Addressing Modeling Uncertainties in Sensor Placement for Community Water Systems

Robert D. Carr* Harvey J. Greenberg† William E. Hart‡
Cynthia A. Phillips§

Abstract

We present a model for optimizing the placement of sensors in municipal water networks to detect contaminants that are maliciously or accidentally injected. An optimal sensor configuration is desirable to minimize the cost and maximize the information provided by the sensors. We formulate sensor placement problems as mixed-integer programs, for which the objective coefficients are not known with certainty. We present three robust optimization models that differ in how the coefficients in the objective vary. Under one set of assumptions there exists a sensor placement that is optimal for all realizations of the coefficients. Under other assumptions, we apply sorting to solve each worst-case realization. The most difficult case is where the objective parameters are bilinear, for which we propose a branch-and-bound method.

1 Introduction

Recent terrorist attacks have heightened concerns about whether community water systems are sufficiently well protected to ensure a safe and reliable supply of drinking water in the United States and around the world. Community water monitoring programs, such as weekly sampling, are often inadequate to quickly detect malicious or accidental contamination in a water system. Consequently, there is growing interest in the use of contaminant sensors to provide ongoing monitoring of water quality. Such a monitoring system could significantly mitigate

*Algorithms and Discrete Math Dept, Sandia National Laboratories, Mail Stop 1110, P.O. Box 5800, Albuquerque, NM, 87185-1110; PH (505) 845-8562, FAX (505) 845-7296; rdcarr@sandia.gov.

†Mathematics Department, Univ Colorado at Denver, P.O. Box 173364, Denver, CO 80217-3364l PH (303) 556-8464; Harvey.Greenberg@cudenver.edu

‡Algorithms and Discrete Math Dept, Sandia National Laboratories, Mail Stop 1110, P.O. Box 5800, Albuquerque, NM, 87185-1110; PH (505) 844-2217, (505) 845-7296; wehart@sandia.gov.

§Algorithms and Discrete Math Dept, Sandia National Laboratories, Mail Stop 1110, P.O. Box 5800, Albuquerque, NM, 87185-1110; PH (505) 845-7296, FAX (505) 845-7296;caphill@sandia.gov.

the risk from contamination, and thus it complements other physical security approaches for community water systems.

A good sensor placement minimizes cost and maximizes the information available for contamination containment and remediation across the full range of possible contamination scenarios. Uncertainties in attack risk and water demand and variability in population density can affect the optimal solution considerably. This is a major concern for sensor placement because accurate, detailed information is not available for many community water networks (e.g., how many people consume water at a particular location on a "normal" day). In this paper, we consider "robust" problem formulations for sensor placement that directly account for some types of data uncertainties.

We consider robustness for sensor placement problems that are formulated as mixed-integer programs (MILPs). A wide variety of combinatorial optimization problems can be formulated as MILPs, for which globally optimal solutions can be generated with standard, commercial optimization software. We have recently demonstrated that MILPs can be effectively applied to solve moderately large sensor-placement problems [1] and that they can be used to express many sensor placement quality measures [10].

We show how to exploit the linear structure of a MILP model to quantify the impact of data uncertainty in the objective function on the value of the globally optimal solution. Two sensor placement objectives illustrate our analysis: (1) minimize expected population exposed (PE), and (2) minimize the expected portion of the network that becomes contaminated (NC).

We first consider models in which uncertainty can be expressed as a random variable. In several scenarios, it is possible to use properties of these random variables to formulate deterministic MILPs that minimize the expected performance measure. Thus, we do not need to apply more costly methods that perform optimization under uncertainty (e.g., search methods that use samples from the random variables to identify descent directions).

We also consider the case where data uncertainties are bounded within a range of values. For example, this model of uncertainty might reflect limited data precision. For objectives like NC, we show that the solution to a robust min-max formulation is exactly the solution to the original MILP. More complex objectives, like PE, contain terms with multiplicative uncertainties. If considered independently, these uncertainties can again be quantified in a straightforward manner. Further, we show that assessing the impact of these uncertainties together requires only the solution of an alternative MILP formulation.

Section 2 describes integer programming formulations for the Min Network Contaminated and Min Population Exposed sensor placement problems. Section 3 describes random uncertainties, and Section 4 gives three models of interval uncertainty (see [11] for related analysis of minimum spanning trees under interval uncertainty). Kouvelis and Yu [4] provide general background on robust discrete optimization. The *Mathematical Programming Glossary* [2] provides succinct definitions of mathematical programming terms.

2 Integer Programming Models

We now describe two MILP models of sensor placement that we use to illustrate our robust model analysis. We model an attack as the release of a large volume of harmful contaminant at a single point in the network at a single injection site. We assume that typical water demands throughout a day occur in one of a fixed set of patterns, $P = \{1, \ldots, N_p\}$. The model makes no assumptions about how long each pattern holds, how often it appears, or the order in which the patterns appear. We use EPANET [8] to determine an acyclic water flow given a set of available water sources, assuming each demand pattern holds steady for sufficiently long. Further, we ignore the magnitude of water velocity, requiring only its direction and that it be sufficiently large.

For any particular attack, we assume that all points "downstream" of the release point can be contaminated. That is, water moves quickly to the demand points. This model assumes perfect sensors that raise a general alarm precisely when contamination passes. This is a vast simplification of water transport in real networks, but the issues of data uncertainty still apply in more realistic IP models.

The two MILP models we consider here differ by their objectives. Otherwise, they are defined on a common undirected graph, $G = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ and $E = \{e_1, \ldots, e_m\}$. There are no parallel edges, so we can denote e by its endpoints, $\{v_i, v_j\}$. A pattern defines a direction of flow on edges of this graph such that the graph becomes acyclic, and we denote the start and finish nodes of an edge for pattern p by $V^s(e, p)$ and $V^f(e, p)$, respectively. An edge may have no flow during a particular pattern, in which case it is oriented arbitrarily and the flow is tagged as zero. A *path* from node v_i to v_j for pattern p is a sequence of edges, $e_{j_1}, \ldots, e_{j_k}$, such that they have positive flow and their orientation is downstream: $V^s(e_{j_1}, p) = v_i$, $V^f(e_{j_k}, p) = v_j$, and $V^f(e_{j_l}, p) = V^s(e_{j_{l+1}}, p)$ for $l = 1, \ldots, k-1$. If such a path exists, we say v_j is *reachable* from v_i for pattern p, and we let $\mathcal{R} = \{(i, j, p) : v_j \text{ is reachable from } v_i \text{ for pattern } p\}$.

The primary decision variables for optimization are where to place the sensors: $s_e = 1$ if a sensor is placed at edge e and otherwise $s_e = 0$. We suppose that we have a limited number of sensors, so we have $\sum_{e \in E} s_e \leq S_{max}$. Secondary decision variables are x_{ijp}, for $(i, j, p) \in \mathcal{R}$, such that $x_{ijp} = 1$ if node v_j becomes contaminated, given an attack at node v_i for pattern p; otherwise, $x_{ijp} = 0$.

If an attack occurs at node v_i for pattern p, v_i becomes contaminated, which in turn contaminates its outgoing arcs. The downstream effect continues until an edge is reached that senses the contamination. From that point downward, there is no undetected contamination; otherwise, a node is contaminated whenever any one of its ingoing arcs is contaminated. The following constraints represent this and comprise the MILP feasible region:

$$x \in \{0,1\}^{|\mathcal{R}|},\ s \in \{0,1\}^{|E|},\ x_{iip} = 1,\ \sum_{e \in E} s_e \leq S_{max}, \tag{1}$$

$$x_{ijp} \geq x_{ikp} - s_e \text{ for } \{(i,j,p), (i,k,p) \in \mathcal{R},\ V^s(e,p) = k,\ \text{and } V^f(e,p) = j\}. \tag{2}$$

Constraints (1) require variables to be binary; however, this could be relaxed for the contamination variables x because they will be binary in computed optimal solutions anyway. The constraint $x_{iip} = 1$ (for all i, p) means that node v_i is contaminated if it is attacked when the network flow is using pattern p.

Constraint (2) requires node v_j to be contaminated if it is a descendent (under pattern p) of attacked node v_i, and there is some path from v_i to v_j that has no sensor. To see this, note that such a constraint forces $x_{ijp} = 1$ if, and only if, the following conditions are met: (1) there is a predecessor node v_k that is contaminated by the attack ($x_{ikp} = 1$); and (2) the connecting edge ($e = (v_k, v_j)$) has no sensor ($s_e = 0$). If either of these conditions does not hold, x_{ijp} is not constrained to be 1; the minimization will choose $x_{ijp} = 0$ if all such constraints allow it.

In general, we do not know a priori where or when an attack will occur, so we consider a set of weighted attack scenarios. Each scenario consists of an attack point and a flow pattern. Attack weights may reflect expert knowledge of network vulnerabilities. We consider two objectives for which we wish to minimize a weighted sum of values of some outcome, given the attack weights α_{ip}. The α_{ip} may be considered probabilities, but our analysis does not depend upon that interpretation of the uncertainty. In fact, our robust analysis addresses the uncertainty of these weights.

The first objective is the portion of the extent of contamination, measured by the lengths of the pipes represented by the edges of the network. Edge e becomes contaminated under flow pattern p if node $V^s(e, p)$ is contaminated. Let $E^s(j, p) = \{e \in E : V^s(e, p) = j \text{ and } e \text{ has positive flow}\}$. Then, given the length of a pipe is $L(e)$, let $\lambda_{jp} = \sum_{e \in E^s(j,p)} L(e)$. The (weighted) total amount of contaminated pipe is given by

$$\text{MILP-NC: minimize NC} = \sum_{(i,j,p) \in \mathcal{R}} \alpha_{ip} \lambda_{jp} x_{ijp} \text{ subject to (1), (2).} \tag{3}$$

The second objective is the population exposed to the contaminant, using estimates of populations at nodes. Let δ_{jp} be the population at node v_j for pattern p. Then, the population exposed to contamination is given by

$$\text{MILP-PE: minimize PE} = \sum_{(i,j,p) \in \mathcal{R}} \alpha_{ip} \delta_{jp} x_{ijp} \text{ subject to (1), (2).} \tag{4}$$

Both of these objectives have the structure of two parameters defining the coefficient of the contamination variables x, which depend on the decision variables s_e. The NC objective is defined by parameters α and λ, though we only expect significant uncertainty in the attack probabilities α. The PE objective is a product of two uncertain parameters, since the population values δ_{jp} are also estimated values.

3 Random Uncertainties

A natural method for modelling data uncertainties is to describe the likely values of data using random variables. For example, many sources of data can be effectively characterized by a Normal random variable, whose mean value and standard deviation are empirically estimated. Suppose that we can model the attack probabilities and population densities at different sites and for different patterns as random variables A_{ip} and D_{jp} respectively. Then a standard method to account for data uncertainty is to consider the expected value of the model objective. For example, we can rewrite the PE objective as

$$\min E\left(\sum_{(i,j,p)\in\mathcal{R}} A_{ip} D_{jp} x_{ijp}\right)$$

and the NC objective as

$$\min E\left(\sum_{(i,j,p)\in\mathcal{R}} A_{ip} \lambda_{jp} x_{ijp}\right).$$

These objectives are the only place where this data is used, so the constraint matrix is unchanged. Also, we assume that the length data λ_{jp} is well-known and thus not considered in our robust analysis. In general, these expected-performance objectives may require the use of optimization solvers that directly address these uncertainties. However, we describe assumptions for which these expected value objectives can be re-cast to simple linear objectives by exploiting properties of the probability distributions for A_{ip} and D_{jp}.

Because the λ_{jp} parameters are certain (constant) and the x_{ijp} are fixed for a given sensor placement, we can rewrite the NC objective as

$$\min \sum_{(i,j,p)\in\mathcal{R}} E(A_{ip}) \lambda_{jp} x_{ijp}.$$

Consequently, our robust formulation can be simply recast as the nonrandom NC MILP model using $\alpha_{ip} = E(A_{ip})$.

We now consider the PE objective. Suppose that we model population density at each site with a stable, baseline component plus an uncertain component that is uniformly distributed across all sites. This may be a reasonable model in contexts where most of the population is stable (e.g. where it can be estimated with Census data) and we have limited information to predict the distribution of the uncertain component of the population. Let $\hat{\delta}_{jp}$ be the baseline population estimate, which accounts for a fraction b_p of the total population $D_p = \sum_j \delta jp$ (that is, $\sum_j \hat{\delta}_{jp} = b_p D_p$). Let $\hat{D}_{jp}$ be the random variables that account for the remaining uncertainty.

In a malicious attack, the A_{ip} could depend on the baseline population values $\hat{\delta}_{ip}$, but they will be independent of the random fluctuations $\hat{D}_{jp}$. For example, an attacker may be more likely to attack a mall because a lot of people go there, regardless of the exact population on a given day. Then the PE objective becomes

$$\sum_{p=1}^{P} \sum_{i=1}^{n} \sum_{j:(i,j,p)\in\mathcal{R}} E\left[A_{ip}\left(\hat{\delta}_{jp} x_{ijp} + \hat{D}_{jp} x_{ijp}\right)\right].$$

Using linearity of expectations, the objective is

$$\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j:(i,j,p)\in\mathcal{R}}\left(\hat{\delta}_{jp}x_{ijp}E[A_{ip}]+x_{ijp}E[A_{jp}*\hat{D}_{jp}]\right).$$

Because the A_{jp} are independent of the population fluctuations, this becomes

$$\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j:(i,j,p)\in\mathcal{R}}\left(\hat{\delta}_{jp}x_{ijp}E[A_{ip}]+x_{ijp}E[A_{jp}]*E[\hat{D}_{jp}]\right).$$

We have $\sum_j \hat{D}_{jp}=(1-b_p)D_p$ for all instances of the $\hat{D}_{jp}$ variables. If we assume that the $\hat{D}_{jp}$ are uniformly distributed with this constraint, then for a given p they are symmetric with respect to j so they all have the same mean. Thus, $E(\hat{D}_{jp})=(1-b_p)D_p/n_p$ (where n_p is the number of nodes j for which there is population uncertainty in time period p). Our objective function simplifies to

$$\sum_{p=1}^{P}\sum_{i=1}^{n}\sum_{j:(i,j,p)\in\mathcal{R}}\left(\hat{\delta}_{jp}x_{ijp}E[A_{ip}]+x_{ijp}E[A_{jp}](1-b_p)D_p/n_p\right),$$

which again has no explicit uncertainties.

4 Interval Uncertainty

Our MILP models for sensor placement have the general form min $cx : x \in X$, where c is uncertain. It is often the case that there is uncertainty in the coefficients c that cannot be captured by a probability distribution. For example, these coefficients may be measurements that have limited precision, or they may reflect quantities that have a definite value that is not known precisely. In these contexts, we wish to find a solution with minimum worst-case cost over the interval of uncertainties. We make the general assumption that the total deviation of our uncertainties is zero; that is, the sum of the parameter values is a known constant. In particular, this assumption captures the fact that our estimate of total population is likely to be much better than our estimate of δ_{jp}, the population consuming water at site j in period p.

The next three sections consider three robust optimization models that differ by how c is restricted. Define the *sum-restricted ball* about a vector y by $\mathcal{B}(y,L,U)=\{y' : L\le y'\le U,\ \sum_k y'_k=\sum_k y_k\}$, where we suppose $L\le y\le U$. This is an *interval of uncertainty* about y with one degree of freedom lost. Thus our robust optimization problem is to find

$$\min_{x\in X}\ \max_{c\in\mathcal{B}(\hat{c},L,U)} cx,$$

where X is the feasible region for the MILP defined by constraints (1) and (2).

The coefficients in our MILP models can be rewritten as $c_{ijp}=\alpha_{ip}\theta_{jp}$, where $(i,j,p)\in\mathcal{R}$; $\theta_{jp}=\lambda_{jp}$ in MILP-NC, and $\theta_{jp}=\delta_{jp}$ in MILP-PE. We show that robust optimization is significantly less expensive if the coefficients θ_{jp} are known with certainty, as is the case for MILP-NC, and if the interval of uncertainty is expressed as percentage deviations.

We exploit the particular structure of our robust MILP models in the following analysis. In particular, we consider the linear programming (LP) relaxation of the inner maximization problem, for which we ignore the constraint that x are discrete values. The dual of a maximization LP is another LP with a minimization objective. It is well-known that the minimum value this dual formulation equals the maximum value of the LP itself, and we exploit this fact in our analysis of the inner LP. The following lemma shows the structure of the dual LP relaxation for the inner maximization problem in our robust MILP models.

Lemma 4.1 *Consider the problem* $\max\ yd\ :\ d \in \mathcal{B}(\hat{d}, \underline{d}, \overline{d})$. *The dual linear program uses variables* π *for the constant-sum constraint,* γ *for the lower-bound constraints, and* μ *for the upper-bound constraints, and has the form*

$$\min_{\pi,\gamma,\mu} \left\{ \pi \sum_{r=1}^{N} \hat{d}_r + \mu\overline{d} - \gamma\underline{d}\ :\ \gamma, \mu \geq 0,\ \pi + \mu_r - \gamma_r = y_r \textit{ for } r = 1, \ldots, N \right\}.$$

We use this lemma in the proof of Proposition 4.1, and in solving the more general case in §4.2.

4.1 Unweighted Uncertainty

The **unweighted case with percentage deviations** considers robust models for which $c \in \mathcal{C}(\varepsilon, \hat{c}) \equiv \mathcal{B}(\hat{c},\ (1-\varepsilon)\hat{c},\ (1+\varepsilon)\hat{c})$ for $0 \leq \varepsilon \leq 1$ and $\hat{c} \geq 0$. The following proposition proves that the unweighted case with percentage deviations has a *permanent solution* [11]. That is, the solution for this problem is independent of ε for any fixed $\hat{c}$.

Proposition 4.1 *For each* $\hat{c} \geq 0$*, consider the family of problems for* $\varepsilon \in [0, 1]$*:*

$$P(\varepsilon, \hat{c}) :\ \min_{x \in X}\ \max_{c \in \mathcal{C}(\varepsilon, \hat{c})}\ cx.$$

Then, every optimal solution to $P(0, \hat{c})$ *is optimal for every* $P(\varepsilon, \hat{c}), 0 < \varepsilon \leq 1$.

We omit the formal proof here, but here is the rationale. Choose x^* to minimize $\hat{c}x$, and hence also $P(0, \hat{c})$, the problem with no uncertainty in $\hat{c}$. If there are many sensors, an optimal solution has only a few contamination variables $x^*_r = 1$. Thus, c can only increase the cost *proportionally* (by $\varepsilon \leq 1$). If there are few sensors, an optimal solution has only a few $x^*_r = 0$. The same rationale applies to its complement: $\min cx = S - \max c(1-x)$, where $S = \sum_r c_r = \sum_r \hat{c}_r$ is constant. If x^* is balanced such that $\hat{c}x^* \approxeq \frac{1}{2}S$, this result follows from a simple analysis of the cost of the optimal solutions to $P(0, \hat{c})$ and $P(\varepsilon, \hat{c})$.

Proposition 4.1 demonstrates that the solution to the original MILP is the solution to any robust formulation that allows percentage deviations. Consequently, no additional computational effort is needed to generate a robust solution for these problems.

4.2 Linearly Weighted Uncertainty

The **linearly weighted case** considers robust problems for which $c_r = \eta_r\theta_r$, η_r is known, and $\theta \in \mathcal{B}(\hat{\theta},\, \underline{\theta},\, \overline{\theta})$, where $\underline{\theta} = (1-\varepsilon)\hat{\theta}$ and $\overline{\theta} = (1+\varepsilon)\hat{\theta}$, for $0 < \varepsilon \leq 1$. Thus, our robust problem is

$$\min_{x\in X} \max_{\theta\in\mathcal{B}(\hat{\theta},\underline{\theta},\overline{\theta})} \sum_r \eta_r\theta_r x_r.$$

Let $S = \sum_r \hat{\theta}_r$. We apply Lemma 4.1 to obtain one model with all variables minimizing:

$$\min_{x\in X} \max_{\theta\in\mathcal{B}(\hat{\theta},\underline{\theta},\overline{\theta})} \sum_r \eta_r\theta_r x_r = \min\,\{\pi S + \mu\overline{\theta} - \gamma\underline{\theta} \,:\, x \in X,\ \gamma,\mu \geq 0,\ \pi + \mu_r - \gamma_r = \eta_r x_r\}.$$

Now the sum of the equality constraints in this new formulation, weighted by $\hat{\theta}_r$, gives us $\pi S + \hat{\theta}(\mu-\gamma) = \sum_r \hat{\theta}_r\eta_r x_r$. We thus rewrite this formulation as

$$\min \sum_r \hat{\theta}_r\eta_r x_r + (\overline{\theta}-\hat{\theta})\mu + (\hat{\theta}-\underline{\theta})\gamma \,:\, x \in X,\ \gamma,\mu \geq 0,\ \pi + \mu_r - \gamma_r = \eta_r x_r. \quad (5)$$

We conclude that solving linearly weighted robust formulations simply requires the solution of an augmented MILP formulation, which includes an extended objective and some additional side-constraints on dual variables from the maximization subproblem.

Alternatively, instead of treating this as one minimization, we can decompose it and solve the inner maximization problem to obtain $\theta(x)$ for each x in the outer minimization. This can be done simply by sorting the coefficients $\{\eta_r x_r\}$ so that $\eta_1 x_1$ has the largest value, $\eta_2 x_2$ the next largest and so on. We then transfer weight from the θ_i with the lightest coefficients to the θ_i with the heaviest coefficients, respecting lower and upper bounds. The worst-case value of θ has the form $\theta_r = (1+\varepsilon)\hat{\theta}_r = \overline{\theta}_r$ for $r < k$, $\theta_r = (1-\varepsilon)\hat{\theta}_r = \underline{\theta}_r$ for $r > k$, and $\underline{\theta}_k \leq \theta_k = \sum_r \hat{\theta} - \sum_{r\neq k}\theta_r \leq \overline{\theta}_k$. We can also take advantage of sorting $\{\eta_r x_r\}$ once by putting indices for which $x_r = 0$ last for a particular x. Unlike the unweighted case, this is algorithmic because the result depends upon η (as well as x). Because the inner maximization problem has time complexity $O(|\mathcal{R}|\ln|\mathcal{R}|)$, this decomposition approach may be computationally more efficient than the integrated formulation in Equation (5). (See [2, 7] for descriptions of using decomposition for algorithm design in integer programming.)

4.3 Bilinear Weighted Uncertainty

Finally, we consider the **bilinear weighted case**, where $c_r = \eta_r\theta_r$, $\eta_r \in \mathcal{B}(\hat{\eta},\, \underline{\eta},\, \overline{\eta})$, and $\theta_r \in \mathcal{B}(\hat{\theta},\, \underline{\theta},\, \overline{\theta})$. Our robust optimization problem is

$$\min_{x\in X} \max_{\substack{\eta\in\mathcal{B}(\hat{\eta},\,\underline{\eta},\,\overline{\eta})\\ \theta\in\mathcal{B}(\hat{\theta},\underline{\theta},\overline{\theta})}} \textstyle\sum_r \eta_r\theta_r x_r.$$

Maximizing a bilinear form is a difficult problem, but there are several approaches that make effective use of simple inequalities (e.g., Ryoo and Sahinidis [9]). We sketch here a branch-and-bound approach that we are pursuing to obtain a near-global solution.

Consider for the moment the bilinear subproblem. Let $\sigma(x) = \{r : x_r = 1\}$ and let $\tilde{y}$ denote y restricted to $\sigma(x)$. We can apply McCormick's linear bounds [6, 9],

$$\max\{\overline{\tilde\eta}\tilde\theta + \overline{\tilde\theta}\tilde\eta - \overline{\tilde\eta}\,\overline{\tilde\theta},\ \underline{\tilde\eta}\tilde\theta + \underline{\tilde\theta}\tilde\eta - \underline{\tilde\eta}\,\underline{\tilde\theta}\} \le \tilde\eta\tilde\theta \le \max\{\tilde{\overline{\eta}}\tilde\theta + \underline{\tilde\theta}\tilde\eta - \underline{\tilde\eta}\,\tilde{\overline{\tilde\theta}},\ \underline{\tilde\eta}\tilde\theta + \tilde{\overline{\tilde\theta}}\tilde\eta - \underline{\tilde\eta}\,\tilde{\overline{\tilde\theta}}\},$$

to bound the bilinear objective.

Given this linearization of the bilinear subproblem, we can apply the same duality analysis used for the linearly weighted case. However, the augmented MILP now provides only a bound on the value of a robust solution. To refine this bound, we consider a branch-and-bound process that branches to fix the sensor-placement variables s_e (and hence the decision variables x_r) and to subdivide the values of $\underline{\tilde\theta}$, $\overline{\tilde\theta}$, $\underline{\tilde\eta}$, and $\overline{\tilde\eta}$ (to improve the bound on the bilinear subproblem).

5 Conclusions

There are many possible formal objectives for sensor placement that reflect various costs and risks of an attack on a network [10]. Previous work has considered problem formulations that minimize the volume of water consumed before detection [3], minimize the time to detection [5], and minimize the population exposed to contaminants before detection [1].

Our analysis of randomized uncertainty and intervals of uncertainty applies to many of the MILP formulations described by Watson et al [10]. We have shown that many robust analyses simply require a reformulation or reinterpretation of the initial MILP problem. Consequently, we expect that these analyses will be tractable in practical applications. We have also demonstrated that in some special circumstances the solution to the initial MILP is in fact a robust solution.

However, finding robust formulations with multiple sources of uncertainty remains a significant challenge. Although we have outlined a strategy for dealing with bilinear models with uncertain intervals, the practical application of this strategy remains to be demonstrated. For example, the strategy that we have outlined requires a customized branching strategy. It cannot be directly solved with commercial MILP solvers.

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.

References

[1] J. Berry, L. Fleischer, W.E. Hart, and C.A. Phillips. Sensor placement in municipal water networks. In *Proceedings of the World Water and Environmental Resources Conference*, 2003.

[2] H.J. Greenberg. *Mathematical Programming Glossary.* World Wide Web, http://www.cudenver.edu/~hgreenbe/glossary/, 1996–2004.

[3] A. Kessler, A. Ostfeld, and G. Sinai. Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management*, pages 192–198, 1998.

[4] P. Kouvelis and G. Yu. *Robust Discrete Optimization and Its Applications.* Kluwer Academic Publishers, Norwell, MA, 1996.

[5] A. Kumar, M.L. Kansal, and G. Arora. Discussion of detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management*, pages 308–310, 1999.

[6] G.P. McCormick. *Nonlinear Programming: Theory, Algorithms, and Applications.* Wiley, New York, NY, 1983.

[7] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization.* John Wiley & Sons, New York, NY, 1988.

[8] L. A. Rossman. The EPANET programmer's toolkit for analysis of water distribution systems. In *Proceedings of the Annual Water Resources Planning and Management Conference*, 1999.

[9] H.S. Ryoo and N.V. Sahinidis. Global optimization of nonconvex NLPs and MINLPs with applications in process design. *Computers Chemical Engineering*, 19(5), 1995.

[10] J-P. Watson, H.J. Greenberg, and W.E. Hart. A multiple-objective analysis of sensor placement optimization in water networks. In *this proceedings*, 2004.

[11] H. Yaman, O.E. Karasan, and M.C. Pinar. The robust minimum spanning tree problem with interval data. *Operations Research Letters*, 29:31–40, 2001.