An Analysis of Multiple Contaminant Warning System Design Objectives for Sensor Placement Optimization in Water Distribution Networks

Jean-Paul Watson* William E. Hart† Harvey J. Greenberg‡

Abstract

Recent events have significantly increased the level of interest in using sensors to detect the presence of contaminants in municipal water distribution systems. A key deployment issue is identifying where a limited number of sensors should be placed in order to maximize the level of protection afforded. Researchers have proposed several algorithms for generating such placements, each optimizing with respect to a different design objective. The use of disparate objectives raises several questions, in particular (1) What is the relationship between optimal sensor placements for different design objectives? and (2) Is there any risk in focusing on specific design objectives? To answer these questions, we solve the sensor placement problem using a mixed-integer programming formulation of the well-known p-median problem from facility location theory, which can express a broad range of design objectives. Using three large test networks, we show that optimal solutions with respect to one design objective are often highly sub-optimal with respect to other design objectives. In some instances, however, it is possible to construct solutions that are simultaneously near-optimal with respect to a range of design objectives. For both algorithm developers and decision-makers, the implication is that development of robust sensor placements requires careful and simultaneous consideration of multiple, disparate design objectives.

*Discrete Algorithms and Math Department, Sandia National Laboratories, P.O. Box 5800, MS 1110, Albuquerque, NM 87185-1110; Phone: 505-845-8887; E-Mail: jwatson@sandia.gov.

†Discrete Algorithms and Math Department, Sandia National Laboratories, P.O. Box 5800, MS 1110, Albuquerque, NM, 87185-1110; Phone: 505-844-2217; E-Mail: wehart@sandia.gov.

‡Mathematics Department, University of Colorado – Denver, P.O. Box 173364, Denver, CO 80217; Phone: 303-556-8464; E-Mail: Harvey.Greenberg@cudenver.edu.

1 Introduction

Recent events have highlighted the need to adequately protect drinking water supplies in the United States and around the world. To mitigate this risk, there is growing interest in the design and deployment of contaminant warning systems (CWSs) for water distribution networks, based on real-time sensors that provide continual water quality monitoring. In the short to moderate term, complete protection of distribution networks is unrealistic due to budgetary constraints. Consequently, optimization is required to maximize the level of protection afforded by a limited number of sensors.

The efficacy of a sensor placement can be quantified using a variety of measures. To date, researchers have developed algorithms for optimizing sensor placements with respect to a number of design objectives, including the proportion of population exposed prior to detection (Berry et al., 2005a), the volume of contaminated water consumed prior to detection (Kessler et al., 1998), and the time to detection (Kumar et al., 1998). An implicit assumption underlying these efforts is that the selected design objective is the "best" objective for sensor placement; in some instances researchers have made this argument very explicit, e.g., see (Kumar et al., 1998). Alternatively, it may be the case that no single best design objective exists. The variety of objectives introduced by researchers to date supports this view, as there are valid arguments for the importance of all such objectives. Further, existing objectives are not obviously redundant, e.g., the number of failed detections and the proportion of population exposed.

In this paper, we explore the trade-offs between different design objectives for sensor placement optimization in water distribution networks. Specifically, we pose and answer the following two research questions: (1) What is the relationship between optimal sensor placements for different design objectives? and (2) Is there any risk in focusing on specific design objectives? Our analysis considers the following six design objectives: population exposed, time to detection, volume of contaminated water consumed, mass of contaminant consumed, number of failed detections, and extent of contamination. For each of these objectives, the corresponding sensor placement optimization problem can be cast as the well-known p-median problem from facility location theory. Using mixed-integer programs (MIPs) and commercial MIP solvers, we identify optimal solutions for all design objectives, across a range of sensor budgets, on each of

three large test networks.

Our analysis of the resulting solutions indicates that optimal solutions with respect to one design objective are often highly sub-optimal with respect to complementary design objectives. In other words, there is potentially significant risk associated with focusing *a priori* on specific design objectives. Overall, our results reinforce the view that multiple objectives should at least be considered during the design of sensor placements for CWSs, as there are significant trade-offs that should be exposed to decision-makers during the design process. Fortunately, there is evidence that this risk can be mitigated in some circumstances; by sacrificing optimality in some design objectives, we demonstrate that it is possible to develop solutions that are more robust (i.e., higher-quality) with respect to secondary design objectives.

Finally, we analyze optimal sensor performance for each of the design objectives independently. Most of the design objectives we consider have previously been examined only in the context of small (e.g., on the order of 100 junction) test networks, or on less accurate sensor placement formulations than the p-median problem. Consequently, our results for individual design objectives provide a unique case-study contribution in that they quantify the benefit of various sensor budgets over the unprotected baseline case for CWS design, for large-scale distribution networks.

The remainder of this paper is organized as follows. In Section 2, we document the p-median formulation of the sensor placement problem and detail the computation of the various design objectives we consider in our analysis. In Section 3, we first describe our test networks, contamination scenarios, and aspects of our experimental methodology (Section 3.1). The behavior of individual design objectives on our test networks is then subsequently analyzed (Section 3.2). The remainder of the section addresses multiple-objective analysis, focusing on the relationship between different design objectives (Section 3.3). The implications of our analysis, including a discussion of the extent to which it may be possible to generate solutions that simultaneously yield high-quality solutions with respect to a range of design objectives, are detailed in Section 4. We conclude with a summary of our contributions in Section 5.

2 Problem, Objectives, and Mixed-Integer Formulation

The problem of placing contaminant sensors in a water distribution network to maximize the degree of afforded protection can be expressed as any of a number of standard problems in discrete location theory (Mirchandani and Francis, 1990). The specific selection depends on the outcome of a number of modeling decisions, e.g., whether sensor installation costs should be considered in the course of optimization, or if the objective is to minimize expected or worst-case impact. We base our analysis on the well-known p-median facility location problem (Daskin, 1995), building on our prior research efforts involving sensor placement optimization (Berry et al., 2006).

In the p-median problem, there are n customers and m *potential* facility locations. Exactly p of the m potential facilities are actually "opened", where $1 \leq p \leq m$, and each customer is "served" by the nearest open facility. For any fixed p, the objective is to determine the subset of p facilities that minimizes the sum of the distances between each customer and the nearest open facility. Although the p-median problem is not NP-hard for any fixed p (Garey and Johnson, 1979), it is computationally intensive to determine optimal solutions for instances with even modest n and m (Crainic et al., 2004). The p-median problem is closely related to the p-center problem, in which the objective is to minimize the *maximum* distance between a customer and the nearest open facility; the latter, however, is significantly more difficult to solve in practice.

In the present context of sensor placement for water distribution networks, a "customer" corresponds to an particular contamination scenario, i.e., an injection of contaminant into the network. Mirroring nearly all prior research on sensor placement optimization (see (Berry et al., 2005a) for the sole exception), we assume sensors are placed at network junctions, including tanks and reservoirs. Consequently, the set of potential facility locations corresponds to the set of network junctions. We assume a fixed budget of p general contaminant sensors, each of which is placed at a specific junction. Let $\mathcal{S}$ denote the set of potential contamination scenarios, and let $\mathcal{L}$ denote the set of network junctions. We additionally define a "dummy" network junction q corresponding to an abstract location at which detection occurs via mechanisms external to the sensor network, e.g., through observation of population behaviors. The introduction of q is due to use of finite sensor budgets, as not all contamination scenarios are detectable by a physical

sensor.

Next, let P denote the subset of junctions with installed sensors, where $|P| = p$ and $P \subseteq L$. For each combination of $s \in \mathcal{S}$ and $j \in \mathcal{L}$, we define d_{sj} as the aggregate, network-wide "damage" incurred if scenario s is first detected by a sensor at junction j, *assuming* a sensor is actually located at junction j. We further define d_{sq} for each $s \in \mathcal{S}$ as the network-wide damage incurred if scenario s is not detectable by any sensor in P, but rather through external inference. As discussed below, precise quantification of damage depends on the optimization objective; for illustrative purposes, the d_{sj} can be interpreted as the number of people exposed to the injected contaminant. The objective is then to minimize

$$x = \sum_{s \in \mathcal{S}} d_{sf(s,P \cup \{q\})} \tag{1}$$

where $f(s, P \cup \{q\})$ denotes the $j \in \mathcal{L} \cup \{q\}$ such that d_{sj} is minimized.

To determine an optimal sensor placement P and the corresponding minimal x, we formulate the p-median problem as a mixed-integer (linear) program (MIP), which we then solve using a commercially available MIP solver. The MIP-related terms used throughout this paper are defined in the *Mathematical Programming Glossary* (Greenberg, 2006). A MIP formulation of the p-median problem is given as follows:

$$\text{Minimize} \quad \sum_{s \in \mathcal{S}} \sum_{j \in \mathcal{L} \cup \{q\}} d_{sj} x_{sj} \tag{2}$$

$$\text{Subject to} \quad \sum_{j \in \mathcal{L} \cup \{q\}} x_{sj} = 1 \qquad \forall s \in \mathcal{S} \tag{3}$$

$$x_{sj} \leq y_j \qquad \forall j \in \mathcal{L} \tag{4}$$

$$\sum_{j \in \mathcal{L}} y_j = p \tag{5}$$

$$y_j \in \{0, 1\} \qquad \forall j \in \mathcal{L} \tag{6}$$

$$0 \leq x_{sj} \leq 1 \qquad \forall s \in \mathcal{S}, j \in \mathcal{L} \cup \{q\} \tag{7}$$

The binary y_j variables determine whether a sensor is placed at a junction $j \in \mathcal{L}$. Linearization

of Equation 1 is achieved through the introduction of auxiliary variables x_{sj}, which indicate whether a sensor placed at junction j is the first to detect scenario s. Constraint 4 ensures that detection is possible only if a sensor exists at junction j. The x_{sj} variables are implicitly binary due to a combination of binary y_j, Constraint 4, and the objective function pressure induced by Equation 2. Constraint 3 guarantees that each scenario $s \in \mathcal{S}$ is first detected by exactly one sensor, either at q or in the set $\mathcal{L}$; ties are broken arbitrarily. Finally, the objective function (Equation 2) ensures that detection of a scenario s is assigned to the junction $j \in \mathcal{L} \cup \{q\}$ such that d_{sj} is minimal.

This straightforward formulation of the p-median MIP is computationally tractable for small water distribution networks, but suffers from serious scalability limitations when the number of network junctions exceeds several thousand. To facilitate analysis of large-scale distribution networks, we solve a more complex variant of the above formulation, which yields identical solutions in significantly shorter run-times. The improved MIP formulation, detailed in (Berry et al., 2006), is used to generate the results presented in Section 3.

The impact of a potential contamination scenario is determined via transport simulation. Specifically, EPANET (Rossman, 1999) is used to generate a time-series τ_{sj} of contaminant concentration at each junction $j \in \mathcal{L}$ for each scenario $s \in \mathcal{S}$. The resulting time-series are then used to compute the network-wide impact d_{sj} of the scenario s assuming first detection via a sensor placed at junction j. More formally, let γ_{sj} denote the earliest time t at which a sensor at junction j can detect contaminant due to scenario s, e.g., when contaminant concentration reaches a specific detection threshold. If contaminant from scenario s fails to reach junction j, then $\gamma_{sj} = t^*$, where t^* denotes either the end of the simulation or an appropriate user-specified delay; otherwise, γ_{sj} can easily be computed from τ_{sj}. Next, we define $d_{sj} = d_s(\gamma_{sj})$, i.e., the aggregate, network-wide damage incurred if scenario s is first detected at time γ_{sj}. In our analysis, $d_{sq} = d_s(t^*)$. We assume without loss of generality that a sensor placed at a junction $j \in \mathcal{L}$ is capable of immediately detecting any scenario $s \in \mathcal{S}$ at j once non-zero concentration levels of a contaminant are present. Finally, in the absence of realistic alarm procedures and mitigation strategies, we assume that both consumption and propagation of contaminant is terminated once detection occurs.

By isolating objective-specific information to the d_{sj} coefficients, the p-median MIP seam-

lessly allows for optimization of disparate design objectives. We consider the following objectives in our analysis, variants of which have previously been considered by at least one research group; we briefly consider any key factors in the computation of these objectives from the set of τ_{sj} where necessary:

Population Exposed (*pe*) This objective quantifies the number of people *sickened* by exposure to the injected contaminant, as defined by the demand-based model described in (Murray et al., 2006). Specific values for the numerous parameters in the dosage-response computation can be obtained from the authors. Alternative models of population exposure have assumed the availability of population estimates on a per-junction basis (Berry et al., 2005a; Watson et al., 2004). While correcting the obvious deficiency of demand-based models, reliable estimates of population density are generally unavailable.

Time to Detection (*td*) This objective quantifies the time, measured in minutes, between the initiation of an injection and the earliest presence of non-zero contaminant concentration at a junction with a sensor. This objective was previously considered in the context of a flow-averaged model of the sensor placement problem (Watson et al., 2004).

Volume of Contaminated Water Consumed (*vc*) This objective quantifies the total volume of water, measured in gallons, extracted from the system prior to detection of non-zero contaminant concentration at a junction with a sensor. Extraction occurs at any junction – excepting tanks and reservoirs – with a positive, non-zero demand; the computation is independent of the magnitude of contaminant concentration. This objective is among the most widely studied, having previously been examined in (Kessler et al., 1998; Watson et al., 2004; Ostfeld and Salomons, 2004).

Mass of Contaminant Consumed (*mc*) This objective quantifies the total "mass" of injected contaminant, quantified in terms of the number of biological organisms, extracted from the system prior to the presence of non-zero contaminant concentration at a junction with a sensor. In contrast to *vc*, the *mc* objective is sensitive to the concentration of the contaminant. This objective has previously been considered in (Berry et al., 2006; Watson et al., 2004).

Number of Failed Detections (*nfd*) This objective quantifies the proportion of contamination scenarios for which no sensor detects non-zero contaminant concentration, i.e., the contamination scenario is "detected" by a sensor at the dummy q junction. This objective was previously examined in (Watson et al., 2004).

The Extent of Contamination (*ec*) This objective quantifies the length of pipe in a distribution system, measured in linear feet, that has been directly exposed to non-zero contaminant concentration. The entire length of an individual pipe is considered to be contaminated if (1) non-zero contaminant concentration is present at an end-point junction j and (2) water flow (as determined via contaminant transport simulation) enters the pipe from j; consequently, the measure is conservative. This objective was introduced in (Watson et al., 2004).

The *pe*, *vc*, and *mc* objectives are arguably related, in that they all attempt to quantify – either implicitly or explicitly – the impact of a contamination scenario on a population. However, the relationship between these and the remaining objectives, e.g., *ec* (which abstractly captures the cost of cleanup), is far less clear.

We conclude by discussing various assumptions underlying our p-median formulation of the sensor placement problem. The most critical and unrealistic assumption is the availability of general-purpose, perfect contaminant sensors. We base our decision to proceed under this assumption on two factors. First, sensor performance characteristics are not currently well-understood (although research is underway in this area, e.g., see (McKenna et al., 2006)) and data for real-world distribution networks is limited. Second, the tractability of computational techniques for imperfect-sensor variations of the p-median problem lags that of the perfect-sensor formulation (Berry et al.,), making extensive studies of the form presented in Section 3 infeasible. Although not considered here, we observe that the p-median formulation can be easily extended to handle a number of real-world constraints and factors, including fixed and/or invalid sensor locations, delays in raising a general alarm after detection by a sensor, thresholds on contaminant concentration, and installation costs (Berry et al., 2005b; Berry et al., 2006).[1]

[1]Many of these extensions are facilitated via straightforward modification of the computation of the d_{sj} damage coefficients.

3 Experimental Results and Analysis

We now use the p-median model introduced in Section 2 to analyze the relationship between different design objectives for three large, real-world water distribution networks. The p-median formulation of the sensor placement optimization problem is codified using the AMPL modeling language (Fourer et al., 2002); the resulting MIPs are solved to optimality using ILOG's AMPL/CPLEX 9.1 commercial solver package.[2] We defer to (Berry et al., 2006) for a discussion of the computational characteristics of the p-median MIP model and alternative heuristic techniques for its solution. The test networks and contamination scenarios are first detailed in Section 3.1. We then discuss in Section 3.2 the nature of individual design objectives for the test networks given a range of sensor budgets. Section 3.3 analyzes the impact of optimization for a single design objective on complementary design objectives.

3.1 The Test Networks and Contamination Scenarios

We report computational results for three real, large-scale municipal water distribution networks. The networks are denoted simply as Network1, Network2, and Network3; the identities of the corresponding municipalities are withheld due to security concerns. Network1 consists of roughly 400 junctions, 500 pipes, and a small number of tanks and reservoirs. Network2 consists of roughly 3000 junctions, 4000 pipes, and roughly 50 tanks and reservoirs. Network3 consists of roughly 12000 junctions, 14000 pipes, and a handful of reservoirs; there are no tanks or well sources in this municipality. All of the models are skeletonized, although the degree of skeletonization in Network1 and Network2 is much greater than in Network3.

Graphical representations of Network1, Network2, and Network3 are respectively shown in Figures 1, 2, and 3. Each figure was produced by manually "morphing" or altering (e.g., through pipe lengthening or coordinate translation/rotation) key topological features of the original network structure to further inhibit identification of the source municipality. Local topologies were largely preserved in this process, such that the graphics faithfully capture the overall characteristics of the underlying network structures. Sanitized versions of all three networks, in the form of EPANET input files, are available from the authors. While these files contain no coordinate

[2] `http://www.ilog.com/products/cplex`

information, all data other than that relating to labels (which have been anonymized) are unaltered. Consequently, all computed hydraulic and water quality information accurately reflect (within the fidelity limits of the data and the computational model) the dynamics of the source municipality. Our goals in making these files available to the broader research community are to facilitate independent replication of our results and to introduce larger, more realistic networks into the currently limited suite of available test problems.

Network hydraulics are simulated over a 96 hour duration, representing multiple iterations of a typical daily demand cycle. For each junction with non-zero demand, a single contamination scenario is defined. Each scenario starts at time $t = 0$ and continues for a duration of 12 hours. Scenarios are modeled as biological mass injections with a constant rate of $5.78e + 10$ organisms per minute. We note that the p-median formulation, via the d_{sj}, allows for the use of arbitrarily complex attack scenarios, e.g., multiple simultaneous injection sites with different contaminants at variable injection strengths and durations.

We assume uniform scenario probabilities, such that all results (defined by Equation 1) are normalized by the number of non-zero demand junctions to obtain an expectation. Water quality simulations are performed for each scenario, with a time-step resolution of 5 minutes. The resulting τ_{sj} are then used to compute the impact parameters d_{sj} for the various design objectives, as previously described in Section 2. All hydraulic and water quality simulations are performed using EPANET (Rossman, 1999).

3.2 Characteristics of Individual Design Objectives

Many of the design objectives introduced in Section 2 have either been considered only in the context of small test networks or on less accurate formulations of the sensor placement problem. Further, performance case-studies of sensor configurations for large-scale, real-world distribution networks are of interest to practitioners and researchers, and are absent in the broader literature. Consequently, we first consider the nature of the individual design objectives on our test networks.

Table 1 reports the performance of optimal sensor placements for each of our design objectives on Network1, over a range of sensor budgets p. In the absence of sensors, the mean impact of an contamination scenario is significant, especially when measured in terms of *pe*,

vc, and *ec*. Addition of even $p = 5$ sensors yields order-of-magnitude or larger reductions in many of the objectives, including *pe*, *vc*, *nfd*, and *ec*. Independent of objective, a budget of only $p = 10$ is sufficient to yield impacts of at most 28% that of the $p = 0$ solution; for many objectives, e.g., *pe* and *nfd*, the impact is approximately 5% of the $p = 0$ solution. For all but the *mc* objective, $p = 50$ is sufficient to achieve near-prefect protection. In all cases, excellent performance can be achieved with a budget p equal to a small fraction that of the total number of network junctions.

In Table 2, we report the performance of optimal sensor placements for each of our design objectives on Network2; recall that Network2 is roughly an order of magnitude larger in terms of the number of network elements than Network1. Considering the $p = 0$ solution, there is a marked growth in impact relative to Network1, which is consistent given the differences in network size and roughly equivalent degrees of skeletonization. In terms of *pe*, over 14,000 individuals are sickened *on average* across the range of possible contamination scenarios, while large numbers of specific scenarios (not reported) impact far larger numbers of individuals. Similarly, roughly 66 miles of pipe are exposed to contaminant on average, while 11.7 million gallons of contaminated water are extracted from the distribution network prior to detection. Despite the network size, $p = 5$ is still sufficient to yield an order-of-magnitude or greater reduction in damage relative to the $p = 0$ solution for the *pe*, *vc*, and *ec* objectives. Relative to the results for Network1, a large ($p = 100$) number of sensors are required to reduce impacts to 28% or less than that of the $p = 0$ solution for all objectives, while slightly over 1000 sensors are required to achieve near-perfect protection. We observe that both results are consistent with the differences in network size.

Finally, we consider individual design objective results for Network3, reported in Table 3. Despite the larger size relative to Network2, the mean impacts under $p = 0$ for the network-dependent performance measures (*pe*, *vc*, *mc*, and *ec*) are comparable to those reported for Network1. This is due to the lesser degree of skeletonization used in the development of the Network3 model. Given the absolute network size, a very small budget of 5 sensors yields significant protection relative to the baseline $p = 0$ solution. However, very large numbers of sensors ($p \approx 1000$) are required to yield impacts of at most 28% that of the $p = 0$ solution. While small relative to the total number of network junctions, such large budgets are clearly

unrealistic in the near-to-moderate term.

Overall, the results presented in Tables 1 through 3 demonstrate that independent of objective, a very small number of sensors can yield very large – and often order-of-magnitude – reductions in impacts relative to the $p = 0$ solution. In all of our test networks, a sensor budget equal to at most 10% of the total number of network junctions yields reductions in the mean impact of an contamination scenario, quantified by any design objective, of at least 80% or more relative to the $p = 0$ solution. In other words, a very small number of sensors, in both relative and absolute terms, provide a significant degree of protection. As the number of sensors grows, however, the per-sensor benefit diminishes greatly.

3.3 The Impact of Optimization on Complementary Objectives

Several researchers have argued for the use of specific design objectives to develop sensor placements in water distribution networks (Kessler et al., 1998; Kumar et al., 1998). Populations would undoubtedly prefer minimization of *pe* as the primary objective. In contrast, potential economic impacts influence both adversaries and decision-makers, leading to an argument for minimization, or at least consideration, of *ec*. Similarly, it is not unreasonable to expect an CWS to detect a large proportion of possible contamination scenarios within a reasonable timeframe (*nfd*). In any case, no single view is likely to prevail, and planners will realistically have to understand the trade-offs between the various design objectives. Ideally, the objectives of interest are highly correlated, such that optimal solutions with respect to one objective yield near-optimal solutions with respect to the others. [3] Unfortunately, as we now discuss, we do not necessarily observe this behavior on our test networks.

Before proceeding, it is important to note that when optimizing any individual objective, solution quality with respect to secondary objectives is entirely ignored. Consequently, when given a problem with multiple globally optimal solutions with respect to the primary objective, we are not selecting among the "best" with respect to any particular secondary objective. Rather, we are simply using the particular solution returned by our MIP solver. However, there is no evidence that such consideration would influence the results presented below.

[3] We use this informal notion of correlation throughout, as opposed to the more familiar concept of statistical correlation.

We begin with an example in which we investigate how minimization of the *nfd* objective on Network1 impacts solution quality with respect to the other design objectives. Given $p = 25$, *nfd* $= 0.0$, such that all contamination scenarios are detected within the 96-hour simulation period. However, the solution that yields an optimal value of *nfd* given $p = 25$ also results in *pe* $= 234$, which represents a 1070% deviation from the optimal value of *pe* $= 20$ given $p = 25$ (as shown in Table 1). Similarly large deviations are observed for *ec* and *vc* (677% and 1017%, respectively), while the values of *td* and *mc* represent "only" 180% and 223% deviations from optimality, respectively. Although such large deviations were unexpected, a qualitative understanding of the underlying causes were easily determined via straightforward *a posteriori* analyses. For example, minimization of *nfd* subject to a limited sensor budget tends to yield sensor placements near the leaves or end-points of the distribution network. In doing so, many upstream nodes are exposed to contaminant for longer durations, resulting in greater overall ingestion.

Characterizing and analyzing the interactions between all of the design objectives introduced in Section 2 is beyond the scope of this paper. Rather, we examine illustrative cases, one for each of our test networks. In the first case, we consider how optimization of *pe* on Network1 impacts solution quality, in terms of deviation from optimality, with respect to the complementary design objectives. The results are shown in Table 4 for a range of sensor budgets; for each complementary objective, both the absolute and percentage difference from optimality (e.g., determined in part using the data recorded in Table 1) are reported. We observe that large absolute deviations do not necessarily correspond to large percentage deviations, and vice versa. Values of ∞ in a percentage difference column indicate the optimal value for the corresponding objective given equal p is 0. The results indicate two specific trends. First, for small-to-moderate p ($5 \leq p \leq 50$), *pe* is not significantly correlated with any of the other objectives; both percentage and absolute deviations from the optimal values of the complementary objectives are unexpectedly and uniformly large. Second, for large p ($p \geq 100$), the correlation between *pe* and the complementary objectives begins to improve, reaching near-perfect correlation once $p \approx 200$. The convergence as $p \rightarrow |\mathcal{L}|$ is expected: as a larger number of network junctions are covered by sensors, the similarity between the optimal placements for different objectives necessarily increases. Unfortunately, the differences are greatest in the most likely regime for

CWS deployment in the near-to-moderate term, i.e., when the sensor budgets p are small, such that few network components are covered by sensors.

Next, we consider a similar analysis on the larger Network2 model, in which we analyze how optimization of *nfd* impacts performance in terms of secondary design objectives. The results, presented in an analogous form to those shown in Table 4, are shown Table 5. In contrast to the results for Network1, we do observe fairly strong correlation between some objectives. Specifically, optimal *nfd* solutions yield near-optimal *td* performance, and only moderately worse *mc* performance. For small-to-moderate sensor budgets ($5 \leq p \leq 100$), optimal *nfd* solutions yield significant absolute and percentage deviations from optimality for the *pe*, *ec*, and *vc*. Further, these deviations persist even given very large sensor budgets of $p = 1000$ and greater. Comparison of the Network1 and Network2 results further reinforce the general observation that optimization with respect to a specific design objective can yield highly sub-optimal performance with respect to secondary objectives. However, the results also indicate that the degree of sub-optimality appears to be highly dependent upon both the test network and the design objectives under consideration.

Finally, we consider on Network3 the impact of optimization of *vc* on secondary design objectives; the results are shown in Table 6. Relative to the results shown in Tables 4 and 5, the deviations from optimality for the secondary objectives are significantly lower in both percentage and absolute terms (e.g., most deviations are less than 100%) and there is stronger correlation between many of the objectives (e.g., *pe*, *mc*, and *vc*). In contrast to Network1 and Network2, Network3 is supplied strictly through reservoirs. The lack of tanks strongly limits the flow dynamics, which partially explains both the lower deviations from optimality and the stronger correlations observed between many of the objectives.

Overall, our results support three general conclusions. First, there are significant risks associated with optimization of sensor placements with respect to any particular design objective. The results shown in Tables 4 through 6 demonstrate that optimal solutions with respect to any specific design objective can be far from optimal with respect to a range of complementary objectives. Second, and counter-intuitively, the lack of significant correlation between objectives may not improve with small-to-moderate increases in p. In other words, a large sensor budget does not necessarily mitigate the risk associated with optimization with respect to a sin-

gle design objective. Third, the nature of the correlation between various objectives is highly problem-dependent, suggesting that a comprehensive analysis is required on a per-network basis.

4 Discussion

While there is significant risk associated with focusing on any individual design objective in sensor placement optimization, it is unclear whether or not it may be possible to construct a solution that more carefully balances a range of design objectives, or whether there exist objectives that are strictly complementary. To explore this question, we again consider minimization of *pe* on Network1 given $p = 10$, but impose additional constraints on the corresponding p-median MIP described in Section 2 such that the values of *td*, *nfd*, and *vc* are respectively constrained to respectively no greater than 50%, 20%, and 30% of their optimal values (as recorded in Table 1). While the optimal value of *pe* increases 60% relative to the baseline MIP lacking the side constraints, the deviations from optimality for all but one of the other design objectives are significantly reduced relative to the baseline MIP results, as reported in Table 4. Specifically, the deviation from optimality given the additional side constraints is 29% for *td* (down from 96%), 28% for *vc* (down from 42%), 20% for *mc* (down from 58%), and 180% for *nfd* (down from 440%). However, the deviation from optimality for *ec* grows from 46% to 78%, indicating that at least for Network1, the *pe* and *ec* are strongly complementary.

Due to space limitations, we omit a broader analysis of the trade-offs between the various design objectives. Rather, we simply observe that by sacrificing solution quality with respect to a primary design objective, it is possible to gain significant improvements with respect to secondary objectives and avoid some of the brittleness associated with solutions that are optimal with respect to individual design objectives. However, due to the competing nature of some design objectives on some networks, it is not always possible to improve the performance of all secondary objectives simultaneously. Ultimately, a detailed understanding of the relationship between various design objectives is required for decision-makers to develop robust sensor placements for CWS deployment. Such understanding is even more crucial in the early phases of CWS deployment, where sensors budgets are small and the correlation between the optimal-

ity of different design objectives is weakest. As the sensor budgets grow to cover a significant fraction of the network, this correlation tends to increase, and the need for multiple-objective analysis is less crucial.

5 Conclusions

To date, all research on sensor placement optimization in water distribution networks presupposes a given, fixed design objective. Several disparate objectives have been proposed, and there are associated arguments — both implicit and explicit — for why one particular objective should be preferred over another. Yet, preference for any fixed objective is potentially risky, given the current lack of understanding of the relationships among the proposed objectives. We have characterized some of the inter-dependencies among a range of optimization objectives on three large-scale test networks. The majority of these objectives are uncorrelated, in that optimal solutions with respect to any one objective are often highly sub-optimal with respect to complementary objectives. Further, increasing the number of sensors frequently fails to improve the correlation. However, the risks can be mitigated in some circumstances by considering solutions that are sub-optimal with respect to all performance objectives, which in turn requires a thorough understanding of how different objectives are related. Overall, the implications of our results for both researchers and planners is clear: robust algorithms for the sensor placement problem must carefully and simultaneously consider multiple design objectives.

Acknowledgements

Sandia is a multipurpose laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000. This work was funded through an Interagency Agreement with the United States Environmental Protection Agency.

References

Berry, J., Carr, R., Hart, W., Leung, V., Phillips, C., and Watson, J. On the placement of imperfect sensors in municipal water networks. In *Proceedings of the 2006 Symposium on Water Distribution Systems Analysis*. Submitted.

Berry, J., Fleischer, L., Hart, W., Phillips, C., and Watson, J. (2005a). Sensor placement in municipal water networks. *Journal of Water Resources Planning and Management*, 131(3).

Berry, J., Hart, W., Phillips, C., Uber, J., and Walski, T. (2005b). Water quality sensor placement in water networks with budget constraints. In *Proceedings of the ASCE/EWRI Congress*.

Berry, J., Hart, W., Phillips, C., Uber, J., and Watson, J. (2006). Sensor placement in municipal water networks with temporal integer programming models. *Journal of Water Resources Planning and Management: Special Issue on Drinking Water Distribution Systems Security (To Appear)*.

Crainic, T., Gendreau, M., Hansen, P., and Mladenovic, N. (2004). Cooperative parallel variable neighborhood search for the p-median. *Journal of Heuristics*, 10:293–314.

Daskin, M. (1995). *Network and Discrete Location: Models, Algorithms, and Applications*. John Wiley & Sons, New York, NY.

Fourer, R., Gay, D., and Kernighan, B. (2002). *AMPL: A Modeling Language for Mathematical Programming*. Duxbury Press, 2nd edition.

Garey, M. and Johnson, D. (1979). *Computers And Intractability: A Guide To The Theory Of NP-Completeness*. W.H. Freeman and Company, New York.

Greenberg, H. (1996–2006). *Mathematical Programming Glossary*. World Wide Web, http://www.cudenver.edu/~hgreenbe/glossary/.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management*, 124(4):192–198.

Kumar, A., Kansal, M., and Arora, G. (1998). Discussion of "Detecting accidental contaminations in municipal water networks". *Journal of Water Resources Planning and Management*, 124(4):308–310.

McKenna, S., Hart, D., and Yarrington, L. (2006). Impact of sensor detection limits on protecting water distribution systems from contamination events. *Journal of Water Resources Planning and Management: Special Issue on Drinking Water Distribution Systems Security (To Appear)*.

Mirchandani, P. and Francis, R., editors (1990). *Discrete Location Theory*. John Wiley & Sons, New York, NY.

Murray, R., Uber, J., and Janke, R. (2006). Modeling acute health impacts resulting from ingestion of contaminated drinking water. *Journal of Water Resources Planning and Management: Special Issue on Drinking Water Distribution Systems Security (To Appear)*.

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.

Rossman, L. (1999). The EPANET programmer's toolkit for analysis of water distribution systems. In *Proc. of the Annual Water Resources Planning and Management Conference*.

Watson, J., Greenberg, H., and Hart, W. (2004). A multiple-objective analysis of sensor placement optimization in water networks. In *Proceedings of the ASCE/EWRI Congress*.

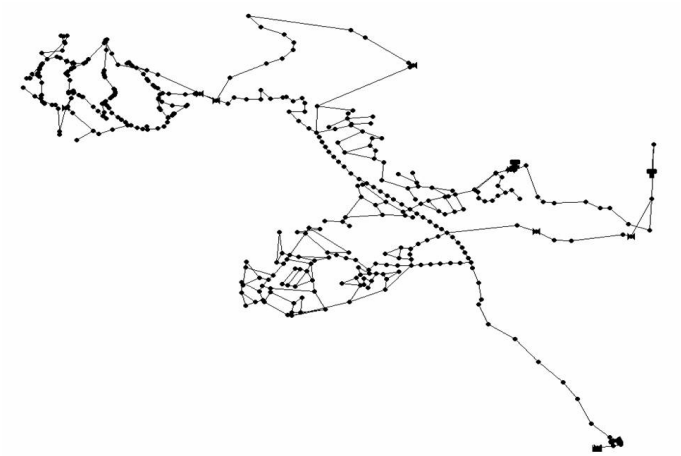

Figure 1: Graphical depiction of Network1 topology. See text for details.

Figure 2: Graphical depiction of Network2 topology. See text for details.

p	*pe*	*td*	*vc*	*mc*	*nfd*	*ec*
0	2445	5760	1288000	3.95e+13	1.0	41268
5	143	1600	8357	1.69e+13	0.13	4084
10	63	985	3010	1.10e+13	0.05	2444
25	20	219	660	3.43e+12	0.0	982
50	5	41	158	3.79e+11	0.0	375
100	0	0	2	4.36e+08	0.0	0
200	0	0	0	0	0.0	0
400	0	0	0	0	0.0	0

Table 1: Optimal values of design objectives for a range on p on Network1. The units of measure for the various design objective are respectively number of individuals, minutes, gallons, organisms, proportion of total contamination scenarios not detected, and linear feet.

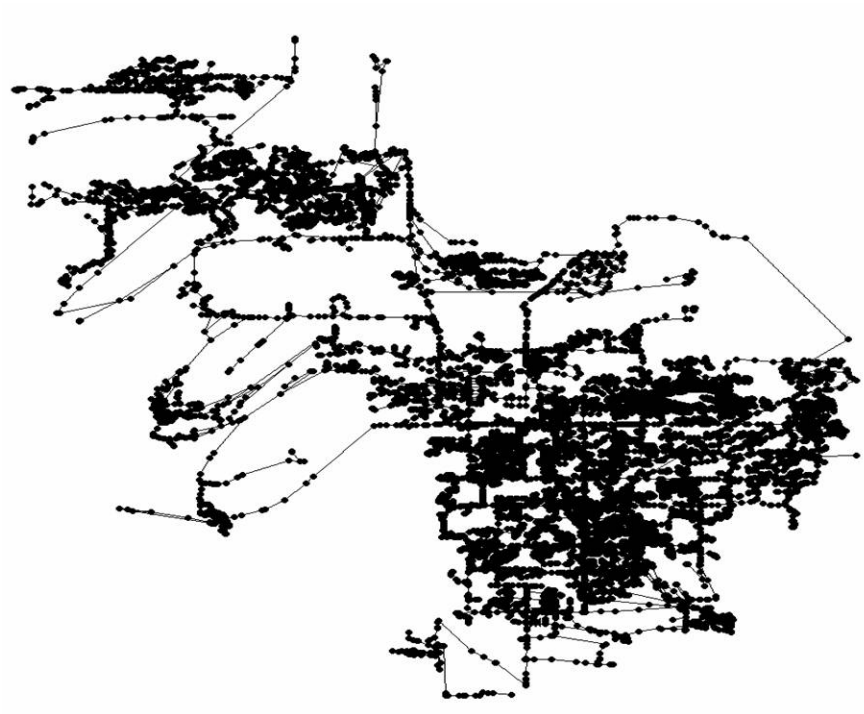

Figure 3: Graphical depiction of Network3 topology. See text for details.

p	*pe*	*td*	*vc*	*mc*	*nfd*	*ec*
0	14217	5760	11667200	3.90e+13	1.0	344376
5	1709	3218	162640	2.71e+13	0.47	38822
10	1061	2860	66241	2.44e+13	0.41	22062
50	347	2028	13675	1.74e+13	0.29	6382
100	205	1632	7549	1.42e+13	0.23	3604
500	50	124	1527	2.51e+12	0.0	754
1000	14	11	272	7.18e+10	0.0	84
2000	0	0	0	0	0.0	0

Table 2: Optimal values of design objectives for a range on p on Network2.

p	*pe*	*td*	*vc*	*mc*	*nfd*	*ec*
0	2249	5760	978487	4.15e+13	1.0	138543
5	764	4523	98751	3.63e+13	0.69	41623
10	498	4134	52354	3.41e+13	0.62	26973
50	169	3224	10112	2.72e+13	0.46	8801
100	103	2832	5305	2.43e+13	0.39	5424
500	34	1642	1311	1.58e+13	0.20	1820
1000	20	987	672	1.08e+13	0.09	1153
2000	11	310	287	5.14e+12	0.0	684

Table 3: Optimal values of design objectives for a range on p on Network3.

	td		*vc*		*mc*		*nfd*		*ec*	
p	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.
5	73	1160	206	17212	30	5.10e+12	246	0.32	42	1707
10	96	944	42	1263	58	6.40e+12	440	0.22	46	1119
25	400	875	31	205	174	5.98e+12	∞	0.15	68	663
50	1685	513	58	91	1396	5.29e+12	∞	0.11	107	403
100	∞	74	700	14	87385	3.81e+11	0	0.0	∞	106
200	0	0	0	0	∞	1.16e+8	0	0.0	∞	26
400	0	0	0	0	0	0	0	0.0	0	0

Table 4: Percentage and absolute deviations from optimality for complementary design objectives on Network1, given a *pe*-optimal solution.

	pe		*td*		*vc*		*mc*		*ec*	
p	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.
5	472	8062	11	342	819	1332100	28	7.60e+12	567	220233
10	773	8204	12	338	1651	1093519	37	9.10e+12	980	216231
50	933	3237	10	203	1779	243245	44	7.60e+12	1467	93647
100	1030	2111	8	127	1473	111161	37	5.30e+12	1742	62793
500	844	422	72	89	824	12578	91	2.20e+12	1490	11234
1000	1586	222	1100	121	2526	6872	3897	2.80e+12	6636	5574
2000	∞	58	∞	53	∞	2385	∞	1.10e+12	∞	2077

Table 5: Percentage and absolute deviations from optimality for complementary design objectives on Network2, given a *nfd*-optimal solution.

	pe		*td*		*mc*		*nfd*		*ec*	
p	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.
5	4	30	3	115	< 1	3.00e+11	10	0.07	10	4324
10	7	33	5	190	< 1	2.00e+11	15	0.09	17	4518
50	14	24	9	283	2	5.00e+11	24	0.11	35	3101
100	22	23	12	344	2	6.00e+11	33	0.13	46	2475
500	35	12	37	612	12	1.90e+12	80	0.16	75	1371
1000	50	10	73	723	25	2.70e+12	200	0.18	94	1089
2000	55	6	240	743	63	3.25e+12	∞	0.16	121	829

Table 6: Percentage and absolute deviations from optimality for complementary design objectives on Network3, given a *vc*-optimal solution.