<b>66</b>	37
Mean EC	44	22	43	15	15	<b>47</b>
Regret Score	0.43	0.27	0.27	0.38	0.38	0.45



**Figure 5-2.** Sensor network design tradeoff curve for EPANET Example 3 network based on the number of people exposed (PE) objective. The blue diamonds correspond to the left axis which shows the number of people exposed decreases with the number of sensors. The pink squares correspond to the right axis which shows the benefit of the sensor network (percent reduction in the number of people exposed) increases with the number of sensors.



**Figure 5-3.** Histograms of the percentage of incidents resulting in a given number of people exposed for the case with no sensors (left side) and the case with a five-sensor design (right side).

Policy makers could make reasonable yet conflicting conclusions from Figure 5-2. For instance, a policy maker could say that given budget concerns, placing five sensors provides substantial public health benefit to the system but that no additional costs can be justified (because the marginal benefits decrease dramatically after 5 sensors). Another policy maker could say that designing a system that would still expose more than 1,600 people (or 1.4% of the population) on average is not acceptable in any circumstance.

If the utility selected a sensor network with five sensors, the number of people exposed is reduced by 85%. **Figure 5-3** shows the distribution of PE over the 59 contamination incidents, first in the absence of sensors and then for the five-sensor design. The shape of the two distributions is quite different — the mean and maximum number of people

exposed is reduced significantly by five sensors. For this design, 81% of the incidents are detected with the five sensors (i.e., 19% are not), and an average of 18,000 pipe feet are contaminated. Is this level of risk reduction acceptable? Are there additional criteria that should be considered?

In order to answer this question, eight real-world water distribution networks that serve from 6,000 to 1.2 million people are examined. The goal is to look for trends among the networks that would help inform the question of how many sensors are needed. Trends are considered for multiple metrics of acceptable risk that impose specific limits on public health objectives (PE), coverage objectives (NFD), and economic objectives (EC). The characteristics of the additional networks can be found in Murray et al. (2008a).

**Table 5-2** lists the number of sensors needed to meet each of several public health metrics for the 8 networks based on results from TEVA-SPOT version 1.2. The results show the number of sensors needed to meet specific public health objectives. The first row shows the number of sensors needed to ensure that public health impacts will be less than 10,000 people on average. If a utility is only concerned with limiting exposures to less than 10,000 people, the results show that only 1 or 2 sensors might be necessary. However, for lower levels of risk, the number of sensors needed might be dependent upon population.

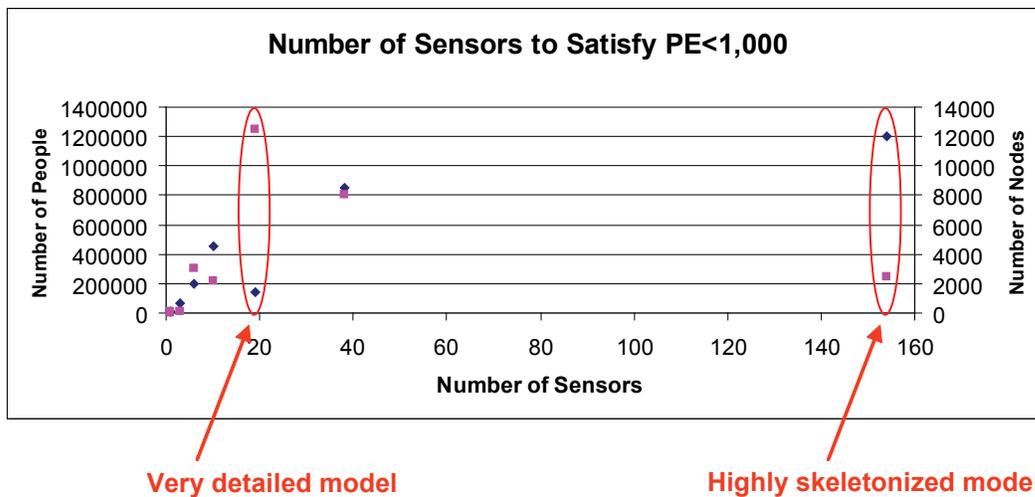
**Figure 5-4** plots the results for PE<1,000. The number of sensors needed to satisfy this metric is plotted against the population of each network (blue diamonds). There does appear to be a trend although there are two obvious outliers. Upon further examination, it appears that the level of detail in the model might be affecting these results. The right axis

is the number of nodes in the model. The two outlier cases represent an extremely detailed model (high number of nodes compared to population – Net 4) and an extremely skeletonized model (low number of nodes compared to population – Net 8).

Similarly, **Table 5-3** lists the number of sensors needed to meet several coverage objectives for the networks. The coverage metrics measure the percentage of contamination incidents detected (i.e., 1 – NFD) by the sensor network, from 40 to 90% of incidents. Net 4 gives anomalous results for these metrics, requiring significantly more sensors than the other networks. This is the extremely detailed network that produced anomalous results in **Figure 5-4**. As the number of nodes increases, so does the number of incidents, therefore, this metric should vary with the level of skeletonization.

**Table 5-2.** Number of sensors needed to achieve public health objective. \*This metric is beyond the resolution of the utility network model because of skeletonization.

Metric/Network	Net 1	Net 2	Net 3	Net 4	Net 5	Net 6	Net 7	Net 8
Population	6.2K	7.6K	114K	142K	200K	450K	840K	1,200K
Mean PE < 10,000	0	0	1	0	0	0	1	2
Mean PE < 1,000	1	1	3	19	6	10	38	154
Mean PE < 500	1	2	5	85	21	47	125	*
Mean PE < 100	5	5	11	*	*	*	*	*
Mean PE < 10	24	7	24	*	*	*	*	*



**Figure 5-4.** Number of sensors needed to satisfy PE<1,000 for each of the 8 networks versus population (blue diamonds) and number of network nodes (pink squares).

**Table 5-3.** Number of sensors needed to achieve coverage objective (number of incidents detected). +Note that the sensor placements were only calculated for up to 100 sensors, and these metrics required more than 100 sensors.

Metric/Network	Net 1	Net 2	Net 3	Net 4	Net 5	Net 6	Net 7	Net 8
Incidents	79	9	90	11,000	1,800	2,200	7,000	1,400
Mean ID > 40%	1	1	1	8	2	1	1	1
Mean ID > 50%	2	1	1	30	2	2	3	2
Mean ID > 60%	2	1	1	90	2	4	6	5
Mean ID > 70%	2	1	2	+	4	5	21	11
Mean ID > 80%	3	1	2	+	25	15	75	28
Mean ID > 90%	5	2	6	+	+	+	+	80

**Table 5-4.** Number of sensors needed to achieve economic objective. \*This metric is beyond the resolution of the utility network model given existing pipe lengths.

Metric/Network	Net 1	Net 2	Net 3	Net 4	Net 5	Net 6	Net 7	Net 8
Total pipe miles	123K	64K	216K	5.6M	4.1M	2.7M	9.4M	7.5M
EC < 100 miles	0	0	0	0	1	0	1	1
EC < 10 miles	0	0	0	7	12	10	25	27
EC < 1 mile	7	4	16	*	*	*	*	*

Finally, **Table 5-4** lists the number of sensors needed to meet several economic objectives for the networks. The economic metric is measured in terms of the length of pipe contaminated, from 1 to 100 miles.

Typically, water utilities use a combination of budget constraints and sensor network design performance curves in order to determine the appropriate number of sensor stations to install in a distribution network. An analysis of **Figure 5-2** might lead one to determine that 5 sensors is the most appropriate number for Net 3. However, **Tables 5-2, 5-3** and **5-4** show that with only 5 sensors, a contamination incident would be likely to result in more than 1 mile of contaminated pipe, 10% of incidents not detected, and 300 people exposed. Using acceptable risk criteria might persuade the utility to install additional sensors.

The number of sensors needed in a water distribution system is a question of acceptable risk. Acceptable risk must be defined by the water utility, and thus is highly dependent on the detection goals of the community. The risk reduction goals of communities can vary widely from striving to detect only catastrophic incidents, to detecting as many incidents as possible (including accidental cross connections). The utility might have broad goals, such as widespread coverage of the distribution system (for example, sensors in every pressure zone), detection of a large number of contaminants, and specific goals, such as preventing events that would be expected to impact more than 100 people. Using a multi-objective analysis might help to improve the performance of sensor designs across several objectives; however, there will always be a tradeoff in performance when balancing performance with costs. In order to design and implement an effective contamination warning system, utilities must explicitly consider the performance tradeoffs of the system they design.

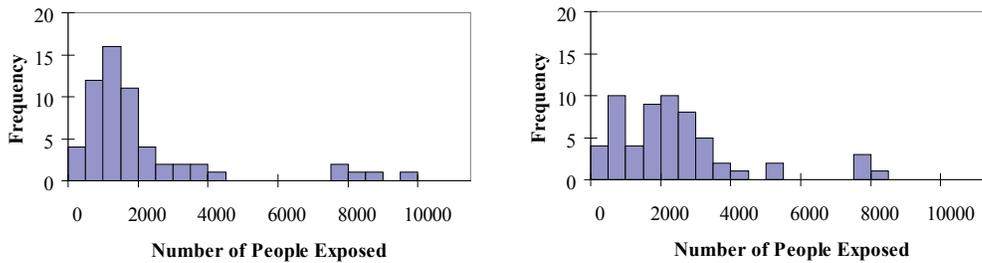
## Sensor Network Design Based on High-impact Incidents

Frequently, water utilities wonder why most sensor placement strategies focus on reducing mean consequences; they ask, “Why not design for high-impact contamination incidents only?” An optimal sensor network design based on minimizing the mean value of a performance measure can still allow many high-impact contamination incidents to occur. Further, most sensor placement optimization is done with the assumption that all incidents are equally likely (uniform event probabilities). This assumption is made because, typically, one does not have information about terrorist intentions; however, this results in an unintended de-emphasis of high-impact incidents.

It is possible to develop sensor networks based on high-impact contamination incidents. Rather than minimizing the mean, the optimization process can attempt to minimize the maximum value, or other robust statistic. A robust statistic is insensitive to small deviations from assumptions (Huber 2004). For example, the mean statistic is not robust to outliers because a single large value can significantly change the mean.

Although the final determination of the design statistic ultimately rests with policy-makers at a utility, the aforementioned factors strongly suggest that, at a minimum, there is a need to understand the differences between and implications of both mean-based and robust sensor designs.

To illustrate the relative tradeoffs that are possible between mean-case and robust sensor network designs, sensor placement designs that minimize PE with five sensors were examined for EPANET Example 3 network (for a full treatment of this topic on real-world networks, see Watson et al. 2009). Histograms showing the simulated number of



**Figure 5-5.** Histograms of the frequency of incidents resulting in a given number of people exposed for the case with a five-sensor network design designed by minimizing the mean-case (left side) or the max-case (right side).

people exposed during each contamination incidents (in this case, there are 59 incidents) if the mean-case or max-case sensor network designs is in place are shown in **Figure 5-5**. The distribution of impacts under the mean-case design is shown on the left side of **Figure 5-5**. With 5 sensors selected to minimize mean impacts, the mean was reduced from 11,000 to 1,600 people and the max-case reduced from 32,000 to 9,200 people. The distribution on the left side of **Figure 5-5** exhibits a key feature of sensor network designs that minimize the mean-case: the presence of non-trivial numbers of contamination incidents that yield impacts that are much greater than that of the mean. Even with these five sensors in place, there was one contamination incident that exposed more than 9,000 people, and an additional 15 contamination incidents that exposed between 2,000 and 9,000 people.

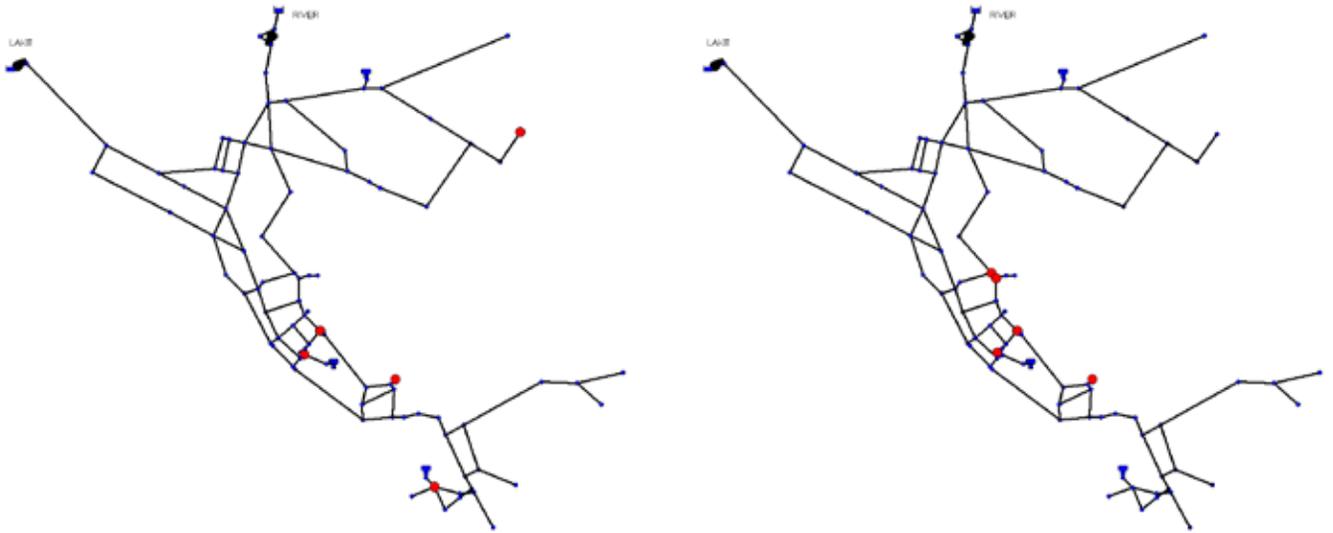
**The right side of Figure 5-5** shows the distribution of the number of people exposed for the case when a sensor network design was in place that minimized maximum impacts. With this design, there were not as many high impact incidents as there were with a sensor network design that minimized the mean number of people exposed. In particular, the highest-impact incident exposed 7,600 individuals, in contrast to 9,200 individuals under the optimal mean-case sensor design (**Table 5-5**). However, there were more small-to-moderate impact incidents. The max-case design yielded a mean impact of 2,300 people exposed, representing a 42% increase relative to the mean-case design which only impacted 1,600 people.

Thus, there is a tradeoff involved in switching from the mean to max case statistic for optimization — if the mean value is reduced, high impact incidents can still occur; if the max case value is reduced, the mean value will increase. In this case, the question for decision-makers in water security management is then: Is an 18% reduction in the max-case impact worth the 42% increase in the mean?

**Table 5-5.** Performance of mean-case and max-case for the five-sensor network designs in terms of the number of people exposed.

Performance Measure	Objective to Minimize	
	Mean	Max
Mean	1605	2287
Max	9223	7600

The qualitative characteristics of mean-based and max-case designs for this network can also be compared and contrasted. The locations of the respective sensor network designs are respectively shown in the left and right sides of Figure 5-6. To compare the two sensor network designs, characteristics such as the size and number of pipes connected to the sensor junctions, the demand at sensor junctions, the number of contamination events that are detected by each design, the average impact of these contamination events, and the time of detection are considered.



**Figure 5-6.** Diagrams of the five sensor network designs for the mean-case (left side) and max-case (right side).

In both designs, the majority of sensors were located at junctions along relatively large diameter pipes, which are often connected to more than 2 pipes; 9 of the 10 sensors were located at junctions with large demand. Specifically, all sensors were located on junctions connected to 8-inch or larger diameter pipes. Moreover, the majority of sensor junctions were connected to 24-inch pipes or greater (7 of the 10). One difference in the two designs, however, is that the max-case design put all 5 of the sensors on junctions connected to 24-inch or larger pipes, while only 2 of the sensors in the mean-case design were connected to 24-inch pipes or larger.

It appears from examination of **Figure 5-6** that sensors in the max-case design were somewhat closer together, possibly resulting in less spatial coverage of the distribution network. Forty-eight contamination events out of 59 were detected by the mean-case design; 31 events were detected by the max-case design. The average detection time of each design was different; 120 hours for the mean-case design and 270 hours for the max-case design. In addition, the average impact of the contamination events at the time of detection by the mean-case design was about 1,600 people, in contrast to 2,300 for the max-case design.

It should be noted that it is possible to gain significant reductions in the number and degree of high-consequence events at the expense of moderate increases in the mean impact of a contamination event. This can be accomplished through the use of side-constraints during the optimization process (see Chapter 7 for more information about side constraints). For example, if the mean is minimized, the max-case can be constrained to be less than some maximum value, so that the resulting sensor network design performs well both in minimizing mean and max-case consequences.

## Sensor Placement for Large Networks

Many optimization methods for sensor placement were developed and tested on small test networks; however, applying them to large real-world networks has proven to be a challenge. TEVA-SPOT has a number of effective strategies available to assist users in developing sensor network designs for large networks. When a sensor placement problem is so large that TEVA-SPOT runs out of memory using standard approaches, there are a number of strategies to produce sensor network designs using less memory. These strategies might result in designs that are not optimal but close to optimal. This section contains a qualitative discussion of the options, followed by a case study on runtimes for large networks. Refer to Chapter 7 for more quantitative descriptions of methods to reduce memory usage. The discussion in this section refers to optimizing the mean of a single objective function.

### Options for Reducing Memory

There are two main strategies for handling large networks:

1. Carefully choose the optimization solver.
2. Reduce the size of the problem by shortening the list of potential sensor locations, the list of contamination incidents simulated, or by using skeletonization or aggregation.

These methods are described in more detail in what follows.

One approach to managing large sensor placement problems is to carefully choose an optimization solver. There are three solvers available in TEVA-SPOT: an integer programming solver (IP), a heuristic solver (GRASP), and a Lagrangian solver (LAG). Chapter 7 describes these solvers and their tradeoffs in more detail. The heuristic solver is generally a good first choice, since it runs quickly and has been proven to produce good designs. If the heuristic fails on a real-world network, but only needs a small amount of additional

memory to run, then running the heuristic in sparse mode might be sufficient (see the TEVA-SPOT toolkit User Manual for more information: Berry et al. 2008b). If this does not work, try the Lagrangian solver. LAG uses the least amount of memory. The sensor designs it produces, however, are not as close to optimal as those produced by GRASP. Yet, as described in Chapter 7, LAG gives a lower bound on the optimal value of the objective. No sensor placement can be better than this lower bound. If the lower bound is close to the value of the objective for the sensor placement LAG finds, then this is a good solution.

If running the Lagrangian solver on the large network still requires too much memory, the next step is to create a smaller problem. Most methods to create smaller problems will remove information or restrict options. That means the solution, even if optimal for the reduced problem, will only approximately solve the full-sized problem. The first approach to creating a smaller problem is to change the input to TEVA-SPOT. Reducing the number of potential sensor locations reduces the memory requirements for all the solvers. This size reduction introduces no error if the locations that are removed in reality cannot host sensors. For example, if some nodes cannot host a sensor because they are on large mains or are otherwise inaccessible, these nodes should be marked infeasible. Utility owners may initially choose to consider all locations infeasible except for those explicitly evaluated and deemed feasible based on cost, access, or other considerations.

Another way to change the input is to reduce the number of contamination incidents in the design basis threat. The selected incidents should represent the original set as much as possible. For example, injection locations should cover all the geographic regions of the network. Currently TEVA-SPOT does not automate this process of reducing the number of contamination incidents. However, Chapter 7 describes one special case in which TEVA-SPOT can recognize an extremely similar pair of incidents and merge them into one.

Users can also change the input by coarsening the network through skeletonization, using, for example the techniques in Walsk et al. (2004) or a commercial skeletonization code. This merges pipes and nodes that are geographically close to create a smaller graph that approximates the hydraulic behavior of the original. However, it will introduce error by dropping some pipes of sufficiently small diameter.

TEVA-SPOT also provides an option, called aggregation, for automatically reducing the size of the problem. As described in Chapter 7, aggregation methods group potential sensor locations based on their performance for each incident. This effectively reduces the amount of memory needed to solve the sensor network design problem. When simulations are run with a coarse reporting step, aggregation can save some space without introducing error. The IP solver, for example, will do this automatically. However, if that is not sufficient then users can direct TEVA-SPOT to group nodes with differing, but approximately similar quality. The loss of information means the solver can only approximately solve the full problem. Aggregation is only available for the IP and LAG solvers. By

selecting *ratio aggregation* with ratio  $\rho$ , the resulting sensor network design could have an objective as much as a factor of  $\rho$  higher than the optimal. A user will need to use trial and error to determine the smallest value of  $\rho$  that produces a problem that can be solved.

Finding and evaluating methods for effectively solving large problems is an area of ongoing research. There are planned improvements for TEVA-SPOT in the near future. For example, TEVA-SPOT will have a built-in skeletonizer. Future aggregation methods may involve several steps, using solvers such as GRASP to do sensor placements on compressed and/or restricted instances. Future versions of TEVA-SPOT will allow the users more freedom in specifying how aggregated values are computed, allowing more options for approximately solving large instances. Users with difficult large instances should consult the TEVA-SPOT release notes and documentation to learn about new options as they become available.

### Case Study on Runtimes for Large Networks

The execution of TEVA-SPOT on large utility network models can be time consuming. **Figure 1-2** showed the data flow for the TEVA-SPOT software. Each of the major computational steps takes time: simulating incidents, assessing consequences, and optimizing sensor placement. Computational runtimes for all three steps are determined by: (1) network topology and hydraulics (e.g., the number of nodes or junctions and the flow paths); (2) EPANET simulation options (e.g., simulation length, water quality and hydraulic time steps, and reporting time step); (3) the design basis threat (e.g., the number of contaminants, the number of injection locations and times).

Here a case study is presented for a large utility network model using the TEVA-SPOT User Interface which contains a distributed processing capability. The software distributes EPANET simulations, consequence assessment calculations, and sensor placement optimization when sufficient memory and processors are available. A minimum of two gigabytes (GB) of random access memory (RAM) are required per processor.

For the case study, the runtimes are reported for a single processor computer, a dual processor (dual core), and a dual, quad core processor. The large utility network model consists of approximately 50,000 nodes, which includes approximately 10,000 non-zero demand nodes, about 10 reservoirs, and numerous valves, pumps, and tanks. A single water quality simulation in EPANET 2.00.12 takes about a minute.

Sensor placement is a challenge for this network because it requires large amounts of memory. The problem size must be reduced in order to use either the GRASP or LAG algorithms. The first strategy used in this case study was to reduce the number of feasible sensor locations from 50,000 to about 1,000 locations. The problem size was also reduced by skeletonizing the network. MWH Soft's Skeletonizer was used to preferentially remove pipes and connected nodes by the "Trim," "Reduce," and "Merge," skeletonizer routines.

Using a single processor, **Table 5-6** shows runtimes are reported for each step of the computation: EPANET simulations, consequence assessment, and sensor placement optimization. The total simulation time is calculated by multiplying the number of injections (10,000) by the sum of the second and third columns (EPANET simulation and consequence assessment runtimes) plus the fourth column (GRASP runtime) plus the sixth column (sensor placement summary runtime). Total runtimes are on the order of tens of days on a single processor. It should be noted that in this case, only a single sensor placement analysis was completed; in practice, several are usually analyzed which would further elongate runtimes.

The performance objective used here was PD — the number of people receiving a dose above a fixed threshold. Results are presented in **Table 5-6** where runtimes are reported in seconds. This table shows how the runtimes

are reduced when the problem size is reduced by: (1) reducing the number of potential sensor locations, (2) reducing the EPANET simulation duration, or (3) reducing the number of nodes and pipes in the network through skeletonization. Increasing the number of feasible sensor locations increased the corresponding runtime for both the GRASP and LAG algorithms.

In **Table 5-7**, total runtimes are reported for 3 single workstation configurations: (1) single processor, (2) a dual processor with 4 or more GB RAM, and (3) a dual, quad core processor with 8 or more GB RAM. Reported runtimes are for the mean statistic using the PD objective and the GRASP algorithm and 100 sensor network designs were generated (in which the number of sensors, the response time, and detection time were varied). Finally, the runtimes reported should be considered as only likely estimates of run times; newer and faster processors will have shorter runtimes.

**Table 5-6.** Runtimes for each component of the sensor network design process using the TEVA-SPOT User Interface. The simulation options include the simulation duration (168 or 240 hours), the number of potential sensor locations (10,000 or 1,000), and the original model or a skeletonized model. The sensor placement summary step is unique to the User Interface and is the time required to report results to files and output tables.

Simulation Duration/Injection Nodes/Number of Feasible Sensor Locations/Skeletonization	Single EPANET Simulation (seconds)	Cons. Assess. Simulation (seconds)	Sensor Placement GRASP vs. LAG (seconds)		Sensor Placement Summary (seconds)	Total Simulation Time (seconds)
240 Hours; 50,000 nodes; 1,000 locations; original model	479	93	1,600	4,800	210	5,721,810 (66 days)
240 Hours; 10,000 nodes; 10,000 locations; original model	479	93	1,000	1,600	152	5,721,152 (66 days)
168 Hours; 10,000 nodes; 10,000 locations; original model	338	93	831	1,400	116	4,310,947 (50 days)
240 Hours; 10,000 nodes; 10,000 locations; 8-inch skeletonized	239	50	610	1,224	80	2,890,690 (33 days)
240 Hours; 10,000 locations; 12-inch skeletonized	203	44	463	1,156	65	2,470,528 (29 days)
240 Hours; 10,000 nodes; 10,000 locations; 16-inch skeletonized	185	59	414	1,134	73	2,440,487v (28 days)

**Table 5-7.** Approximate total runtimes (in days) for three different computing platforms: a single processor, a dual processor, and a dual quad core processor.

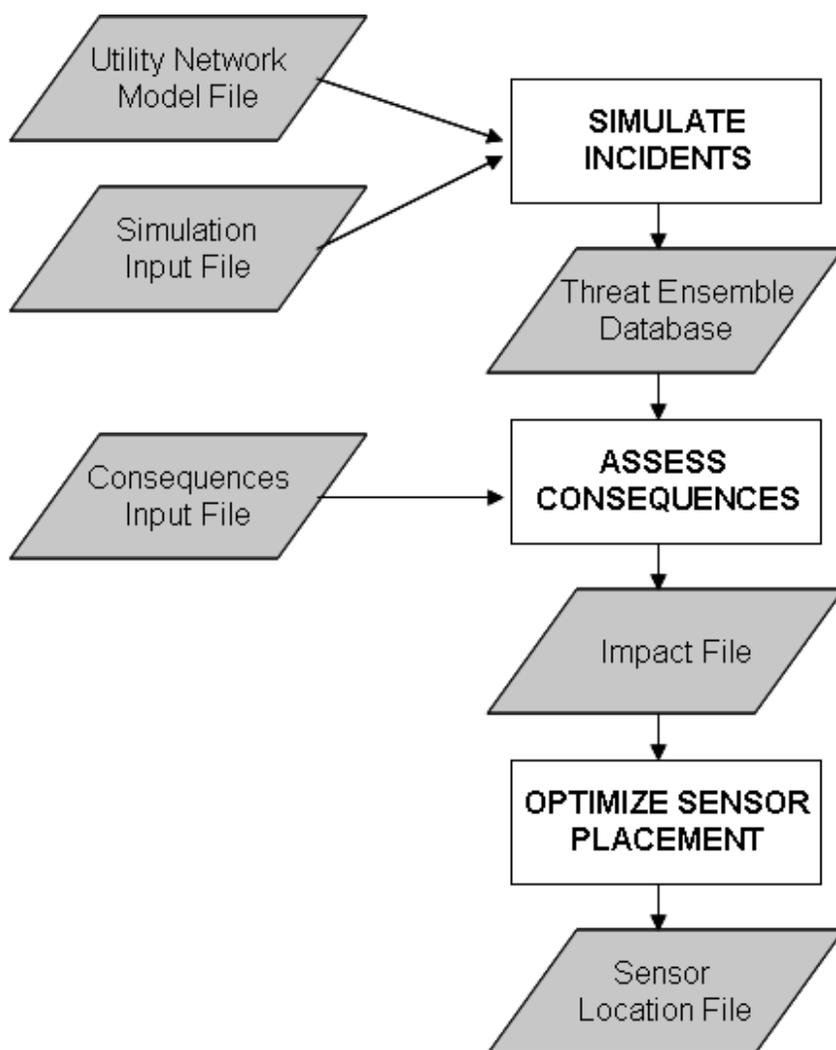
Simulation Duration/Injection Nodes/Number of Feasible Sensor Locations/Skeletonization	Total Computational Time		
	Single Processor	Dual Processor (4 GB +)	Dual, Quad Processor (8 GB +)
240 Hours; All Nodes; 1,000 locations; original model	321	134	54
240 Hours; NZD Nodes; 10,000 locations; original model	62	26	10
168 Hours; NZD Nodes; 10,000 locations; original model	47	18	8
240 Hours; NZD Nodes; 10,000 locations; 8-inch Skeletonized model	27	11	5
240 Hours; NZD Nodes; 10,000 locations; 12-inch Skeletonized model	23	10	4
240 Hours; NZD Nodes; 10,000 locations; 16-inch Skeletonized model	23	9	4

# Impact Assessment Methodology

This chapter presents the technical details of the methodology underlying the simulation and consequence assessment modules of TEVA-SPOT. **Figure 6-1** shows the data flow in the TEVA-SPOT software.

The simulation module of the TEVA-SPOT software simulates a set of contamination incidents in a specific water utility distribution system network. The user must provide a utility network model (e.g., an EPANET input file), and

an input data that defines the set of contamination incidents. Incidents are specified by a single injection location in the distribution system, the assumed volume and concentration of the contaminant, and the start and stop time of contaminant introduction. The EPANET software is used to simulate the transport of each contaminant through the water distribution network. Concentration profiles from each contamination incident are stored in an output database for further analysis.



**Figure 6-1.** Data flow diagram for the TEVA-SPOT software.

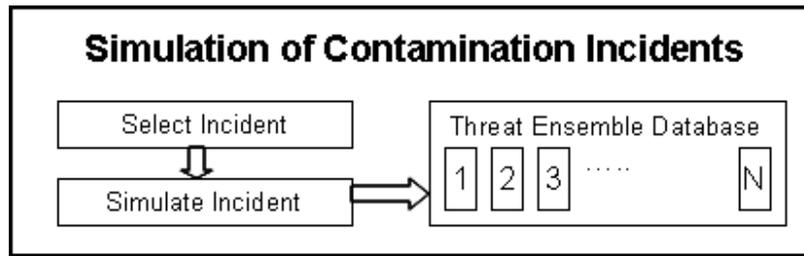


Figure 6-2. Data flow diagram for the simulation module of TEVA-SPOT.

The consequence assessment module of TEVA-SPOT reads in the database output from the simulation module and calculates the potential impacts of each contamination incident. This module calculates impacts in terms of the number of people becoming ill from exposure to a contaminant, the volume or mass of contaminant removed from the network, or the length of contaminated pipe in the distribution system. The results of this analysis are stored in an output file for further analysis. The rest of this chapter describes these methodologies in more detail and refers the reader to additional background material when needed.

### The Simulation Module of TEVA-SPOT

Given a utility network model, the simulation module simulates a set of contamination incidents. The set of incidents make up the “design basis threat” for the sensor network design — the set of contamination incidents that the water utility would like to be able to detect with a sensor network. Given that there are a wide variety of potential contamination threats to water distribution systems, and it is difficult to predict the exact incident adversaries might try to enact, TEVA-SPOT supports the simulation of a large number of threat incidents (as shown in **Figure 6-2**).

The utility must provide a network model as input (see Chapter 2 for more discussion on the model requirements). Incidents are defined by the location at which a contaminant is introduced, the start and stop time for contaminant introduction, and the mass injection rate. When using the TEVA-SPOT User Interface, this data is input in a window (U.S. EPA 2009); if using the TEVA-SPOT toolkit, this information is specified in an input file (Berry et al. 2008b).

#### Selecting Incidents

**Location.** Contaminant injections can be simulated for a single location, a set of locations, or at all possible locations (all the nodes defined in a utility network model).

**Start and Stop Time.** The start and stop time, or the duration (D) of the contamination injection must be specified. In practice, the authors generally have used durations between 1 and 24 hours. For more information about the influence of the timing of the contamination incidents on the consequence assessment, see Murray et al. (2006b) and Davis and Janke (2008).

**Mass Injection Rate.** The mass injection rate is the rate at which mass enters the distribution system. One can choose an arbitrary value (e.g., 1,000 mg/min), or one can calculate this value based on assumptions about a specific contaminant.

Contaminants of interest for water security could include chemicals (household, toxic industrial, and chemical warfare agents), biotoxins (such as botulinum toxin or ricin), biological pathogens (bacteria, viruses, or protozoa), and radiological (e.g., Cs-137).

The mass injection rate can be calculated based on a contaminant stock concentration (C) and volume (V) and the duration (D) over which the contaminant is introduced. The concentration and volume can be estimated based on the availability and technical feasibility of acquiring or producing the contaminants. For example, some toxic industrial contaminants can be purchased at large quantities at a known concentration. Some bacterial cultures are known to require a relatively low level of skill and equipment to produce at a particular concentration and volume.

A target mass release rate (MR) can be calculated by:

$$MR = \frac{VC}{D} \quad (6-1)$$

#### Simulating Incidents with EPANET

TEVA-SPOT simulates contamination incidents using EPANET (Rossman 2000). EPANET utilizes the system specific data related to utility operations and customer demands provided in the utility network model to simulate the hydraulics of pipe flow and water quality throughout the distribution network.

Contaminant injections are simulated as mass sources, thereby adding mass to the system without directly changing the hydraulics at the point of introduction. Typically, all contaminants are treated as conservative tracers. This results in both overestimation and underestimation of contaminant concentrations at specific locations, because fate and transport processes such as hydrolysis, oxidation, adsorption, and attachment to biofilm are not considered. It is possible to assume constant first order decay for contaminants, although it is difficult to determine appropriate decay constants that lump together all of these processes. (Later versions of TEVA-SPOT will run with EPANET-MSX (Shang et al. 2007), which allows for more complex fate and transport modeling.)

Each incident is run separately and the contaminant concentration time series (averaged over each reporting time interval) for each node in the network model are stored in an output database. The TEVA-SPOT User Interface supports distributed processing of the EPANET runs. For a dual-core

machine with sufficient memory, TEVA-SPOT can run two simulations simultaneously, thereby reducing the run time by a factor of two. Quad-core or dual quad-core workstations would offer even greater computational efficiency. The TEVA-SPOT software can also be run on a distributed server-based computing system (U.S. EPA 2009).

### Output Database

Simulation results are stored in a binary database for later analysis by the consequence assessment module (see **Figure 6-2**). This is a structured database that efficiently stores a large volume of numerical data. The database includes header information, hydraulic information, and the concentration matrix. The concentration matrix combines the time series of contaminant concentrations at all nodes in the network model. For a more detailed description of the output database, see the TEVA-SPOT User's Manual (Berry et al. 2008b).

### Consequence Assessment Module

The Consequence Assessment module of TEVA-SPOT calculates the potential consequences of each incident simulated. In particular, the module calculates the potential public health impacts, the extent of contamination in the pipe network, and the mass or volume of contaminant that has been removed from the pipe network. The results of the consequence assessment calculations are then stored in impact files. The impact files are utilized by the sensor placement optimization module.

### Public Health Impacts

Public health impacts can be estimated by combining the contamination concentration time series with exposure models. Contaminant-specific data is needed to accurately estimate the health endpoints. For many threat agents, reliable data are lacking, and the ensuing uncertainty in the results must be understood.

**Population models.** In order to calculate exposure to contaminants, an estimate of the population consuming water at each node is required. In TEVA-SPOT, the default is to calculate the population at each network node based on the total amount of water consumed at that node over a 24-hour period:

$$pop(x_i) = \frac{\sum_{j=1}^{24} q(x_i, t_j)}{R_{pc}} \quad (6-2)$$

where  $q$  is the demand (or total water usage) and  $R_{pc}$  is the per capita consumption rate per day. A USGS report provides usage rates by state and gives a nationwide average of 179 gallons per capita per day (USGS 2004). For simplicity, 200 gallons per capita per day is often used for TEVA-SPOT calculations.

If detailed population information is available for each node in a network model, users can input a population file (see U.S. EPA 2009). The file-based approach allows users to input accurate numbers from utility billing records or from census data.

Population is assumed to be constant over time. Population mobility is ignored, and so effects related to commuting to work and attending school or daycare are not evaluated.

**Routes of exposure.** Exposure to contaminants in domestic drinking water supplies is possible through multiple routes depending on water usage and the specific characteristics of the contaminant. Municipal water is used for drinking, showering, washing clothes, brushing teeth, cooking, bathing, cleaning, watering the lawn, and more. Through such activities, there is the potential for exposure to contaminants in drinking water through three primary routes: inhalation, dermal contact with the skin or eyes, and ingestion. An individual could be exposed to some contaminants through all three exposure routes.

Inhalation exposure might occur if a contaminant is volatilized or aerosolized. Pathogens, biotoxins, chemicals, and other contaminants could be inhaled in the form of finely dispersed mists, aerosols, or dusts during showering, bathing, cooking, or lawn work. Household appliances such as dishwashers and washing machines may also contribute significantly to the inhalation exposure pathway for volatile organic compounds (VOCs) (Howard-Reed et al. 1999; Jacobs et al. 2000). Highly water-soluble gases and vapors and larger mist or dust particles (greater than 10 microns in diameter) generally are deposited in the upper airways. Less soluble gases and vapors and smaller particles can be inhaled more deeply into the respiratory tract. Inhaled substances can be absorbed into systemic circulation, causing toxicity to various organ systems (ATSDR 2001).

Skin and eye contact can occur when handling contaminated water or by using contaminated water for laundry, recreational activities, bathing, or washing. Corrosive agents can cause direct damage to tissues by various mechanisms including low or high pH, chemical reaction with surface tissue, or removal of normal skin fats or moisture. Chemicals also can be absorbed systemically through the skin. This is more likely to occur when the normal skin barrier is compromised through injury or when the chemical is highly fat-soluble such as organophosphate and organochlorine pesticides (ATSDR 2001).

Ingestion is the most likely route of human exposure to contaminants from the drinking water supply. Ingestion of a corrosive agent can cause severe burns to the mouth, throat, esophagus, and stomach. Chemicals also can be aspirated into the lungs (e.g., liquid hydrocarbons), causing a direct chemical pneumonia (ATSDR 2001). A study from England reported that pathogens ingested from contaminated water are a major contributor to the estimated 1 in 5 people in the general population that develop an infectious intestinal disease each year (Wheeler et al. 1999). Many of the biological agents can also be dangerous ingestion risks.

Some studies indicate that even small doses of contaminants could result in higher combined inhalation, oral, and dermal exposures from daily water use (Shehata 1985; Weisel et al. 1996). Several studies have concluded that skin absorption or inhalation of contaminants in drinking water has been underestimated and that ingestion might not constitute the sole or even primary route of exposure (Andelman 1985a, 1985b; Brown et al. 1984). Another study estimated that the uptake of VOCs from household inhalation may be from 1-6 times the uptake of ingestion exposure. In addition, the uptake of VOCs from dermal exposure during baths and showers could be from 0.6-1 times the uptake of ingestion exposure (McKone 1989).

The design basis threat for sensor placement is often based on high impact contamination incidents that would involve contaminants that have rapid and/or acute health impacts. It is assumed that the volume and concentration of contaminants introduced into the drinking water system would be selected to maximize the health impacts to the population; therefore, the quantities would be sufficient to cause harm from ingestion alone. Long-term exposure to low levels of contaminants through multiple exposure routes would certainly increase the overall public health impacts, but currently the consequence assessment module only estimates exposure to contaminants through ingestion. Future versions of TEVA-SPOT could include the capability to model exposure from inhalation and dermal routes.

**Modeling exposure.** The Consequence Assessment Module estimates exposure to contaminants at each node,  $x_i$ ,  $i=1..N$ , where  $N$  is the total number of nodes, in a drinking water distribution system. At each node, there are many people being served water, the total number is given by  $pop(x_i)$ . The cumulative dose of a contaminant ingested by the population at  $x_i$  at time  $t$  is calculated according to:

$$d(x_i, t) = \sum_{j=1}^T C(x_i, t_j) P_W(x_i, t_j) V_W, \quad (6-3)$$

where  $d$  is measured in number of organism or number of milligrams,  $C$  is the contaminant concentration in water at node  $x_i$  at time  $t$  as predicted by EPANET,  $P_W$  is the probability of a person consuming water at time  $t$ ,  $V_W$  is the volumetric rate of water consumption, and  $T$  is the time period of interest.

The assumption that the dose is accumulating over the entire simulation period can result in an overestimation of the health impacts. Some toxic chemicals, such as cyanide, are effectively removed from the body quite rapidly; thus, a lethal or harmful dose would need to be accumulated over a very short period of time — before the body had time to render the substance harmless.

The volumetric rate of water consumption,  $V_W$ , is commonly assumed to be 2 Liters/day for risk assessment purposes (U.S. EPA 1997); however, studies show that the average quantity of tap water ingested in the U. S. is less than 2 Liters/day (Jacobs et al. 2000). A survey conducted in 2002 by the

International Bottled Water Association found that the mean daily water consumption by Americans is 1.25 Liters per day. This figure takes into account variations between age group, sex, and regions. The survey found that adults over 24 years old drank more water than those 18–24, women drank more water than men, and those residing in the western part of the country drank more than those in the northeast, south, and Midwest (IBWA 2002). The Consequence Assessment Module allows users to select a fixed value for  $V_W$ , or to select a probabilistic model for volumetric rate that selects from a distribution (Jacobs et al. 2000).

The probability that an individual at node  $x_i$  consumes water at time  $t$ ,  $P_W$ , can be estimated by one of three “timing” models (see **Table 6-1**). The simplest model, labeled D24 in **Table 6-1**, assumes that the timing of water consumption is proportional to the timing of network demands. The probability of consuming water at time  $t$  is assumed to be proportional to the ratio of the demand  $q$  at time  $t$  to the average demand over the time period  $T$ ,

$$P_W(x_i, t) = \frac{q(x_i, t)}{\frac{1}{T} \sum_{j=1}^T q(x_i, t_j)} \quad (6-4)$$

This demand-based timing model is probably not accurate for a single person, but instead reflects the average usage patterns of all the people being served at a particular node. This model was used in Janke et al. (2006) and Murray et al. (2006b).

Network demands quantify the total amount of water used over time. However, demand accounts for both ingestion of water as well as water usage for washing dishes, laundry, showering, and watering the lawn. It is estimated that less than 1% of water demand is actually consumed, and the timing of consumption might not be correlated with the timing of water usage overall (Jacobs et al. 2000). Therefore, timing models based on other information than demands could be more accurate.

Little information has been collected on the times of day at people ingest tap water. Studies in the U.S. and England have shown that 68 to 78% of the daily intake of water is consumed when people eat (de Castro 1988; Engell 1988; Phillips et al. 1984). The quantity of water ingested is determined primarily by how much food is ingested, and this does not vary with age among 20 to 80 year-olds (de Castro 1988, 1992). Models for the timing of eating, therefore, might be useful for predicting the timing of water consumption. A simple ingestion model is based on three conventional meals per day (Ma et al. 2005). In 2003 and 2004, the American Time Use Survey (ATUS), sponsored by the Bureau of Labor Statistics (BLS) and conducted by the U.S. Census Bureau, reported on the starting times for eating (BLS et al. 2005).

Another timing model, labeled F5 in **Table 6-1**, assumes that tap water is ingested at five fixed times a day corresponding to the typical starting times for the three major meals on weekdays (7:00, 12:00, and 18:00 hours) and times halfway

between these meals (9:30 and 15:00 hours). A third model, P5 in **Table 6-1** also assumes that tap water is ingested five times per day at major meals and halfway between them, but uses a probabilistic approach to determine the actual times. Both of these models are based on the ATUS data. For more information about these models, see Davis and Janke (2008; 2009).

**Table 6-1.** TEVA-SPOT consumption timing models.

Name	Description
D24	Demand based, every time step
F5	Ingestion based, fixed times (5 events)
P5	Ingestion based, probabilistic (5 events)

**Dose-response models.** Equation 6-3 predicts the dose received by each individual at a given node. Dose-response curves can be used to predict the percentage of people who might experience a particular health outcome after receiving a specific dose. For chemicals, often the outcome of utmost concern is fatalities. For biologicals, the outcome can be infection or fatalities. Dose-response curves can be considered the probability of a representative individual dying as a function of exposure.

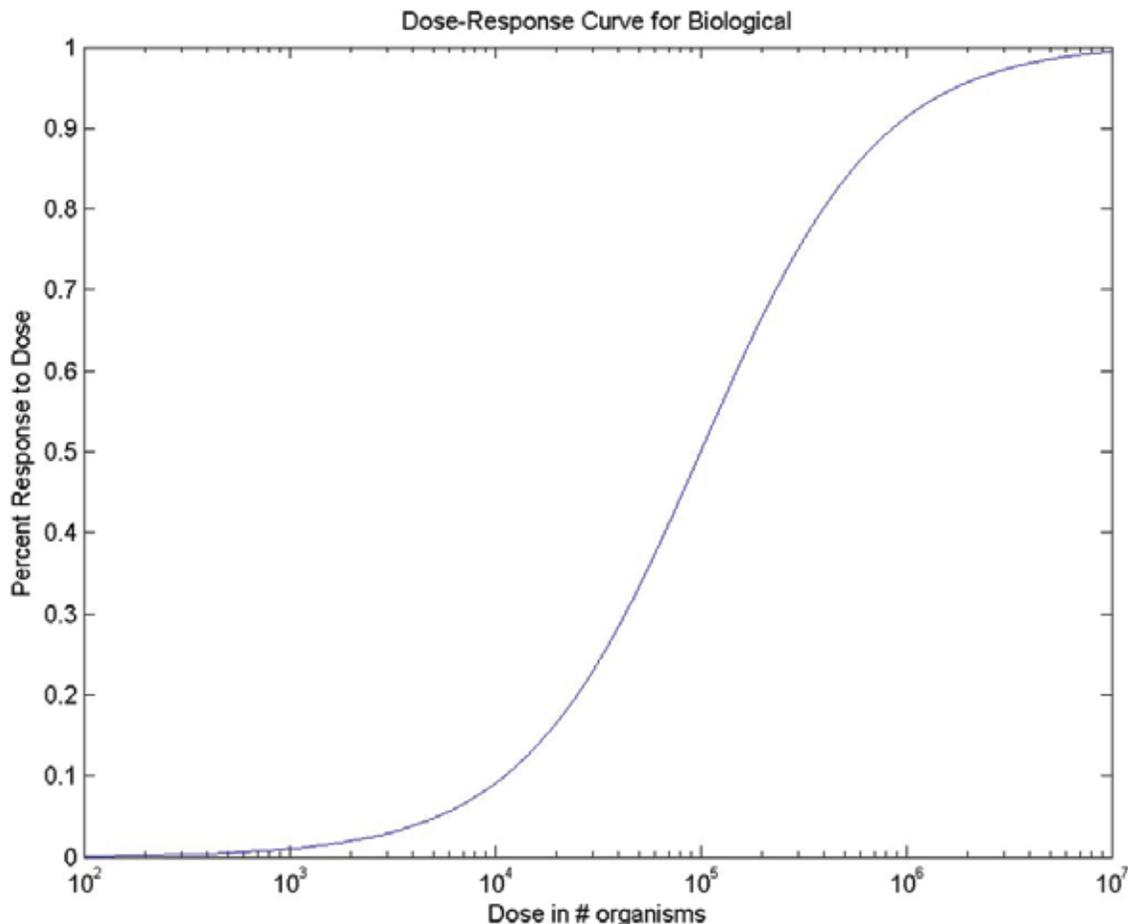
An example dose response curve is given in **Figure 6-3**; note that the  $ID_{50}$  shown is 100,000 organisms.

The consequence assessment module includes two dose-response functions. The first is the log-probit model, which is a toxicity model frequently used for a wide range of contaminants. Results of toxicity studies of both chemical and infectious agents often fit the shape of this model (Covello et al. 1993); however, this model does not work well for biological contaminants with health outcomes that occur at very low doses (Haas et al. 1999). The model is based on the assumption that the tolerance (dose at which response is first observed, or threshold) to exposure to a harmful agent of members of a population follows a log-normal distribution. The model is also referred to as the log-normal dose-response model. The log-probit model predicts the probability of fatality at a given dose  $d$  by:

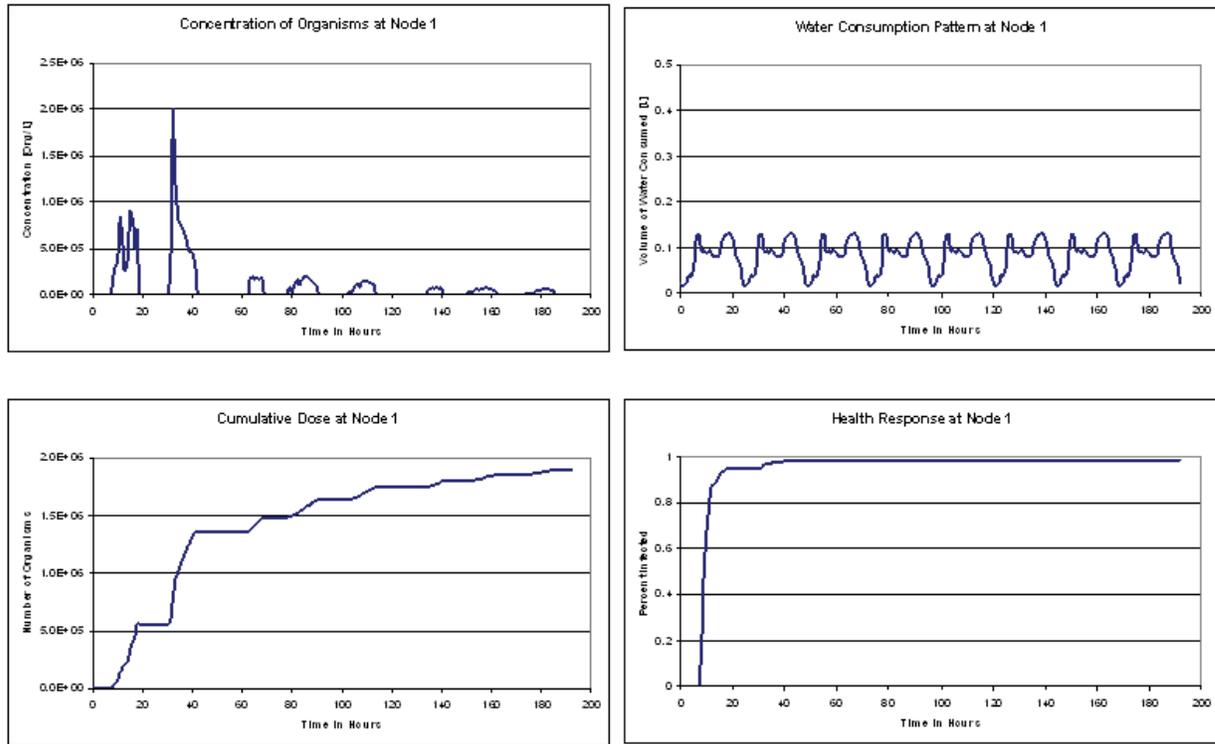
$$r(d) = \Phi(\beta_1 + \beta_2 \ln d) \quad (6-5)$$

where  $\Phi$  is the cumulative distribution function of a standard normal random variable,  $\beta_2$  is related to the slope of the curve,  $\beta_1$  is the product of  $\beta_2$  and the log of the  $LD_{50}$  (the dose at which 50% of the population would die).

The log-probit model produces a symmetrical sigmoidal curve when log dose is plotted against cumulative response, with the  $LD_{50}$  lying at the inflection point of the curve. When this curve is put through a probit transformation, which



**Figure 6-3.** Example dose-response curve for a biological agent.



**Figure 6-4.** At one node, the concentration of contaminant, the consumption patterns, the cumulative dose, and the percent health response over time.

converts cumulative response to probit units, or number of standard deviations, a straight line is formed, the slope of which is represented by beta. Probit plots are used for comparing the relative sensitivity (slopes) of populations to different toxic agents.

The second dose response curve is a more generic logistic function with a sigmoidal shape given by

$$r(d) = \frac{1 - e^{-d/\tau}}{1 + (e^{LD50/\tau} - 2)e^{-d/\tau}} \quad (6-6)$$

where  $\tau$  is a parameter that controls the slope of the response curve. This parameter can be used to fit this model to available data. Other dose response models might be more appropriate for specific contaminants; however, at this time equations (6-5) and (6-6) are the only models contained in the consequence assessment module.

**Figure 6-4** shows plots of several of the quantities used in Equations 6-2, 6-3, and 6-4 as predicted in one particular example incident. The figure shows how one location in the system (downstream of the introduction location) would experience a specific contamination event. The figure shows four plots: the concentration of contaminant ( $C$ ) that passes by the consumers at one node, the water consumption patterns of consumers ( $P_w$ ), the cumulative dose received by consumers ( $d$ ), and the response function ( $r$ ) (cumulative

percent of population experiencing a health response) over time. Note that the concentration profile is very complicated since the spatial location is under the influence of a nearby tank. The contaminant is drawn inside the tank as the tank fills and is transported out as the tank drains.

**Dynamic disease progression models.** Equations (6-2)–(6-6) are used to predict the number of people at each node who become infected or ill. For water security applications, knowledge of the timeline of events is critical in order to provide rapid response to reduce the impacts. Understanding the timeline of public health impacts can allow utilities and public health departments to plan for effective interventions that reduce further exposures and/or treat the people who have been exposed. To this end, TEVA-SPOT combines the dose-response model with dynamic disease progression models in order to predict how the illnesses progress over time.

Given the percentage of people at each node who become ill after being exposed to the contaminant, disease transmission models predict how the disease progresses over time. Disease models are used to predict the number of people at each node susceptible ( $S$ ) to illness from the contaminant, exposed to a lethal or infectious dose ( $I$ ), experiencing symptoms of disease ( $D$ ), and either recovering ( $R$ ) or being fatally impacted ( $F$ ). These quantities are predicted at each node over time according to the following differential equations:

$$\begin{aligned}
\frac{dS}{dt} &= \gamma R - \lambda S \\
\frac{dI}{dt} &= \lambda S - \sigma I(t) \\
\frac{dD}{dt} &= \sigma I - (\alpha + \nu) D \\
\frac{dR}{dt} &= \nu D - \gamma R \\
\frac{dF}{dt} &= \alpha D
\end{aligned}
\tag{6-7}$$

where  $\nu$  is the per capita recovery rate ( $1/\nu$  is the mean duration of illness),  $\sigma$  is the inverse of the mean latency period,  $\alpha$  is the per capita untreated death rate, and  $\gamma$  is the per capita rate of loss of immunity. Parameter  $\lambda$  is the per capita rate of acquisition of illness. In general, for any route of transmission, it can be written as the product of the rate of exposure to the contaminant and the probability of illness given that exposure. The rate of exposure to the contaminant is the partial derivative of the dose function with respect to time, and the probability of illness given that exposure is the partial derivative of the response function with respect to dose. This formulation of  $\lambda$  is a generalization of that used by Chick et al. (2001).

Equations (6-7) are applied at each spatial node  $x_i$  in the network model. If the number of births in the population is assumed to exactly balance the number of deaths not due to exposure to contamination over the time period of interest, then the total population at each node is given by:

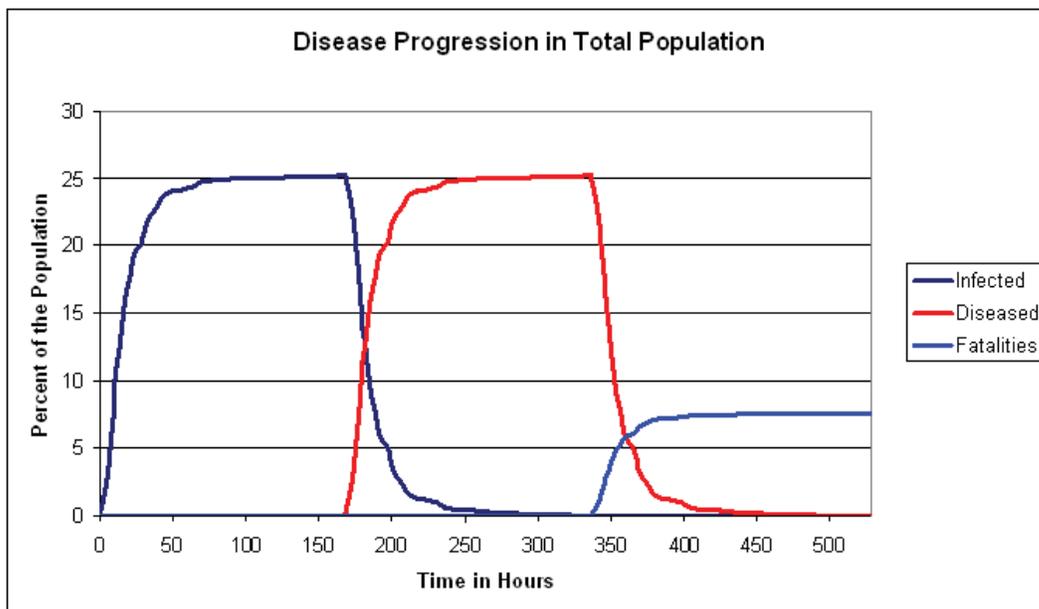
$$Pop_i = S + I + D + R + F \tag{6-8}$$

The populations can be summed in order to estimate the total number of infected, diseased, recovered, and fatally impacted in the total population at any point in time, for example:

$$\bar{I}(t) = \sum_{i=1}^N I(x_i, t) \tag{6-9}$$

**Figure 6-5** shows the output from modeling equations (6-7) for a biological agent over time. The curves show the percentage of infected people (I), the number of symptomatic people (D), and the percentage of fatalities (F) over the entire network. The slope of the infections curve is directly related to  $\lambda$ , the rate of acquisition of illness. This was calculated through equations (6-2)–(6-6) which incorporate all of the hydraulics of the contamination incident. The number of susceptible people who become infected quickly increases and then drops off to a very small number (not shown). The number of infected people increases rapidly, sustains itself as the disease is latent (for one week), and then drops quickly as the infected people transition into the diseased stage. Similarly, the number of symptomatic people increases rapidly, sustains itself for the duration of the illness (an additional week), and then a proportion of the symptomatic population recovers, and the remaining die (30% untreated fatality rate). Over the entire network, 25% of the population is infected after consuming contaminated water.

The health impacts methodology described here allows users to estimate the spatial and temporal distribution of health impacts resulting from ingestion of contaminated drinking water. The method is flexible enough to accommodate most types of acute illnesses from chemical or biological sources. The model could be extended to incorporate exposure through dermal and inhalation routes, and to incorporate person to person transmission. For more information about this methodology, see Murray et al. (2006b).



**Figure 6-5.** The spread of disease over time in a population exposed to a biological agent through drinking water.

**Modeling Other Consequences.** In addition to estimating the public health consequences, three other consequence measures are included in TEVA-SPOT.

The extent of contamination, or the number of feet of pipe contaminated during a contamination incident, is a useful measure of the economic impacts of an incident. It is an indication of the length of pipe that might need to be super-chlorinated, decontaminated, re-lined, or replaced following a contamination incident with a persistent contaminant. This consequence metric can be estimated according to:

$$EC = \sum_{i=1}^N L(x_i, t_j) \text{ if } C(x_i, t_j) > 0 \text{ for any } j \quad (6-10)$$

where  $L$  is a pipe with flow starting at node  $x_i$ .

The mass consumed metric is the mass of contaminant that is removed from the distribution system by consumer demand. This includes the mass of contaminant that is ingested by consumers, and also the mass of contaminant present in the water used for watering lawns, washing clothes, or any other consumer use. Mass consumed for each incident is calculated according to:

$$MC = \sum_{j=1}^T \sum_{i=1}^N C(x_i, t_j) q(x_i, t_j) \Delta t \text{ if } C(x_i, t_j) > 0 \text{ for any } i, j \quad (6-11)$$

where  $C$  is the concentration of the contaminant,  $q$  is the demand, and  $\Delta t$  is the time step.

The volume consumed is the volume of contaminant that is removed from the distribution system by consumer demand. Volume consumed for each incident is given by:

$$VC = \sum_{j=1}^T \sum_{i=1}^N q(x_i, t_j) \Delta t \text{ if } C(x_i, t_j) > 0 \text{ for any } i, j \quad (6-12)$$

## List of Variables

$q$	Demand at a node [Volume/Time]
$MR$	Mass injection rate [Mass/Time]
$V$	Volume of the contaminant [Volume]
$C$	Concentration of the contaminant [Mass/Volume]
$D$	Duration of the contaminant injection [Time]
$Pop$	Population at a node [.]
$R_{pc}$	Per capita daily rate of water consumption [Volume/Day]
$d$	Cumulative dose of contaminant ingested by consumers at a node [Mass]
$P_w$	Probability of a consumer ingesting water at time $t$ [.]
$V_w$	Volumetric rate of water consumption at time $t$ [Volume/Time]
$r$	Response at a given dose [.]
$\Phi$	Cumulative distribution function of a log-normal distribution
$\beta_1$	Parameter in the log-probit dose response curve
$\beta_2$	Parameter in the log-probit dose response curve
$\tau$	Parameter in the logistic dose response curve
$S$	Number of people at each node susceptible to illness [.]
$I$	Number of people at each node exposed to a lethal or infectious dose [.]
$D$	Number of people at each node experiencing the symptoms of illness [.]
$R$	Number of people recovered from illness [.]
$F$	Number of fatalities resulting from illness [.]
$v$	Per capita recovery rate ( $1/v$ is the mean duration of illness), [1/Time]
$\sigma$	Inverse of the mean latency period, [1/Time]
$\alpha$	Per capita untreated death rate, [1/Time]
$\gamma$	Per capita rate of loss of immunity [1/Time]
$\lambda$	Per capita rate of acquisition of illness [1/Time]
$L$	Pipe link in model
$EC$	Extent of contamination [Length]
$MC$	Mass consumed [Mass]
$VC$	Volume consumed [Volume]

# Optimization Methodology

This chapter describes several fundamental sensor placement methods that are included in the TEVA-SPOT software. The model formulations are presented without going into extensive detail regarding the actual solution techniques, and references to additional information are provided. The implications of algorithmic choice are also considered in terms of running time, memory (size of the machine necessary to run the optimization), and confidence in the final sensor placement solution.

As described in Chapter 6, TEVA-SPOT simulates contamination incidents in the Simulation Module and calculates impacts in the Consequence Assessment module. The Sensor Placement Module, then, optimizes sensor locations. Appendix A discusses other possible approaches to the sensor placement problem, including some that model contamination movement as part of the optimization problem. To date, such models have used average velocities or other approximations that are likely to be much less realistic than the approach used in TEVA-SPOT.

## Sensor Placement Problem

The Consequence Assessment Module output file contains a list of all the contamination incidents and the calculated impacts of those incidents over time in terms of a specific performance measure. As described in Chapter 2, performance measures can include the number of incidents detected, the number of people exposed to contaminants, the length of pipe contaminated, among others. The sensor placement problem is described as locating a set of sensors in order to best minimize this set of impacts; e.g., minimizing detection times.

When selecting sensor locations that minimize the mean impacts over a set of contamination incidents, this problem is equivalent to a well-known problem from the facility location literature: the *p*-median facility location problem (Mirchandani et al. 1990), in which *p* facilities must be located in such a way that the distance from each facility to its customers is minimized. The specific structure of sensor placement problems in water distribution networks leads to *p*-median problems that are relatively easy to solve, even if the networks have tens of thousands of junctions (Berry et al. 2006b). This is fortunate, since there are examples in the *p*-median literature of much smaller instances using other applications that have proven much harder to solve in practice.

The classic *p*-median facility location problem can be illustrated as follows. Consider the layout of a city, and imagine that *p* fire stations must be located in order to best serve the city's residents and infrastructure. Each house and building in the city is a *customer*, and each fire station a *facility*. Given a proposed set of locations, the *p*-median

objective is to minimize the average distance from a customer to the nearest facility. One could assign fire stations using nothing more than eyesight and a city map, but the optimization techniques described below do much better.

For the drinking water sensor placement problem, the sensors are facilities analogous to the fire stations. However, the analogue to customers is more subtle. Each contamination incident is a single "customer." A contamination incident propagates contaminated water through the network and is "served" from the network users' point of view, by the first sensor facility that detects the contamination. By modeling sensor placement as a *p*-median problem, the actual network topology (which pipes are connected to which junctions) is not required for optimization. These topological details are only considered during the water quality simulations that produce the impact information (the Simulation and Consequence Assessment modules). The *p*-median formulation merely requires a list of potential facilities for each customer (locations where sensors could observe an incident) and the associated service costs. For the fire station example, these costs are distances and for the water sensor placement problem, the costs are contamination impacts to people and/or infrastructure. TEVA-SPOT measures these impacts in terms of performance objectives like the time of detection or the number of people exposed.

## Solution Options

Given a *p*-median problem, there are many possible ways to solve it. TEVA-SPOT provides three general optimization methods: mixed-integer programming (MIP), a Greedy Randomized Adaptive Search Procedure (GRASP) heuristic, and a Lagrangian relaxation method. These optimizers vary in runtime, the amount of computer memory required, and the guarantee provided for solution quality. Generally, a method that gives a stronger quality guarantee will require more time and/or memory.

The MIP solvers for the *p*-median algorithms are *exact*. They produce solutions that are provably optimal, given the input data. The GRASP solvers are *heuristic*, meaning that their solutions tend to be good, but not provably optimal. The Lagrangian method produces a *lower bound*, a value guaranteed to be no larger than the optimal objective. A lower bound can provide higher confidence in the quality of a heuristic solution. For example, when a heuristic method returns a value with a small percentage difference from a lower bound, then decision makers can be confident the heuristic solution is good.

The great challenge of the drinking water sensor placement problem is that the set of contamination incidents can be much larger than the set of customers in a more conventional facility location problem. Threat ensembles that attempt

to be comprehensive for location, time of day, season, day of the week, and contamination type can be very large. Consequently, solution methods applied to corresponding  $p$ -median problems can easily exceed the memory available on a standard desktop computer or Unix/Linux workstation. TEVA-SPOT includes methods to reduce memory requirements; however, this is usually at the price of reduced solution quality.

Because the Simulation Module and Consequence Assessment Module are distinct from the Optimization Module, users can try multiple types of solution methods on any particular large problem without repeating these simulation/assessment runs. For example, one can experiment with different objectives, different solvers, or search over error parameters with a single objective until the system returns a satisfactory solution. Even if simulation methods or incident generation methods improve, the optimization methods remain viable, since the optimization can be rerun with the new input.

### Mixed-integer programming

A MIP is the optimization (minimization or maximization) of a linear objective function subject to linear constraints. Some of the variables must take on integer values (no fractional parts), but others can take on continuous values. There is a large body of theoretical work in operations research supporting MIP solution technology. When usable, this technology will do the best possible job of optimization — it will find optimal solutions.

MIP technology is usable if the problem instances are not too large, and if MIP solvers of sufficient power are available. Commercial MIP solvers are generally the fastest and most reliable. However, they cost tens of thousands of dollars for a license. Typically, free MIP software like the PICO solver available in TEVA-SPOT is sufficient to optimize  $p$ -median problems for moderate-sized water networks.

The MIP formulation for sensor placement (SP) is essentially a  $p$ -median formulation:

$$(SP) \text{ minimize } \sum_{a \in A} \alpha_a \sum_{i \in L_a} d_{ai} x_{ai} \quad (7-1)$$

Where:

$$\begin{aligned} \sum_{i \in L_a} x_{ai} &= 1, \quad \forall a \in A \\ x_{ai} &\leq s_i, \quad \forall a \in A, i \in L_a \\ \sum_{s_i \in \{0,1\}} s_i &\leq p, \quad \forall i \in L \\ 0 &\leq x_{ai} \leq 1, \quad \forall a \in A, i \in L_a \end{aligned} \quad (7-2)$$

and  $A$  is the set of contamination incidents. The SP MIP (equations (7-1)–(7-2)) minimizes the expected impact of a set of contamination incidents; this is a weighted average. Since dividing by the number of incidents does not change

the optimal set of locations (only the value), that division is not represented here.

For each incident,  $a \in A$ ,  $\alpha_a$  is the weight of incident  $a$ . This weight could be a probability of that incident occurring or it might represent a relative frequency (e.g., an incident that could occur on Monday–Friday might have a weight of 5/7 while a weekend incident might have a weight of 2/7). The set of locations denoted  $L$  is the set of network junctions with nonzero concentrations for one or more incidents (as determined by EPANET simulations). For each incident  $a$ ,  $L_a \subseteq L$  is the set of locations that can be contaminated by  $a$ . Thus, a sensor at a location  $i \in L_a$  can detect contamination from incident  $a$  when contamination first arrives at location  $i$ . Each incident is said to be “witnessed” by the best (lowest impact) sensor that sees it. For each incident  $a \in A$  and location  $i \in L_a$ ,  $d_{ai}$  is the impact of the contamination incident  $a$  if it is witnessed by location  $i$ . This impact measure assumes that as soon as a sensor witnesses contamination, any further contamination impacts are mitigated, perhaps after a suitable delay that accounts for the response time of the water utility. The  $x_{ia}$  variable is 1 if incident  $a$  is witnessed by a sensor at location  $i$  and 0 otherwise. The  $s_i$  variables indicate where sensors are placed in the network ( $s_i = 1$  if there is a sensor placed at location  $i$  and 0 otherwise). There is a sensor budget  $p$  (place at most  $p$  sensors). SP allows placement of, at most,  $p$  sensors;  $p$ -median formulations generally enforce placement of all  $p$  facilities. In practice, the distinction is irrelevant unless  $p$  approaches the number of possible locations.

All contamination incidents might not be witnessed by a given set of sensors (unless every node has a sensor). To account for this,  $L$  contains a dummy location. This dummy location is in all subsets  $L_a$ . The impact for this location for a given incident is the impact of the contamination incident after the entire contaminant transport simulation has finished, which corresponds to the impact that would occur without an online CWS.

The first constraint in equations (7-2) ensures that exactly one location witnesses each incident. This could be the dummy location. The second constraint in (7-2) ensures that only locations with sensors can witness incidents. The third constraint in (7-2) enforces the sensor budget. The sensor placements variables are binary: they can only be 0 or 1. The witness variables are continuous, but must be between 0 and 1. In practice, there is always an optimal solution where the witness variables are binary. If they are ever fractional, then two or more equally good (best) locations have sensors.

### Linear-programming lower bound

A linear program (LP) is a MIP with no integrality constraints. That is, all variables can take continuous values not just integer values. If the binary constraints for the  $s_i$  variables in the SP formulation (i.e.,  $s_i = 0$  or 1) are replaced with linear constraints ( $0 \leq s_i \leq 1$ ) then the problem becomes linear. It is called the *LP relaxation* of the MIP, because the integrality constraints have been “relaxed.” Linear programs can usually be efficiently solved. The LP relaxation of SP

is not directly useful, because one cannot place a fractional portion of a sensor at a location and then receive a fractional portion of the benefit. However, any real sensor placement is also feasible for the LP, so the LP relaxation method can be used to find a lower bound on the optimal value for any integer solution.

### Reducing MIP size via aggregation

The size of the SP formulation is largely a function of the number of non-zero values in the impact matrix,  $d$ . This number is determined by the number of contamination incidents simulated and the number of locations contaminated by each incident. It is the dominant term in the number of constraints, the number of variables, and the number of non-zeros in the constraint matrix. Typical water distribution network models have 1,000s to 100,000s of pipes and junctions. The number of locations contaminated by an incident can be highly variable. Although many incidents impact a small number of locations, some large networks have many incidents that contaminate a large fraction of the network. Many of the SP analyses performed by the TEVA Research Team have had millions of impact values. Even with relatively small numbers of times per day in the threat ensemble — and not accounting for other sources of variability — typical problems can have tens of millions of impacts. More comprehensive threat ensembles will be considerably larger.

The SP MIP model provides a generic approach for performing sensor placement with a variety of design objectives. However, the size of this MIP formulation can quickly become prohibitively large, especially for 32-bit computers (yielding a maximum of 4GB of RAM for UNIX systems, and, in practice, 3GB of RAM for Windows systems).

For any given contamination incident  $a$ , there are often many impacts  $d_{ai}$  that have the same value. If a contaminant reaches two junctions at about the same time, then the total impacts across the network would be identical for both junctions. Arrival times can be indistinguishable when using a typical reporting time-step, such as a small number of minutes, for the water quality simulation. Even though the contamination plume may pass nodes at different times within a 5-minute period, EPANET reports them all as occurring at the end of the 5-min water quality time-step.

This observation leads to a revised formulation that treats sensor placement locations as equivalent if their corresponding contamination impacts are the same for a given contamination incident. Define  $L_{ai}$  as a maximal set of locations in  $A$  that all have the same impact for incident  $a$  (that is, this set contains all the locations with a particular shared impact value for incident  $a$ ). Recall that a witness is a sensor that can detect a contamination incident because it is on the same travel path. By considering any witness in  $L_{ai}$  as equivalent to any other, the set of effective witness “locations” for incident  $a$  is reduced to a new set  $\hat{L}_a$ . Each group of equivalent locations (for an incident) is a *superlocation* for that incident. The locations grouped

in a superlocation for an incident are not necessarily located physically close in the network even though the contamination for incident  $a$  reaches them at approximately the same time. The new MIP formulation is:

$$(\text{waSP}) \text{ minimize } \sum_{a \in A} \alpha_a \sum_{i \in \hat{L}_a} d_{ai} x_{ai} \quad (7-3)$$

Where:

$$\begin{aligned} \sum_{i \in \hat{L}_a} x_{ai} &= 1, \quad \forall a \in A \\ x_{ai} &\leq s_i, \quad \forall a \in A, i \in \hat{L}_a \\ \sum_{s_i \in \{0,1\}} s_i &\leq p, \quad \forall i \in L \\ 0 &\leq x_{ai} \leq 1, \quad \forall a \in A, i \in \hat{L}_a \end{aligned} \quad (7-4)$$

The waSP model (equations (7-3)–(7-4)) revises SP to exploit structure that can make the MIP formulation smaller. The “wa” stands for “witness aggregation,” the term that describes this type of problem compression. This MIP selects both a superlocation to witness an incident and an actual sensor from the group in the superlocation. The fundamental structure of this formulation changes only slightly from SP, but in practice this MIP often requires significantly less memory. Specifically, grouping  $k$  equivalent locations removes  $k-1$  entries from the objective,  $k-1$  variables, and  $k-1$  constraints. Every feasible solution for SP has a corresponding solution in waSP with the same sensor placement. The selected observation (witness) variable can always be mapped to a real sensor with the same impact. Because the impact for each incident is the same, the objective value is the same, so waSP can be used to find optimal sensor placements.

The waSP formulation can be generalized to consider location values as equivalent if their impact values are *approximately* equal. For each incident  $a$ , consider a list of locations in  $L_a$  sorted by impact. A superlocation is a contiguous sublist of this sorted list. Generally, locations are grouped into a superlocation if the difference in their impact values meets a given threshold. For waSP, that threshold was equality. Berry et al., (2006b), describes two ways for creating superlocations: (1) the ratio of largest to the smallest impact in the superlocation is small [*ratio aggregation*], and (2) the total number of witnesses for any incident is small. The first type keeps the error low, but might not provide a lot of compression. The second type guarantees compression, but might introduce large errors.

TEVA-SPOT also allows grouping with an absolute threshold, where the difference between the largest and smallest impact is small. Recall  $\hat{L}_{ai}$  is the set of superlocations for incident  $a$ , and  $\tilde{L}_{ai} \subseteq L_a$  is the set of (real) locations in the  $i$ th superlocation for incident  $a$ .

Define  $\tilde{d}_{ai}$  to be the largest impact value for incident  $a$  if witnessed by any location in  $\tilde{L}_{ai}$  (that is,  $\tilde{d}_{ai} = \max_{i \in \tilde{L}_{ai}} d_{ai}$ ).

Then, define  $x_{ai}$  as a binary variable that is 1 if incident  $a$  is witnessed by some location in  $\tilde{L}_{ai}$  and 0 otherwise. Then the MIP for general witness aggregation is the waSP formulation where  $d_{ai}$  is replaced by  $\tilde{d}_{ai}$  and  $\hat{L}_{ai}$  by  $\tilde{L}_{ai}$ .

Berry et al. (2006b) proved that the optimal solution to a problem with ratio aggregation is guaranteed to be an approximation for the original problem with quality proportional to the ratio. However, a user must determine a good threshold via careful experimentation.

### Incident Aggregation

In some cases, one can replace a pair or a group of contamination incidents with a single new incident that is equivalent. Berry et al. (2006b) describes one such strategy (called scenario aggregation in that paper for historical reasons). This aggregation strategy combines two incidents that contaminate the same locations in the same order, although one incident might stop before the other. For example, two injected contaminants should travel in the same pattern if they differ only in the nature of the contaminant, though one might decay more quickly than the other. Two such incidents can be combined into one by simply averaging their impacts and adding their incident weights.

### Effectiveness of Aggregation

These aggregation techniques significantly improved the ability to apply MIP solvers to real-world sensor placement applications. One might need to use the waSP formulation to solve large sensor placement problems, even on high-end workstations with large memory. For example, Berry et al. (2007), describe the use of witness aggregation on sensor placement models derived from water networks with over 3,000 pipes and junctions. These results are summarized in **Table 7-1**. The  $\rho$  value varies from 0 to 1 and indicates the ratio used to control witness aggregation. When  $\rho$  is nonzero, witnesses are aggregated into groups such that the ratio of best-to-worst impact values does not exceed  $\rho$ . (Note that when  $\rho$  is one, all of the witnesses are aggregated together.) When  $\rho$  is zero, witnesses with the same impacts are aggregated, which can reduce the number of non-zeros in the MIP model by almost a factor of three. Similarly, the runtime is reduced by a factor of three. An appropriate level of aggregation significantly reduces the size of the MIP model and the corresponding runtime. However, the solution quality deteriorates as the sensor placement model becomes more approximate.

**Table 7-1.** Reduction of MIP problem size using witness aggregation with different ratios ( $\rho$ ). The IP value shows the value predicted by the aggregated problem, and the true value is the value of that solution evaluated in the original non-aggregated  $\rho$ -dependent problem.

$\rho$	# variables	# constraints	# nonzeros	Runtime (sec)	IP value	True value
None	16854011	16850654	67334870	79504	1186	1186
0	2506339	2502982	23770968	22415	1186	1186
0.125	31323	27966	12169827	722	25	2060
0.25	18025	14668	9842434	322	6	2743
0.5	7179	3822	3416662	17	0.1	9302

### The GRASP Heuristic

A combinatorial heuristic exploits properties of combinations of objects. In our context, these objects are sensors and the combinations are the possible ways to place those sensors in a water network. TEVA-SPOT contains the current state-of-the-art combinatorial heuristic for p-median problems, an adaptation of Resende and Werneck's GRASP algorithm (Resende et al. 2004). GRASP finds good solutions to p-median problems by systematically exploring the space of possible sensor layouts. It usually (experimentally) produces solutions as good as MIP solutions, but much faster. However, there is no provable performance guarantee.

GRASP randomly constructs a set of starting points, using greedy bias to make these solutions reasonable approximations. Then for each candidate solution, it considers ways to move a single sensor to a location that improves the objective. It makes the best swap of this type repeatedly until no improving swap exists. The general GRASP technique normally considers combinations of these local optima, but generally taking the best solution suffices for this sensor placement application.

The GRASP heuristic can find solutions to very large p-median instances (with over 10,000 facilities and 50,000 customers) in approximately ten minutes on a modern workstation-class computer (Ostfeld et al. 2008). This is approximately 5 to 10 times faster than the commercial MIP code CPLEX® (CPLEX Optimization, Inc.) could solve the waSP MIP formulation. The GRASP solutions were often optimal, as verified by comparison with exact solutions to the MIP formulation. The only drawback to the GRASP heuristic is the memory requirements, which reached 16GB of RAM for these large instances. This capacity is beyond the limits of what is available in most end-user environments for which CWS design is targeted.

Because the cost of determining the decrease in total impact during a local search move is dominated by the lookup cost of specific  $d_{ai}$  impact values, the GRASP heuristic creates a dense matrix of all impacts. The dense matrix represents unnecessary zeros, but it gives fast (constant-time) lookup of the  $d_{ai}$ . An alternative sparse representation simply stores, for each  $a \in A$ , a tree containing pairs  $(i, d_{ai})$  for all  $i$  touched by incident  $a$ . The trees require logarithmic (in the number of defined  $d_{ai}$  for a given  $a$ ) time to look up an impact value. In practice the slow-down is less than 50%, and the memory requirements are reduced by a factor of four or more.

TEVA-SPOT provides variants of the GRASP heuristic using the dense and sparse storage schemes for the  $d_{ai}$ . Even with the sparse representation, there are large real-world problems too large for 32-bit workstations. Users can reduce the problem size further by, for example, restricting the number of locations for sensors. This can help the GRASP heuristic considerably, since it reduces the search space during iterations of the swapping portion. This space-reducing measure requires the users to expend effort to determine infeasible locations, rather than determining feasibility as necessary during network design.

### The Lagrangian Heuristic

A Lagrangian method works by removing a set of “difficult” constraints, leaving behind a problem that is easy to solve. It then applies pressure to satisfy the relaxed (dropped) constraints by adding penalties to the objective function. These penalties are proportional to the constraint violations. Thus there is no penalty if a constraint is met, a small penalty for a small violation, and a larger penalty for a larger violation. By manipulating the penalty weights (called *Lagrange multipliers*), an iterative algorithm can drive the solution towards feasibility. Using the TEVA-SPOT Lagrangian solver, each optimal solution to such a relaxed problem gives a lower bound for the original  $p$ -median problem.

The Lagrangian solver is composed of a Lagrangian-based lower-bounding procedure and an approximation heuristic. This solver requires memory proportional to  $n + D$ , where  $n$  is the number of sensor locations and  $D$  is the total number of impacts. This is within a constant factor of the smallest possible memory requirement for a program that does not explicitly move data back and forth from secondary memory (like disk farms).

The Lagrangian-based lower-bounding method is based on the method described by Avella et al. (2007). Given a set of Lagrange multipliers, one can compute the optimal solution for that particular relaxation quickly. Based on work for a similar problem by Barahona and Chudak (2005), the Barahona and Anbil’s subgradient search method, called the Volume Algorithm (Barahona et al. 2000), is used to find Lagrangian multipliers that produce progressively higher lower bounds. This search converges to a set of Lagrange multipliers for which the optimal solution to the relaxed problem is an optimal solution to the  $p$ -median LP relaxation. Thus the Lagrangian solver computes the LP relaxation using considerably less memory than an LP solver would. Finally, the Lagrangian solver uses a constrained rounding algorithm to randomly select  $p$  sensor locations biased by the LP relaxation.

The Lagrangian relaxation model relaxes the first set of constraints in the SP formulation — those that require each incident be witnessed by some sensor. Recall that this might be the dummy sensor which indicates a failure to detect the incident. This constraint is written as an equality, because that is a more efficient integer programming formulation. However, the difficult part of the constraint is insuring that at

least one sensor witnesses each incident. The objective will prevent over-witnessing, so for the sake of the Lagrangian relaxation, these constraints are treated as inequalities. For some incident  $a$ , this constraint is violated for a proposed setting of the  $s_i$  and  $x_{ai}$  variables if  $\sum_{i \in L_a} x_{ai} < 1$  giving a violation of  $(1 - \sum_{i \in L_a} x_{ai})$ . Each such violation is weighted with its own Lagrange multiplier  $\lambda_a$ , which allows some violations to be penalized more than others.

Adding a penalty term  $\lambda_a(1 - \sum_{i \in L_a} x_{ai})$  to the objective for

each incident  $a$ , the Lagrangian model becomes:

$$(\mathbf{LAG}) \text{ minimize } \sum_{a \in A} \left( \alpha_a \sum_{i \in \hat{L}_a} (d_{ai} - \lambda_a) x_{ai} \right) + \sum_{a \in A} \alpha_a \lambda_a \quad (7-5)$$

Where:

$$\begin{aligned} x_{ai} &\leq s_i, \quad \forall a \in A, i \in \hat{L}_a \\ \sum_{s_i \in \{0,1\}} s_i &\leq p, \quad \forall i \in L \\ 0 &\leq x_{ai} \leq 1, \quad \forall a \in A, i \in \hat{L}_a \end{aligned} \quad (7-6)$$

For a fixed set of  $\lambda_a$  the optimal value of LAG can be quickly computed using low memory with a slight variation on the method described by Avella et al. (2007). The optimal solution to LAG gives a valid lower bound on the value of an optimal solution to the  $p$ -median (SP) problem. This is because any feasible solution to the  $p$ -median problem is feasible for LAG. It has a zero violation for each of the lifted (relaxed) constraints and a value equal to the original  $p$ -median value.

Given a fractional solution to the  $p$ -median LP, the fractional values are treated as probabilities and sensors are selected randomly according to this probability. However, one is unlikely to get precisely  $p$  sensors this way. A variant of conditional Poisson sampling is used to efficiently sample over the “lucky” distribution where exactly  $k$  sensors are selected. If necessary, the dummy location is selected.

In preliminary tests with a moderate-sized problem (the same size as those Avella et al. (2007) call “large-scale”), the Lagrangian method required approximately one third the memory of the GRASP heuristic, and usually found a solution almost as good while running up to 2.5 times longer. For example, on a problem with 3358 locations, 1621 incidents, and 5 sensors, considering four different types of objectives, the Lagrangian solver required 45 megabytes (MB) of memory and the GRASP heuristic required 154 MB of memory. The GRASP heuristic found the optimal solution in all four cases as verified by the MIP. The Lagrangian heuristic was within 0.5% of this for three out of the four objectives (PE, EC, MC). Running times for GRASP ranged from 33.8 seconds to 44 seconds. The Lagrangian ran in less than 86 seconds for these 3 objectives. For the fourth

objective (VC), Lagrangian ran for 105 seconds and had a gap of 64%, showing that the Lagrangian behavior can be less stable than GRASP.

Witness aggregation can be used to further reduce the memory required for the Lagrangian method, particularly aggregation of locations that have the same impact values. However, the set-cover constraints (the second set of constraints in the waSP formulation) cannot be used without altering the Lagrangian model. The current version in TEVA-SPOT runs the heuristic with the aggregated witnesses where the superlocations are not directly associated with their constituent locations. This creates a straight p-median problem for the Lagrangian solver that now no longer has the same optimal solution. Because there are fewer opportunities to witness incidents, this revised formulation has a higher optimal impact, and therefore the current Lagrangian solver does not give a valid lower bound. However, a heuristic solution can still be computed by solving this modified problem and mapping superlocations back to real locations. The current version simply selects the first real location in a superlocation list.

For a large-scale problem with 42,000 junctions, the Lagrangian heuristic required only 100Mb for the aggregated problem where we equated only witnesses of equal impact. This is a considerable reduction from the 1.8GB the Lagrangian method required with no witness aggregation, even of equal impact (the SP version). The GRASP heuristic required 17GB; there is no value for witness aggregation in the GRASP heuristic, so this is the memory requirement for the SP version. However, the objective of the Lagrangian solution is 60% worse than the solution found by GRASP.

### Alternative Objectives and Multiple Objectives

TEVA-SPOT also provides solvers for variations on the average-impact objective function. This includes simultaneously considering multiple impact types and considering objectives over the distribution of impact values that are arguably more robust.

For any particular network and set of contamination incidents, there can be many types of damage to people and/or to the water distribution network. Some initial research has shown that optimizing for one particular objective, such as minimizing the average number of people exposed to lethal levels of a contaminant, can lead to solutions that are highly suboptimal with respect to other objectives, such as minimizing the total pipe feet contaminated (Ostfeld et al. 2008; Watson et al. 2004).

SPOT allows users to seek compromise solutions among multiple types of average impacts with *side constraints*. Users choose an objective, say PE (population exposed). They can also put a bound on the average impact for another

measure, say EC (extent of pipeline contamination). For example, the user can ask for a sensor placement that minimizes the average PE subject to a constraint that at most 1000 feet of pipe are contaminated.

The MIP solver treats side constraints as hard. That is, it does not consider a sensor placement feasible unless it meets the side constraint bound. For the MIP solver, the side constraint is simply an additional linear constraint. The GRASP and Lagrangian solvers treat the side constraints as soft *goal constraints*. They might return a solution that violates one or more side constraints, but it tries to meet the goals. They do this by adding another penalty term to the objective, this time penalizing violation of the side constraint. Currently the GRASP solver cannot handle more than one side constraint. The other solvers can handle an arbitrary number, but currently the Lagrangian solver's solution quality degrades considerably with more than one side constraint. In all cases, the side-constrained case will take longer to solve than the single-objective case. The GRASP solver might have trouble finding a feasible solution. The user will generally have to use trial and error to find side-constraint bounds that produce good compromise solutions.

One solution  $X_1$  dominates another  $X_2$  if the average impact of  $X_1$  is no worse than the average impact of  $X_2$  in all measurement categories. In general, there might be many non-dominated solutions (*pareto optimal points*), points for which there is no feasible solution that dominates it. None of the solvers will currently produce multiple pareto-optimal points at once, but they all can produce different non-dominated points by varying which impact measure is the objective and which is the side constraint, and by varying the bounds of the side constraints.

### Robust Objectives

In general, the budget for placing sensors will be limited. For a reasonably comprehensive suite of incidents, there will be some incidents that are not covered well and usually some that are not covered at all. The network designer must decide where they are willing to take risks. TEVA-SPOT offers three other objectives over the distribution of incident impacts to give the designer more flexibility in controlling risk. The first is minimizing the max impact taken over all incidents.

The second robust objective is called VaR, which stands for "value at risk." Given a percentage  $\gamma$ , VaR  $v$  is the impact value such that a  $(1 - \gamma)$  fraction of the incidents have impact no larger than  $v$ . For example, if  $\gamma = 0.05$  and  $v = 450$ , that means that 95% of the events have impact no more than 450. This means that the designer is choosing to ignore the *tail* ( $\gamma$  fraction) of the highest-impact incidents, but expects a minimum-quality coverage for all the others.

The final robust objective is called CVaR, for conditional value at risk. Given a tail percent  $\gamma$ , CVaR minimizes the average of the worst  $\gamma$  fraction. In the example above, this objective finds a solution that minimizes the average impact of the worst 5% of the incidents.

Currently, all of these robust measures are currently significantly harder to compute in practice than the average

impact. Optimizing any of these objectives will almost certainly increase the average impact. See Watson et al. (2009) for discussion of some of these issues.

**Table 7-2** summarizes the capabilities of the three solvers in TEVA-SPOT.

**Table 7-2.** TEVA-SPOT solver capability summary.

Feature	Integer Program	GRASP	Lagrangian
Min mean impact	yes	yes	yes
Min max impact	yes	yes	no
Min # sensors	yes	no	no
Robust impact measures	yes	yes	no
Side constraints	yes	yes	yes
Fixed/invalid locations	yes	yes	yes
Imperfect sensors	yes	yes	no
Computes lower bound	yes	no	yes
Aggregation	yes	no	yes

### List of Variables

- $A$  Set of contamination incidents
- $a$  Single incident
- $\bar{\alpha}_a$  Weight of contamination incident  $a$
- $i$  Location in network (junction or node)
- $L$  Set of all locations in network
- $L_a$  Set of locations contaminated by incident  $a$
- $d_{ai}$  Impact of contamination incident  $a$  at location  $i$
- $x_{ai}$  Witness indicator: 1 if incident  $a$  is witnessed at location  $i$  and 0 otherwise
- $s_i$  Sensor indicator: 1 if  $a$  sensor is at location  $i$  and 0 otherwise
- $p$  Total number of sensors allowed
- $L_{ai}$  Set of locations with same impact from incident  $a$
- $\lambda_a$  Lagrangian Multiplier



# Appendix A.

## Literature Review

A variety of technical challenges need to be addressed to make contamination warning systems (CWSs) a practical, reliable element of water security. A key aspect of CWS design is the strategic placement of sensors throughout the distribution network. Given a limited number of sensors, a desirable sensor placement minimizes the potential impact to public health of a contaminant incident.

The following sections describe how authors have defined sensor placement problems and then review methods used to solve these problems. There has been a large volume of research on this topic in the last several years, including a Battle of the Water Sensor Networks (Ostfeld et al. 2008) that compared 15 different approaches to solving this problem. This review largely focuses on optimization methods for sensor placement, since the majority of published sensor placement techniques use optimization; 50 papers on sensor placement optimization are reviewed here.

### Contamination Risks

There are a large number of potentially harmful contaminants and a myriad of ways in which a contaminant can be introduced into a water distribution system. Physically preventing all such contamination incidents is generally not possible. Consequently, the overall goal of sensor placement is to minimize contamination risks.

Expert opinion and ranking strategies do not explicitly quantify contamination risks. For example, these methods do not compute the consequences of different contamination incidents or use this information in a risk comparative risk assessment. Instead, these strategies rely on human judgment to assess how a sensor network would minimize contamination risks. For example, a human expert can predict the likelihood of contamination injections occurring at different locations throughout the network based on local knowledge of the physical layout of the water distribution system. This information can guide the evaluation of effective sensor locations.

In contrast, optimization strategies generally rely on some form of computational risk assessment to guide sensor placement optimization. An optimization strategy uses a model of the water distribution network to predict how a contaminant flows through the network. This information is then used to assess the impact of contamination incidents (e.g., health effects or extent of contamination), which will vary depending on the contaminant type (including fate and transport characteristics), contaminant injection characteristics (e.g., source location, mass flow rate, time of day, and duration), and network operating conditions. All sensor placement optimization strategies developed to date

assume a particular finite set of contamination incidents, which define the threat basis for evaluating and mitigating contamination risk.

Optimization strategies can be categorized based on how the water distribution system network model is used for risk assessment. Early sensor placement research computed risk using simplified network models derived from contaminant transport simulations. For example, hydraulic simulations can be used to model stable network flows (Berry et al. 2005c; Lee et al. 1992; Lee et al. 1991), or to generate an averaged water network flow model (Ostfeld et al. 2004).

Most subsequent optimization research has directly used contaminant transport simulations to minimize contamination risks (Berry et al. 2006b; Ostfeld et al. 2004; Propato et al. 2005). Simulation tools, like EPANET (Rossman 1999, 2000), perform extended-period simulation of the hydraulic and water quality behavior within pressurized pipe networks. These models can evaluate the expected flow in water distribution systems, and they can model the transport of contaminants and related chemical interactions. Thus, the CWS design process can directly minimize contamination risks by considering simulations of an ensemble of contamination incidents, which reflect the impact of variables including contamination at different locations and times of the day.

There have been few direct comparisons of optimization strategies based on simplified versus detailed network model simulations (see Ostfeld et al. 2008; Berry et al. 2005b). Optimization strategies using contaminant transport simulations are clearly attractive because they provide a detailed risk assessment that accurately integrates the impacts of distinct contamination incidents. For example, optimization methods using simplified network models can fail to capture important transient dynamics. However, a potentially large number of contamination incidents might need to be simulated to perform optimization with contamination transport simulation. Consequently, it is very expensive to apply generic optimization methods like evolutionary algorithms (Ostfeld et al. 2004) when simulations are performed to evaluate each new sensor placement. A variety of authors have discussed how to perform simulations efficiently in an off-line preprocessing step that is done in advance of the optimization process (Berry et al. 2006b; Chastain 2006; Krause et al. 2008; Propato 2006). Thus, the time needed for simulation does not impact the time that a user spends performing optimization. This is a general strategy for managing simulation data that can be used by many different optimizers; for example the TEVA-SPOT Toolkit integrates a variety of optimizers that

employ this strategy (Berry et al. 2008a; Berry et al. 2007; Berry et al. 2006a; Berry et al. 2009; Berry et al. 2005a; Berry et al. 2006b; Berry et al. 2008b; Hart et al. 2008a; Murray et al. 2006a; Watson et al. 2005).

## Sensor Characteristics

Characterization of sensor behavior is required to predict the performance of a CWS. Researchers developing optimization strategies have commonly assumed a perfect sensor: a sensor with a detection limit of zero that is 100% reliable. Although this is clearly unrealistic, the assumption of perfect sensors can provide an upper bound on CWS performance. A slightly more realistic modeling assumption is to assume a detection limit for sensors: above a specified concentration, the sensor is 100% reliable, and below that concentration the sensor always fails to detect the contaminant. This approach allows users to model sensors that are not contaminant-specific, such as chlorine sensors that might indirectly detect the presence of a contaminant.

Few researchers have developed sensor network design optimizers that allow for sensors that sometimes fail to detect contaminants. A simple way to characterize sensor failures is to include a likelihood factor, which could be dependent on the sensor detection limit. Berry et al. (2006a; 2009) describe optimizers that allow for sensors with known false negative (FN) and false positive (FP) rates. Recently, McKenna et al. (2008) have used Receiver Operating Characteristic (ROC) curves to characterize the performance of sensors, and a sensor's FN and FP rates can be directly derived from ROC curves. In general, the FN and FP rates could depend on the location at which the sensor is being placed, the type of sensor, and the detection threshold.

## Sensor Placement Objectives

There are many competing design objectives for placing sensors in an online sensor network. Although minimizing impacts to public health is a widely accepted goal, there are several types of health impact objectives:

- Population exposed: The number of individuals exposed to a contaminant.
- Population dosed: The number of individuals exposed to a specified dose of contaminant.
- Population sickened: The number of individuals sickened by a contaminant.
- Population killed: The number of individuals killed by a contaminant.

Further, researchers have developed optimization methods for a variety of other objectives:

- Extent of contamination: The total feet of pipes contaminated before a contaminant is detected

- Mass of contaminant consumed: The mass of contaminant that has left the network via demand at junctions in the network.
- Percent detected: The fraction of contamination incidents that are detected by the sensors.
- Time to detection: The time from the beginning of a contamination incident until the first sensor detects it.
- Volume consumed: The volume of contaminated water that has left the network via demand at junctions in the network.

There are several modeling decisions that affect these design objectives. The first concerns how a utility responds when a sensor detects a contaminant. Computational models of CWS performance typically make the assumption that there is a response time after which contaminants are no longer consumed or propagated through the network (Murray et al. 2008b; Ostfeld et al. 2005b). Response time is often viewed as the time between initial detection of an incident and effective warning of the population (Bristow et al. 2006), and the response time used for optimization can be factored into the computation of these design objectives.

The second modeling decision concerns how detection failures are handled. Most design objectives compute the impact of each contamination incident after it has been detected. But if an incident has not been detected by the end of simulation, then the appropriate impact of that incident is unclear, since it might have been detected later if the simulation had run longer. Most optimization strategies compute the impact at the end of the simulation, which is equivalent to penalizing undetected incidents based on their undetected impact.

Several authors have suggested that these undetected incidents can be ignored (Berry et al. 2008b; Ostfeld et al. 2008). For example, when minimizing time-to-detection, this type of penalty scheme can skew the design towards simply detecting all incidents. However, a trivial optimal solution in this case would be to place no sensors; this design would then detect no incidents. This is clearly undesirable, so this type of performance objective only makes sense with the optimizer is constrained to ensure that a given fraction of the contamination incidents are detected.<sup>1</sup>

Finally, it is clear that users need to evaluate tradeoffs for several design objectives. The impact of this on the optimization process is described below.

## Optimization Objective

As was noted earlier, there are many possible contamination incidents that could be used as the design basis threat for a sensor placement problem. Thus, a sensor placement is evaluated using a distribution of impact values for the entire large set of contamination incidents. The mean impact is a

<sup>1</sup> Preliminary experiments with the TEVA-SPOT Toolkit suggest that it is much more difficult to optimize with this formulation than the more commonly used design objectives that penalize undetected incidents.

natural statistic for this optimization problem that is used by many researchers (see below). For example, Berry et al. (2006b) show that minimizing the mean impact for sensor placement is related to the well-known p-median problem for facility location.

Another optimization objective used by a variety of authors is to maximize the percent detected impact statistic, independent of other impacts. Although Berry et al. (2006b) show that this objective can be mathematically expressed as a mean impact, most researchers have developed optimization strategies that are more tailored to this particular objective. Specifically, this can be viewed as a covering problem, for which there is a rich optimization literature.

Watson et al. (2006; 2009) consider optimization strategies that minimize the max-case impact and other robust measures that focus strictly on high-consequence contamination incidents. A key motivation for considering robust optimization objectives is that an optimal sensor placement that minimizes mean impact might still have numerous high-impact contamination events. Watson et al. describe a variety of robust optimization objectives, including well-studied robustness measures from the financial community.

## Optimization Formulations

An optimization formulation is the mathematical definition of an optimization problem, which includes the decision variables, objective and constraints. For sensor placement problems, optimization formulations integrate modeling assumptions concerning how contamination risk is computed, the performance objective(s) that is optimized, the sensor characteristics, and other factors like feasible sensor locations and existing sensor stations. Thus, it is perhaps not surprising that a diverse array of optimization formulations have been developed for sensor placement.

**Table A-1** categorizes the optimization formulations used in the sensor placement literature with respect to the four factors described above. The majority of the research falls into one of nine groups based on these factors (shown in **Table A-1**). This classification highlights several trends and themes in the literature:

- **Contaminant Simulation:** The use of contaminant transport simulations is a consistent theme in recent sensor placement optimization research (groups 6–8). This reflects the fact that these optimization formulations can more accurately assess the impact of dynamic flows on contamination risks, as well as the fact that the necessary computational resources are more generally available.

- **Mean Impact:** Minimizing mean impact has emerged as the standard optimization formulation for sensor placement. Most early research focused on coverage formulations, which were adapted from early research on water quality management. However, the mean impact formulation can model a wide range of important impact measures, like health effects.
- **Multi-Objective Optimization:** The challenge of analyzing multiple objectives was highlighted by the Battle of the Water Sensor Networks challenge (Ostfeld et al. 2008), where four different objectives were used to evaluate sensor placements generated by the participants. A variety of standard multi-objective strategies have been applied for sensor placement:
  - Optimize a weighted-sum of different objectives
  - Optimize one objective while constraining the remaining objectives at goal values
  - Using a search strategy that searches for undominated points
- **Data Uncertainties:** A variety of authors have considered the impact of data uncertainties. For example, Chastain (2006; 2004) has performed sensitivity analysis of sensor placements. Similarly, Ostfeld and Salomons (2005a, 2005b) have used randomly generated data in their optimization formulation and assessed the impact of these uncertainties. A few authors have adapted their optimization to find more robust solutions. Shastri and Diwekar (2006) considered a stochastic optimization formulation that used a recourse model to capture the impact of uncertainties. Carr et al. (2006; 2004) and Watson et al. (2006; 2009) described robust optimization formulations that either minimize or constrain the max-case contamination incident impact values.

**Table A-1.** Summary of sensor placement optimization literature, categorized by: (a) whether contaminant transport simulations were used to compute risk, (b) whether sensor failures were modeled, (c) whether multiple design objectives were used during optimization, and (d) the type of optimization objective.

#	Citations	Risk Calculations with Simulation	Imperfect Sensors	Multiple Objectives	Optimization Objective
1	Al-Zahrani et al. 2001; Al-Zahrani et al. 2003; Kessler et al. 1998a; Kessler et al. 1998b; Kumar et al. 1997, 1999; Lee et al. 1992; Lee et al. 1991; Ostfeld et al. 2001; Uber et al. 2004	No	No	No	COVER
2	Berry et al. 2003; Berry et al. 2005c; Berry et al. 2005d; Rico-Ramirez et al. 2007; Shastri et al. 2006	No	No	No	MEAN
3	Carr et al. 2006; Carr et al. 2004	No	No	No	ROBUST
4	Watson et al. 2004	No	No	Yes	MEAN
5	Chastain 2006; Chastain Jr. 2004; Cozzolino et al. 2006; Ostfeld et al. 2003; Ostfeld et al. 2004, 2005a, 2005b	Yes	No	No	COVER
6	Berry et al. 2008a; Berry et al. 2007; Berry et al. 2004; Berry et al. 2006b; Berry et al. 2005d; Hart et al. 2008a; Kizileniş 2006; Propato 2006; Propato et al. 2005; Romero-Gomez et al. 2008; Watson et al. 2005	Yes	No	No	MEAN
7	Aral et al. 2008; Berry et al. 2008b; Dorini et al. 2006; Eliades et al. 2006; Guan et al. 2006; Gueli 2006; Hart et al. 2008b; Huang et al. 2006; Krause et al. 2008; Krause et al. 2006; Leskovec et al. 2007; Preis et al. 2006a; Preis et al. 2008; Wu et al. 2006	Yes	No	Yes	MEAN
8	Watson et al. 2006; Watson et al. 2009	Yes	No	No	ROBUST
9	Berry et al. 2006a; Berry et al. 2009	Yes	Yes	No	MEAN

A few other sensor placement formulations have been developed, but they do not neatly fall within these categories. Preis and Ostfeld (2006b) describe an optimization formulation that is intended to facilitate the analysis of sensor data to identify the source of a contaminant. Xu et al. (2008) describe an optimization formulation that does not use water quality simulations, but instead analyzes the topology of flows in a water distribution network to identify *interesting* locations for sensor placement. Finally, several sensor placement methods have been published in Chinese (Huang et al. 2007; Wu et al. 2008).

### Sensor Placement Optimizers

A variety of different sensor placement optimizers have been used to analyze the optimization formulations described above, including:

- Integer programming solvers
- Genetic algorithms
- Local search

Other well-known heuristic optimization methods have also been used (e.g., simulated annealing and tabu search), but most researchers have used one of these three optimizers in their research.

The choice of an optimizer for sensor placement is guided by several factors: the performance guarantee for the final solution, the available computer memory, and the runtime available for performing optimization. Integer programming (IP) solvers can guarantee that the best possible sensor placement is found (i.e., one that optimally minimizes the contamination risk). However, IP solvers are well-known to have difficulty solving large applications; on large problems they can run for a long time and require a lot of memory. By contrast, heuristic optimizers like genetic algorithms and local search methods cannot generally guarantee that the final solution is near-optimal. In practice, these methods are well-known to quickly find near-optimal solutions.

Krause et al. (2008; 2006) and Leskovec et al. (2007) describe the only sensor placement heuristic that is known to provide a performance guarantee. They consider a simple greedy local search method that is used to maximize the reduction of impact that a sensor placement provides. This optimization formulation differs from other authors, who focus on minimizing impact; the key observation of Krause et al. (2008; 2006) is that the structure of this formulation guarantees that a solution from this local optimizer is near-optimal.<sup>2</sup>

Similarly, several authors have demonstrated that lower bounds can be computed to evaluate whether solutions generated by heuristics are near-optimal. Berry et al. (2008a) describe a Lagrangian technique that computes a lower bound on the optimal sensor placement, and then uses a rounding heuristic to general a near-optimal solution. Watson et al. (2005) and Berry et al. (2006b) describe a GRASP heuristic for sensor placement. Their sensor placement formulation is equivalent to the well-known p-median facility location problem, and they show that the p-median IP model can be used to compute a lower bound on solutions generated by the GRASP heuristic.

A key issue for sensor placement optimizers is their ability to scale to large, real-world water distribution networks. Here, scalability refers to the ability of optimizers to perform a quick optimization on limited memory workstations. One strategy for ensuring scalability is to reduce the complexity of the water distribution system. This can be as simple as limiting the number of contamination locations and feasible sensor locations, which limits the size of the data need to represent the set of contamination incidents. More generally, the water network itself can be “skeletonized” to include aggregated junctions and pipes (see Perelman and Ostfeld (2008) for a recent review).

Sensor placement optimization can also be adapted to improve the scalability of the optimizer. For example, Preis and Ostfeld (2007) describe a procedure for selecting the key contamination incidents that are critical to evaluate a

sensor placement design. Similarly, Berry and others describe strategies for reformulating an integer programming model to reduce the number of constraints and decision variables (Berry et al. 2007; Hart et al. 2008b). Finally, low-memory optimization methods can be used to help ensure scalability. Hart et al. (2008a) describe optimization heuristics that are motivated by memory scalability concerns, and note that there are tradeoffs between runtime and memory usage that may influence the choice of a sensor placement optimizer.

## Supporting Decision Makers

Designing a CWS is not as simple as performing a single sensor placement analysis. There are many factors that need to be considered when performing sensor placement, including utility response, the relevant design objectives, sensor behavior, practical constraints and costs, and expert knowledge of the water distribution system. In many cases, these factors are at odds with one another (e.g., competing performance objectives), which makes it difficult to identify a single best sensor network design. Consequently, the design process requires informed decision making where sensor placement techniques are used to identify possible network designs that work well under different assumptions and for different objectives. This allows water utilities to understand the significant public health and cost tradeoffs.

Several researchers have focused on the decision-making process for CWS design. Murray et al. (2006a; 2008b) describe a decision framework composed of a modeling process and a decision-making process that employs optimization. This modeling process includes creating a network model for hydraulic and water quality analysis, describing sensor characteristics, defining the contamination threats, selecting performance measures, planning utility response to detection of contamination incidents, and identifying potential sensor locations. The decision-making process involves applying an optimization method and evaluating sensor placements. The process is informed by analyzing tradeoffs and comparing a series of designs to account for modeling and data uncertainties. This approach was applied to design the first EPA Water Security initiative pilot city (U.S. EPA 2005c).

Grayman et al. (2006) describe an interactive decision making framework that can help water utilities assess the strengths and weaknesses of sensor placement designs. This framework can be integrated with optimization strategies to help water utilities gain insight from optimized sensor placements. This is an important exercise because computational optimization methods do not generally tell the user why a design is optimal. Similarly, Isovitsch and VanBriesen (2007; 2008) describe an analysis technique that uses GIS to provide insight into the layout and sensitivity of sensor network designs.

<sup>2</sup> Mathematically, optimal solutions are guaranteed to be the same for sensor placement formulations that minimize impact or maximize reduction of impact. However, the near-optimal sensor placements generated by the method of Krause et al. (2008;2006) are not guaranteed to provide a near optimal minimization of impact. We have discussed this point with various members of the water community, and there is not a clear preference for one type of formulation over the other. Even so, a colleague has suggested a rationale for designing a sensor placement that minimizes impact: “If a contamination event occurs, the newspaper is going to print the number of people killed rather than the number of people saved by the contamination warning system.”



# Appendix B.

## Battle of the Water Sensor Networks

The “Battle of the Water Sensor Networks” (BWSN) (Ostfeld et al. 2008) of 2006 brought together 15 different small teams of researchers who had developed sensor placement capabilities. These teams generated sensor placements for two utility network models under a variety of threats. The first of these datasets was a small, imaginary network with roughly 100 nodes. The second, “Network 2,” was a disguised version of a real network, used with permission of the relevant utility, and consisting of roughly 12,000 nodes. The threat ensembles were sets of contamination incidents, each with different duration of injection, the number of injections per node, and whether or not simultaneous injections were to occur. Readers are referred to the paper itself for more detail.

Although not a perfect competition between methods (there was healthy debate over many aspects of the competition), the BWSN was a remarkable coordination effort, and it generated some meaningful comparison results. TEVA-SPOT’s GRASP solver was one of the entrants and its results will be placed into context here.

There were four sensor placement objectives considered in the BWSN:

- Z1: the expected (mean) time to detection
- Z2: the expected number of people affected by contamination
- Z3: the expected volume of contaminated water consumed
- Z4: the percentage of incidents detected by a sensor

The competition predated the introduction of side constraints into TEVA-SPOT, so the TEVA Research Team submitted solutions that minimize Z3, knowing that Z1, Z2, and Z3 are strongly correlated. For Network 2, placing 5 sensors in response to “Case A” (single injection sites, two hour duration of injection), TEVA-SPOT’s GRASP solver found the same sensor placement as the closest competitor, a greedy sensor placement algorithm implemented by Krause, et al. (2006). On the more challenging 20-sensor variant of this problem, for objectives Z1, Z2, and Z3, the solutions obtained by TEVA-SPOT’s GRASP solver were, respectively, 18%, 21%, and 36% better than Krause’s greedy algorithm.

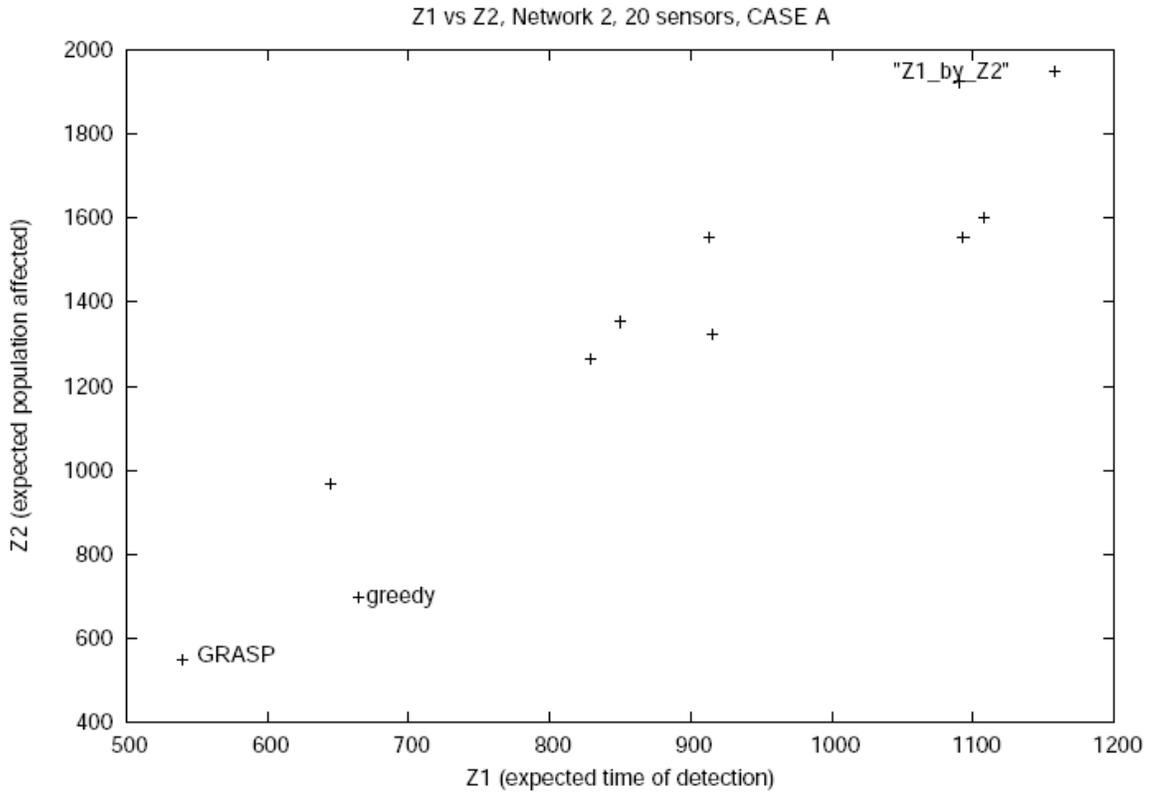
The competition admitted no winner, instead counting the number of “non-dominated solutions” provided by each team. A solution is non-dominated if there is no other solution that is superior in all four objectives simultaneously. The closest

thing to a winner of the BWSN was the entry of Krause et al. (2006), which had the largest number of non-dominated solutions. However, a further look at the data suggests that this non-dominated metric does not adequately capture the relative benefit of sensor placements.

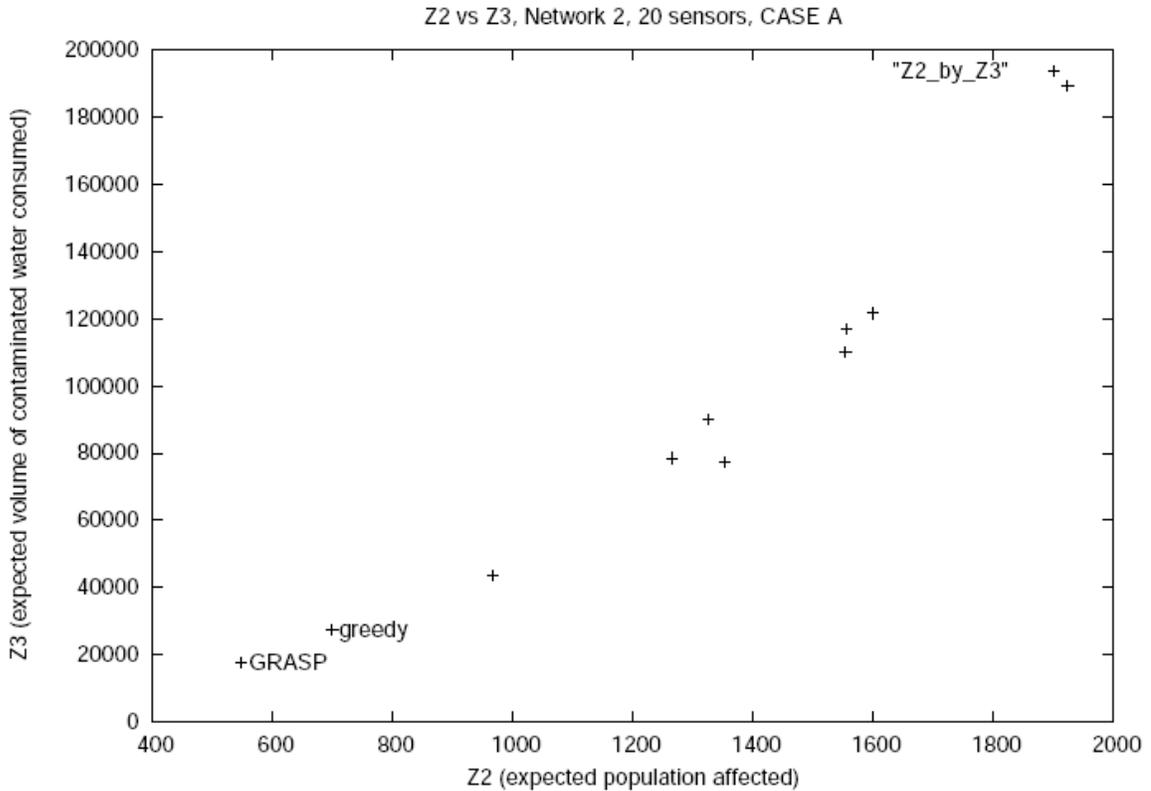
**Figures B-1, B-2, and B-3** show the raw data for Network 2, where 20 sensors are placed based on the assumptions of Case A. Since GRASP does not dominate in Z4 (greedy detects 3% more incidents), the greedy solution is non-dominated. However, the sensor placement computed by TEVA-SPOT is clearly preferable in terms of human costs and timeliness of detection. The network is so large that with only 20 sensors, there is little hope of detecting the large number of incidents that contaminate only a tiny portion of the network. Intuitively, injections near the edges of the network often do not move into large pipes to be dispersed more widely. Yet, Case A includes injections at all such nodes.

One important result of the BWSN is quantitative evidence that optimization has great value in placing sensors. Two competitors submitted designs that were not based on optimization techniques. Ghimire and Barkdoll (2006) use heuristics based on demand (without optimizing over any water quality simulation data), and provide a solution for the same threat ensemble described above (N2A20). This solution is respectively 101%, 251%, and 984% worse in Z1, Z2, and Z3 than the TEVA-SPOT solutions. Trachtman (2006) looked at pressure and flow patterns (again, without considering water quality simulations), and produced a solution for N2A20 that was, respectively, 69%, 183%, and 569% worse in Z1, Z2, and Z3 than the TEVA-SPOT solution.

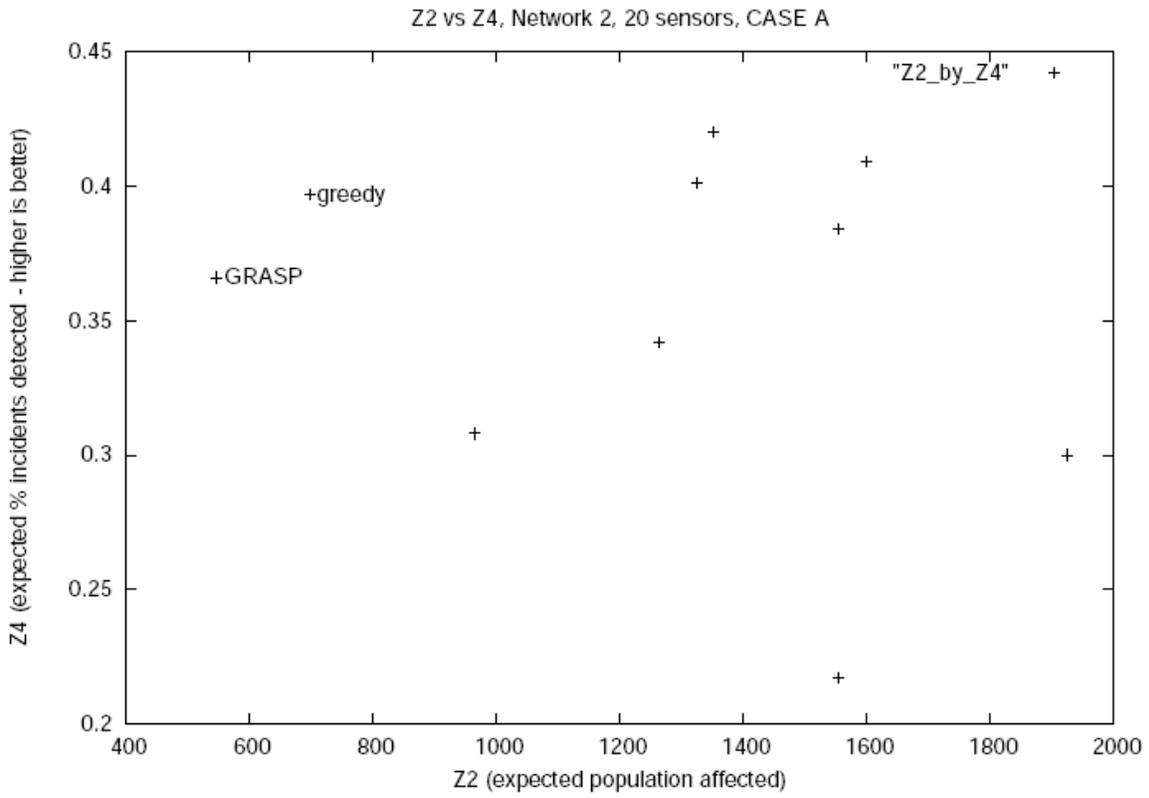
Perhaps indicating a culture clash in the water community, non-simulation-based solutions such as these met with a distinctively warm audience reception at the BWSN session at the Water Distribution Systems Analysis Symposium of 2006. They are, perhaps, more comforting to those distrustful of the hidden details underlying optimization methods. However, the potential consequences of foregoing water quality simulations before making sensor placement decisions were highlighted in ample detail by the BWSN. This competition demonstrates that water quality simulations and subsequent optimization should be a part of any real-world sensor placement application.



**Figure B-1.** Performance of sensor placement methods in terms of the Z1 and Z2 metrics. The GRASP algorithm performs better than the competitors in both objectives.



**Figure B-2.** Performance of sensor placement methods in terms of the Z2 and Z3 metrics. The GRASP algorithm performs better than the competitors in both objectives.



**Figure B-3.** Performance of sensor placement methods in terms of the Z2 and Z4 metrics. The GRASP algorithm performs better in the Z2 metric but not in the Z4 metric.



# Appendix C.

## Quality Assurance

EPA's quality systems cover the collection, evaluation, and use of environmental data by and for the Agency, and the design, construction, and operation of environmental technology by the Agency. The purpose of EPA's quality systems is to support scientific data integrity, reduce or justify resource expenditures, properly evaluate of internal and external activities, support reliable and defensible decisions by the Agency, and reduce burden on partnering organizations.

All research presented in this report performed by the authors was completed under approved EPA and DOE quality practices adapted from the *Advanced Simulation and Computing (ASC) Software Quality Plan* and *EPA guidance for Quality Assurance Project Plans*. The *ASC Software Quality Plan* was generated to conform with the SNL corporate and DOE QC-1 revision 9 standards.

The quality assurance (QA) practices followed under this research included:

- Project Management
- Computational Modeling and Algorithm Development
- Software Engineering
- Data Generation and Acquisition
- Model and Software Verification
- Training

Project management is the systematic approach for balancing the project work to be done, resources required, methods used, procedures to be followed, schedules to be met, and the way that the project is organized. The project management QA practices included: performing a risk-based assessment to determine level of formality and applicable practices; identifying stakeholders and other requirements sources; gathering and managing stakeholders' expectations and requirements; deriving, negotiating, managing, and tracking requirements; identifying and analyzing project risk events; defining, monitoring, and implementing the risk response; creating and managing the project plan; and tracking project performance versus project plan and implementing needed corrective actions.

Modeling and algorithm development are often closely related activities; modeling is the process of mathematically formulating a problem, while algorithm development is the process of finding a method to solve the problem computationally. These activities can be distinguished from software engineering efforts, which are more specifically focused on ensuring that software generated has high quality itself. The modeling and algorithm development QA practices included: documenting designs for models and algorithms; conducting peer reviews of modeling assumptions and algorithmic formulations; documenting preliminary software implementation; documenting sources of uncertainty in modeling and algorithmic methods; and completing peer-review of modeling and algorithmic outputs.

Software engineering is a systematic approach to the specification, design, development, test, operation, support, and retirement of software. The modeling and algorithm development QA practices included: communicating and reviewing software design; creating required software and product documentation; identifying and tracking third party software products and follow applicable agreements; identifying, accepting ownership, and managing assimilation of other software products; performing version control of identified software product artifacts; recording and tracking issues associated with the software product; ensuring backup and disaster recovery of software product artifacts; planning and generating the release package; and certifying that the software product (code and its related artifacts) was ready for release and distribution.

Input data for model development and application efforts are typically collected outside of the modeling effort or generated by other models or processing software. These data need to be properly assessed to verify that a model characterized by these data would yield predictions with an acceptable level of uncertainty. The data generation and acquisition QA practices included: documenting objectives and methods of model calibration activities; documenting sources of input data used for calibration; identifying requirements for non-direct data and data acquisition; developing processes for managing data; and documenting hardware and software used to process data.

The purpose of software verification is to ensure (1) that specifications are adequate with respect to intended use and (2) that specifications are accurately, correctly, and completely implemented. Software verification also attempts to ensure product characteristics necessary for safe and proper use are addressed. Software verification occurs throughout the entire product lifecycle. The software verification QA practices included: developing and maintaining a software verification plan; conducting tests to demonstrate that acceptance criteria are met and to ensure that previously tested capabilities continue to perform as expected; and conducting independent technical reviews to evaluate adequacy with respect to requirements.

The goal of training practices is to enhance the skills and motivation of a staff that is already highly trained and educated in the areas of mathematical modeling, scientific software development, algorithms, and/or computer science. The purpose of training is to develop the skills and knowledge of individuals and teams so they can fulfill their process and technical roles and responsibilities. The training QA practices included: determining project team training needed to fulfill assigned roles and responsibilities; and tracking training undertaken by project team.



# References

- Al-Zahrani, M. A., and Moeid, K. (2001). "Locating optimum water quality monitoring stations in water distribution system." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Al-Zahrani, M. A., and Moied, K. (2003). "Optimizing water quality monitoring stations using genetic algorithms." *Arabian Journal for Science and Engineering*, 28(1B), 57–75.
- AMSA. (2003). *The vulnerability self assessment tool (VSAT) for water and wastewater utilities*, The Association of Metropolitan Sewerage Agencies. <<http://www.vsatusers.net/index.html>>.
- Andelman, J. B. (1985a). "Human exposures to volatile halogenated organic-chemicals in indoor and outdoor air." *Environmental Health Perspectives*, 62(Oct), 313–318.
- Andelman, J. B. (1985b). "Inhalation exposure in the home to volatile organic contaminants of drinking-water." *Science of the Total Environment*, 47(DEC), 443–460.
- Aral, M. M., Guan, J., and Maslia, M. L. (2008). "A multi-objective optimization algorithm for sensor placement in water distribution systems." *Proc., World Environmental and Water Resources Congress*, ASCE, Reston, VA.
- ASCE. (2004). *Interim voluntary guidelines for designing an online contaminant monitoring system*, American Society of Civil Engineers, Reston, VA.
- ASME-ITI. (2005). *RAMCAP executive summary*, American Society of Mechanical Engineers Innovative Technologies Institute, New York, NY. <<http://files.asme.org/ASMEITI/RAMCAP/12604.pdf>>.
- Agency for Toxic Substances and Disease Registry (ATSDR). (2001). *Managing hazardous material incidents (MHMI), Volume 3*, U.S. Department of Health and Human Services, Public Health Service, Atlanta, GA. <<http://www.atsdr.cdc.gov/mhmi/mmg.html>>.
- Avella, P., Sassano, A., and Vasil'ev, I. (2007). "Computational study of large-scale p-median problems." *Mathematical Programming*, 109(1), 89–114.
- AWWA. (2005). *Contamination warning systems for water: an approach for providing actionable information to decision-makers*, American Water Works Association, Denver, CO.
- AwwaRF. (2003). *Actual and threatened security events at water utilities*, Project 2810, American Water Works Association Research Foundation, Denver, CO.
- AwwaRF, and SNL. (2002). *Risk Assessment Methodology for Water Utilities (RAM-W)*, American Water Works Association Research Foundation (AwwaRF) and Sandia National Laboratories (SNL), Denver, CO and Albuquerque, NM.
- Bahadur, R., Samuels, W. B., Grayman, W., Amstutz, D., and Pickus, J. (2003). "PipelineNet: A model for monitoring introduced contaminants in a distribution system." *Proc., World Water and Environmental Resources Congress 2003 and Related Symposia*, ASCE, Reston, VA.
- Barahona, F., and Anbil, R. (2000). "The volume algorithm: producing primal solutions with a subgradient method." *Mathematical Programming*, 87(3), 385–399.
- Barahona, F., and Chudak, F. (2005). "Near-optimal solutions to large-scale facility location problems." *Discrete Optimization*, 2, 35–50.
- Berry, J., Boman, E., Phillips, C. A., and Riesen, L. (2008a). "Low-memory lagrangian relaxation methods for sensor placement in municipal water networks." *Proc., World Environmental and Water Resources Congress*, ASCE, Reston, VA.
- Berry, J., Carr, R., Hart, W. E., and Phillips, C. A. (2007). "Scalable water network sensor placement via aggregation." *Proc., World Environmental and Water Resources Congress*, ASCE, Reston, VA.
- Berry, J., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., and Watson, J-P. (2006a). "On the placement of imperfect sensors in municipal water networks." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Berry, J., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., and Watson, J-P. (2009). "Designing contamination warning systems for municipal water networks using imperfect sensors." *Journal of Water Resources Planning and Management*, 135(4), 253–263.

- Berry, J., Fleischer, L., Hart, W. E., and Phillips, C. A. (2003). "Sensor placement in municipal water networks." *Proc., World Water and Environmental Resources Congress 2003 and Related Symposia*, ASCE, Reston, VA.
- Berry, J., Hart, W. E., Phillips, C. A., and Uber, J. (2004). "A general integer-programming-based framework for sensor placement in municipal water networks." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G., and Walski, T. M. (2005a). "Water quality sensor placement in water networks with budget constraints." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G., and Watson, J-P. (2005b). "Validation and assessment of integer programming sensor placement models." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- 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., Boman, E., Riesen, L. A., Hart, W. E., Phillips, C. A., and Watson, J.-P. (2008b). *User's manual: TEVA-SPOT toolkit 2.2*, EPA-600-R-08-041, U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.
- Berry, J. W., Fleischer, L., Hart, W. E., Phillips, C. A., and Watson, J-P. (2005c). "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. (2005d). "Scalability of integer programming computations for sensor placement in water networks." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- BLS and U.S. Census Bureau. (2005). *American time use survey: user's guide 2003–2004*, U.S. Department of Labor, Bureau of Labor Statistics and U.S. Department of Commerce, U.S. Census Bureau, Washington, D.C.
- Bristow, E. C., and Brumbelow, K. (2006). "Delay between sensing and response in water contamination events." *Journal of Infrastructure Systems*, 12(2), 87–95.
- Brown, H. S., Bishop, D. R., and Rowan, C. A. (1984). "The role of skin absorption as a route of exposure for volatile organic-compounds (VOCs) in drinking-water." *American Journal of Public Health*, 74(5), 479–484.
- Public Health Security and Bioterrorism Preparedness and Response Act of 2002. (2002). PL 107-188. <<http://www.fda.gov/oc/bioterrorism/PL107-188.pdf>>.
- Burrows, W. D., and Renner, S. E. (1999). "Biological warfare agents as threats to potable water." *Environmental Health Perspectives*, 107(12), 975–984.
- Cameron, C. (2002). *Feds arrest Al Qaeda suspects with plans to poison water supplies*, FoxNews.com, <<http://www.foxnews.com/story/0,2933,59055,00.html>>.
- 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.
- Carr, R. D., Greenberg, H. J., Hart, W. E., and Phillips, C. A. (2004). "Addressing modeling uncertainties in sensor placement for community water systems." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Chastain, J. R., Jr. (2006). "Methodology for locating monitoring stations to detect contamination in potable water distribution systems." *Journal of Infrastructure Systems*, 12(4), 252–259.
- Chastain Jr., J. R. (2004). "A heuristic methodology for locating monitoring stations to detect contamination events in potable water distribution systems." Dissertation, University of South Florida, Tampa, FL.
- Chick, S. E., Koopman, J. S., Soorapanth, S., and Brown, M. E. (2001). "Infection transmission system models for microbial risk assessment." *Science of the Total Environment*, 274(1–3), 197–207.

- Clark, R. M., and Deininger, R. A. (2001). "Minimizing the vulnerability of water supplies to natural and terrorist threats." *Proc., IMTech (Information Management and Technology) Conference*, AWWA, Denver, CO."
- Cook, J., Roehl, E., Daamen, R., Carlson, K., and Byer, D. (2005). "Decision support system for water distribution system monitoring for homeland security." *Proc., AWWA Water Security Congress*, AWWA, Denver, CO.
- Covello, V. T., and Merkhoher, M. W. (1993). *Risk assessment methods: approaches for assessing health and environmental risks*, Plenum Publishing Corporation, New York, NY.
- Cozzolino, L., Mucherino, C., Pianese, D., and Pirozzi, F. (2006). "Positioning, within water distribution networks, of monitoring stations aiming at an early detection of intentional contamination." *Civil Engineering and Environmental Systems*, 23(3), 161–174.
- Davis, M. J., and Janke, R. (2008). "Importance of exposure model in estimating impacts when a water distribution system is contaminated." *Journal of Water Resources Planning and Management*, 134(5), 449–456.
- Davis, M. J., and Janke, R. (2009). "Development of a probabilistic timing model for the ingestion of tap water." *Journal of Water Resources Planning and Management*, 135(5), 397–405.
- de Castro, J. M. (1988). "A microregulatory analysis of spontaneous fluid intake by humans - evidence that the amount of liquid ingested and its timing is mainly governed by feeding." *Physiology & Behavior*, 43(6), 705–714.
- de Castro, J. M. (1992). "Age-related-changes in natural spontaneous fluid ingestion and thirst in humans." *Journals of Gerontology*, 47(5), P321–P330.
- Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S.-T., and Savic, D. (2006). "An efficient algorithm for sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Eliades, D., and Polycarpou, M. (2006). "Iterative deepening of pareto solutions in water sensor networks." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Engell, E. (1988). "Interdependency of food and water intake in humans." *Appetite*, 10(2), 133–141.
- Ghimire, S. R., and Barkdoll, B. D. (2006). "A heuristic method for water quality sensor location in a municipal water distribution system: mass-released based approach." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Grayman, W. M., Ostfeld, A., and Salomons, E. (2006). "Locating monitors in water distribution systems: red team-blue team exercise." *Journal of Water Resources Planning and Management*, 132(4), 300–304.
- Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. (2006). "Optimization model and algorithms for design of water sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Gueli, R. (2006). "Predator - prey model for discrete sensor placement." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Haas, C. N., Rose, J. B., and Gerba, C. P. (1999). *Quantitative microbial risk assessment*, John Wiley & Sons, Inc., New York, NY.
- Hall, J., Zaffiro, A. D., Marx, R. B., Kefauver, P. C., Krishnan, E. R., and Herrmann, J. G. (2007). "On-line water quality parameters as indicators of distribution system contamination." *Journal American Water Works Association*, 99(1), 66–77.
- Hart, D., McKenna, S. A., Klise, K., Cruz, V., and Wilson, M. (2007). "CANARY: a water quality event detection algorithm development tool." *Proc., World Environmental and Water Resources Congress*, ASCE, Reston, VA.
- Hart, W. E., Berry, J. W., Boman, E., Phillips, C. A., Riesen, L. A., and Watson, J-P. (2008a). "Limited-memory techniques for sensor placement in water distribution networks." *Learning and Intelligent Optimization. Second International Conference, LION 2007 II. Selected Papers*, 5313, 125–137.
- Hart, W. E., Berry, J. W., Boman, E. G., Murray, R., Phillips, C. A., Riesen, L. A., and Watson, J-P. (2008b). "The TEVA-SPOT toolkit for drinking water contaminant warning system design." *Proc., World Environmental & Water Resources Congress*, ASCE, Reston, VA.

- Hart, W. E., and Murray, R. (in review). "A review of sensor placement strategies for contamination warning systems." *Journal of Water Resources Planning and Management*.
- Henneberger, M. (2002). "A nation challenged: suspects; 4 arrested in plot against U.S. Embassy in Rome." *The New York Times*, The New York Times Company, New York, NY. <<http://query.nytimes.com/gst/fullpage.html?res=9901E0DD153EF932A15751C0A9649C8B63>>.
- Howard-Reed, C., Corsi, R. L., and Moya, J. (1999). "Mass transfer of volatile organic compounds from drinking water to indoor air: the role of residential dishwashers." *Environmental Science & Technology*, 33(13), 2266–2272.
- Huang, J. J., McBean, E. A., and James, W. (2006). "Multi-objective optimization for monitoring sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Huang, Y.-d., Zhang, T.-q., and Song, J.-r. (2007). "Multi-objective optimization model of sensor placement in water distribution systems considering reliability." *Chinese Journal of Sensors and Actuators*, 20(8), 1888–1893.
- Huber, P. J. (2004). *Robust statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.
- IBWA. (2002). Excerpts from February National Quorum Findings, International Bottled Water Association, <<http://www.bottledwater.org/public/bwsurvey.htm>>.
- Isovitich, S. L., and VanBriesen, J. M. (2007). "Spatial analysis of optimized sensor locations using GIS." *Proc., World Environmental and Water Resources Congress*, ASCE, Reston, VA.
- Isovitich, 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.
- U.S. EPA. (2000). *Estimated per capita water ingestion in the United States: based on data collected by the United States Department of Agriculture's 1994–96 continuing survey of food intakes by individuals*, EPA-822-R-00-008, U. S. Environmental Protection Agency, Office of Water, Washington, D.C.
- Janke, R., Murray, R., Uber, J., and Taxon, T. (2006). "Comparison of physical sampling and real-time monitoring strategies for designing a contamination warning system in a drinking water distribution system." *Journal of Water Resources Planning and Management*, 132(4), 310–313.
- Kessler, A., and Ostfeld, A. (1998a). "Detecting accidental contaminants in municipal water networks: application." *Proc., 25th Annual Conference on Water Resources Planning and Management*, ASCE, Reston, VA, 272–278.
- Kessler, A., Ostfeld, A., and Sinai, G. (1998b). "Detecting accidental contaminations in municipal water networks." *Journal of Water Resources Planning and Management*, 124(4), 192–198.
- Kızileniş, G. (2006). "Optimal sensor locations in water distribution networks." Masters thesis, Sabanci University, Istanbul, Turkey.
- Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., and Faloutsos, C. (2008). "Efficient sensor placement optimization for securing large water distribution networks." *Journal of Water Resources Planning and Management*, 134(6), 516–526.
- Krause, A., Leskovec, J., Isovitich, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., and Fischbeck, P. (2006). "Optimizing sensor placements in water distribution systems using submodular function maximization." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- 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.
- Kumar, A., Kansal, M. L., and Arora, G. (1999). "Detecting accidental contaminations in municipal water networks - discussion." *Journal of Water Resources Planning and Management*, 125(5), 308–309.
- Kunze, D. R. (1997). "Assessing utility threats." *Security Management*, 41(2), 75–77.
- 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.
- Lee, B. H., Deininger, R. A., and Clark, R. M. (1991). "Locating monitoring stations in water distribution-systems." *Journal American Water Works Association*, 83(7), 60–66.

- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. (2007). "Cost-effective outbreak detection in networks." *Proc., The Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007)*, ACM (Association for Computing Machinery), New York, NY, 420–429.
- Ma, Y. S., Bertone-Johnson, E. R., Stanek, E. J., Reed, G. W., Herbert, J. R., Cohen, N. L., Olendzki, B. C., Rosal, M. C., Merriam, P. A., and Ockene, I. S. (2005). "Eating patterns in a free-living healthy U.S. adult population." *Ecology of Food and Nutrition*, 44(3), 255–255.
- McKenna, S. A., Klise, K. A., and Wilson, M. P. (2006). "Testing water quality change detection algorithms." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- McKenna, S. A., Wilson, M., and Klise, K. A. (2008). "Detecting changes in water quality data." *Journal American Water Works Association*, 100(1), 74–85.
- McKone, T. E. (1989). "Household exposure models." *Toxicology Letters*, 49(2–3), 321–339.
- Mirchandani, P. B., and Francis, R. L. (1990). *Discrete location theory*, John Wiley & Sons, Inc., Hoboken, NJ.
- Morley, K., Janke, R., Murray, R., and Fox, K. (2007). "Drinking water contamination-warning systems: water utilities driving water security research." *Journal American Water Works Association*, 99(6), 40–46.
- Murray, R. (2004). "Water and Homeland Security: An Introduction." *Journal of Contemporary Water Research and Education*, 129, 1–2.
- Murray, R., Baranowski, T., Hart, W. E., and Janke, R. (2008a). "Risk reduction and sensor network design." *Proc., Water Distribution Systems Analysis 2008*, ASCE, Reston, VA.
- Murray, R., Hart, W., and Berry, J. (2006a). "Sensor network design for contamination warning systems: tool and applications." *Proc., AWWA Water Security Conference*, AWWA, Denver, CO.
- Murray, R., Hart, W. E., Phillips, C. A., Berry, J., Boman, E. G., Carr, R. D., Riesen, L. A., Watson, J-P., Haxton, T., Herrmann, J. G., Janke, R., Gray, G., Taxon, T., Uber, J. G., and Morley, K. M. (2009). "US Environmental Protection Agency uses operations research to reduce contamination risks in drinking water." *Interfaces*, 39(1), 57–68.
- Murray, R., Janke, R., Hart, W. E., Berry, J. W., Taxon, T., and Uber, J. (2008b). "Sensor network design of contamination warning systems: a decision framework." *Journal American Water Works Association*, 100(11), 97–109.
- Murray, R., Uber, J., and Janke, R. (2006b). "Model for estimating acute health impacts from consumption of contaminated drinking water." *Journal of Water Resources Planning and Management*, 132(4), 293–299.
- NRWA. (2003). *Security and emergency management system (SEMS)*, National Rural Water Association, Duncan, OK. <<http://www.vulnerabilityassessment.org/>>.
- Ostfeld, A. (2006). "Enhancing water-distribution system security through modeling." *Journal of Water Resources Planning and Management*, 132(4), 209–210.
- Ostfeld, A., and Kessler, A. (2001). "Protecting urban water distribution systems against accidental hazards intrusions." *Proc., IWA Second Conference*, IWA (International Water Association), London, UK.
- Ostfeld, A., and Salomons, E. (2003). "An early warning detection system (EWDS) for drinking water distribution systems security." *Proc., World Water & Environmental Resources Congress 2003 and Related Symposia*, ASCE, Reston, VA.
- 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. (2005a). "Optimal early warning monitoring system layout for water networks security: inclusion of sensors sensitivities and response delays." *Civil Engineering and Environmental Systems*, 22(3), 151–169.
- Ostfeld, A., and Salomons, E. (2005b). "Securing water distribution systems using online contamination monitoring." *Journal of Water Resources Planning and Management*, 131(5), 402–405.

- Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J-P., Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S. T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J. H., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y., and Walski, T. (2008). "The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms." *Journal of Water Resources Planning and Management*, 134(6), 556–568.
- Perelman, L., and Ostfeld, A. (2008). "Water distribution system aggregation for water quality analysis." *Journal of Water Resources Planning and Management*, 134(3), 303–309.
- Phillips, P. A., Rolls, B. J., Ledingham, J. G. G., and Morton, J. J. (1984). "Body fluid changes, thirst and drinking in man during free access to water." *Physiology & Behavior*, 33(3), 357–363.
- Pickus, J., Bahadur, R., and Samuels, W. B. (2005). "Integrating the ArcGIS water distribution data model into PipelineNet." *Proc., ESRI International User Conference*, ESRI, Redlands, CA.
- Preis, A., and Ostfeld, A. (2006a). "Multiobjective sensor design for water distribution systems security." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Preis, A., and Ostfeld, A. (2006b). "Optimal sensors layout for contamination source identification in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Preis, A., and Ostfeld, A. (2007). "Efficient contamination events sampling for sensors layout design." *Proc., World Environmental and Water Resources Congress*, ASCE, Reston, VA.
- Preis, A., and Ostfeld, A. (2008). "Multiobjective contaminant sensor network design for water distribution systems." *Journal of Water Resources Planning and Management*, 134(4), 366–377.
- 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., Piller, O., and Uber, J. G. (2005). "A sensor location model to detect contaminations in water distribution networks." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Resende, M. G. C., and Werneck, R. F. (2004). "A hybrid heuristic for the p-median problem" *Journal of Heuristics*, 10(1), 59-88.
- Rico-Ramirez, V., Frausto-Hernandez, S., Diwekar, U. M., and Hernandez-Castro, S. (2007). "Water networks security: a two-stage mixed-integer stochastic program for sensor placement under uncertainty." *Computers & Chemical Engineering*, 31(5–6), 565–573.
- Romero-Gomez, P., Choi, C. Y., Lansey, K. E., Preis, A., and Ostfeld, A. (2008). "Sensor network design with improved water quality models at cross junctions." *Proc., Water Distribution Systems Analysis 2008*, ASCE, Reston, VA.
- Rossman, L. A. (1999). "The EPANET programmer's toolkit for analysis of water distribution systems." *Proc., 26th Annual Water Resources Planning and Management Conference*, ASCE, Reston, VA.
- Rossman, L. A. (2000). *EPANET 2: users manual*, EPA-600-R-00-057, U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Cincinnati, OH. <<http://www.epa.gov/nrmrl/wswrd/dw/epanet/EN2manual.PDF>>.
- Shang, F., Uber, J., and Rossman, L. (2007). *EPANET Multi-Species Extension user's manual*, EPA-600-S-07-021, U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.
- Shang, F., Uber, J. G., and Rossman, L. A. (2008). "Modeling reaction and transport of multiple species in water distribution systems." *Environmental Science & Technology*, 42(3), 808–814.
- 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.
- Shehata, A. T. (1985). "A multi-route exposure assessment of chemically contaminated drinking water." *Toxicology and Industrial Health*, 1(4), 277–298.
- Skadsen, J., Janke, R., Grayman, W., Samuels, W., TenBroek, M., Steglitz, B., and Bahl, S. (2008). "Distribution system on-line monitoring for detecting contamination and water quality changes." *Journal American Water Works Association*, 100(7), 81–94.

Staudinger, T. J., England, E. C., and Bleckmann, C. (2006). "Comparative analysis of water vulnerability assessment methodologies." *Journal of Infrastructure Systems*, 12(2), 96–106.

Trachtman, G. (2006). "A "strawman" common sense approach for water quality sensor site selection." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.

U.S. DHS and U.S. EPA. (2007). *Water: critical infrastructure and key resources sector-specific plan as input to the national infrastructure protection plan*, U.S. Department of Homeland Security and U.S. Environmental Protection Agency, Washington, D.C. <[http://www.dhs.gov/xlibrary/assets/Water\\_SSP\\_5\\_21\\_07.pdf](http://www.dhs.gov/xlibrary/assets/Water_SSP_5_21_07.pdf)>.

U.S. EPA. (1997). *Exposure factors handbook*, EPA-600-P-95-002Fa,b,c, U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Washington, D.C. <<http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=12464>>.

U.S. EPA. (2002a). *Permeation and Leaching*, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C. <[http://www.epa.gov/OGWDW/disinfection/tcr/pdfs/whitepaper\\_tcr\\_permation-leaching.pdf](http://www.epa.gov/OGWDW/disinfection/tcr/pdfs/whitepaper_tcr_permation-leaching.pdf)>.

U.S. EPA. (2002b). *Vulnerability assessment fact sheet*, EPA-816-F-02-025, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C. <[http://www.epa.gov/safewater/watersecurity/pubs/va\\_fact\\_sheet\\_12-19.pdf](http://www.epa.gov/safewater/watersecurity/pubs/va_fact_sheet_12-19.pdf)>.

U.S. EPA. (2003). *EPA needs to assess the quality of vulnerability assessments related to the security of the nation's water supply*, 2003-M-00013, U. S. Environmental Protection Agency, Office of Inspector General, Washington, D.C. <<http://www.epa.gov/oigearth/reports/2003/Report2003M000013.pdf>>.

U.S. EPA. (2004). *Response protocol toolbox: planning for and responding to drinking water contamination threats and incidents*, EPA-817-D-04-001, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C. <[http://www.epa.gov/safewater/watersecurity/pubs/rptb\\_response\\_guidelines.pdf](http://www.epa.gov/safewater/watersecurity/pubs/rptb_response_guidelines.pdf)>.

U.S. EPA. (2005a). *Technologies and techniques for early warning systems to monitor and evaluate drinking water quality: a state-of-the-art review*, EPA-600-R-05-156, U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.

U.S. EPA. (2005b). *WaterSentinel online water quality monitoring as an indicator of drinking water contamination*, EPA-817-D-05-002, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C.

U.S. EPA. (2005c). *WaterSentinel system architecture*, EPA-817-D-05-003, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C.

U.S. EPA. (2006a). *Active and effective water security programs: a summary report of the National Drinking Water Advisory Council recommendations on water quality*, EPA-817-K-06-001, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

U.S. EPA. (2006b). *Initial distribution system evaluation guidance manual for the final stage 2 disinfectants and disinfection byproducts rule*, EPA-815-B-06-002, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C.

U.S. EPA. (2008). *Water security initiative Cincinnati pilot post-implementation system status: covering the pilot period: December 2005 through December 2007*, EPA-817-R-08-004, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water, Washington, D.C. <[http://www.epa.gov/safewater/watersecurity/pubs/rpt\\_post\\_imp\\_system\\_status\\_wsi.pdf](http://www.epa.gov/safewater/watersecurity/pubs/rpt_post_imp_system_status_wsi.pdf)>.

U.S. EPA. (2009). *Tutorial threat ensemble vulnerability analysis – sensor placement optimization tool (TEVA-SPOT) graphical user interface, Version 2.2.0 Beta*, EPA-600-R-08-147, U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.

U.S. GAO. (2003). *Drinking water: experts' views on how future federal funding can best be spent to improve security*, GAO-04-29, U.S. General Accounting Office, Washington, D.C. <<http://www.gao.gov/new.items/d0429.pdf>>.

USGS (2004). *Estimated Use of Water in the United States in 2000*, U.S. Geological Survey, Reston VA. <<http://pubs.usgs.gov/circ/2004/circ1268/pdf/circular1268.pdf>>.

- Uber, J., Janke, R., Murray, R., and Meyer, P. (2004). "Greedy heuristic methods for locating water quality sensors in distribution systems." *Proc., The 2004 World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Walski, T. M., Daviau, J.-L., and Coran, S. (2004). "Effect of skeletonization on transient analysis results." *Proc., The 2004 World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Watson, J-P., Greenberg, H. J., and Hart, W. E. (2004). "A multiple-objective analysis of sensor placement optimization in water networks." *Proc., The 2004 World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Watson, J-P., Hart, W. E., and Berry, J. (2005). "Scalable high-performance heuristic for sensor placement in water distribution networks." *Proc., World Water and Environmental Resources Congress*, ASCE, Reston, VA.
- Watson, J-P., Hart, W. E., and Murray, R. (2006). "Formulation and optimization of robust sensor placement problems for contaminant warning systems." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Watson, J-P., Murray, R., and Hart, W. E. (2009). "Formulation and optimization of robust sensor placement problems for drinking water contamination warning systems." *Journal of Infrastructure Systems*.
- Watts. (2009). *Stop backflow news: case histories and solutions*, F-SBN-0035, Watts Water Technologies Company, North Andover, MA. <<http://www.watts.com/pdf/F-sbn.pdf>>.
- Weisel, C. P., and Jo, W. K. (1996). "Ingestion, inhalation, and dermal exposures to chloroform and trichloroethene from tap water." *Environmental Health Perspectives*, 104(1), 48–51.
- Wheeler, J. G., Sethi, D., Cowden, J. M., Wall, P. G., Rodrigues, L. C., Tompkins, D. S., Hudson, M. J., Roderick, P. J., and Infect Intestinal Dis Study, E. (1999). "Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance." *British Medical Journal*, 318(7190), 1046–1050.
- Wu, X-G., Zhang, T-Q., and Huang, Y-D. (2008). "Optimal algorithm for determining locations of water quality sensors in water supply networks under multi-objective constraints." *Journal of Hydraulic Engineering (China)*, 39(4), 433–439.
- Wu, Z. Y., and Walski, T. (2006). "Multi-objective optimization of sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.
- Xu, J. H., 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.



# RESEARCH CENTER



PRESORTED STANDARD  
POSTAGE & FEES PAID  
EPA  
PERMIT NO. G-35

Office of Research and Development  
National Homeland Security Research Center  
Cincinnati, OH 45268

Official Business  
Penalty for Private Use  
\$300



**Recycled/Recyclable**  
Printed with vegetable-based ink on  
paper that contains a minimum of  
50% post-consumer fiber content  
processed chlorine free