European Symposium on Computer Aided Process Engineering – 15
L. Puigjaner and A. Espuña (Editors)

Water Networks Security: A Two-Stage Mixed-Integer Stochastic Program for Sensor Placement Under Uncertainty

Sergio Frausto-Hernández[a], Urmila M. Diwekar[b], Salvador Hernández-Castro[c] and Vicente Rico-Ramírez[a*]

[a]Instituto Tecnológico de Celaya, Departamento de Ingeniería Química
Ave. Tecnológico y García Cubas S/N, Celaya, Gto., C. P. 38010, México

[b]University of Illinois at Chicago, Departments of Bio & Chemical Engineering
851 S. Morgan St. (M/C 063), Chicago, IL 60607, USA

[c]Universidad de Guanajuato, Facultad de Química
Col. Noria Alta S/N, Guanajuato, Gto., C. P. 36050, México

Abstract

This work describes a stochastic approach for the optimal placement of sensors in municipal water networks to detect maliciously injected contaminants. The model minimizes the expected fraction of the population at risk. Our work explicitly includes uncertainties in attack risks and population density, so that the resulting problem involves optimization under uncertainty. In our formulation, we include the number of sensors as first stage decision variables of a two-stage mixed-integer stochastic linear problem where the costs of sensors are included in the objective function. Since the model is integer in the first-stage, a generalized framework based on the stochastic decomposition algorithm allows us to solve the problem in a reasonable computational time. The paper describes the mixed-integer stochastic model and the algorithmic framework, and compares the deterministic and stochastic optimal solutions. The network used as our case-study has been derived through the water network simulator EPANET; four acyclic water flow patterns are considered. Results show a significant effect of uncertainty in sensor placement and total cost.

Keywords: Mixed-integer stochastic programming, stochastic decomposition, water networks security three to five keywords

1. Introduction

Recently, Berry *et al.* (2003) presented an integer programming (IP) formulation for optimal placement of sensors in municipal water networks to quickly detect accidental or malicious contamination of the system. Such a formulation seeks to minimize the expected fraction of the population at risk. An attack is modelled as the release of a large volume of harmful contaminant at a single point in the network with a single

* Author to whom correspondence should be addressed: vicente@iqcelaya.itc.mx

injection. In general, it is difficult to know a priori where an attack will occur, so a compromise solution across a set of weighted attack scenarios is generated. For each flow pattern, each node is weighted by the number of people potentially consuming water at that point. The resulting IP formulation assumes constant probabilities for attack risks and population densities, making the problem a deterministic IP. The results of the analysis show that the optimal configuration is a strong function of the fixed number of sensors used in the formulation. The model does not include the cost and type of the sensors, which could be major factors in the solution achieved. Further, although uncertainties in the demand and variability in the population density can impact the solution, they are not formally considered.

Our work proposes extensions to such a model to explicitly include uncertainties in attack risks and population density. This changes the problem to an optimization under uncertainty problem. In our formulation, we include the number of sensors as first stage decision variables of a two-stage mixed-integer stochastic problem, where the costs of sensors are included in the objective function.

2. A Two-Stage Stochastic Mixed-Integer Programming Model

Although the fundamentals of this new model are very similar to those proposed by Berry *et al.* (2003), our main contribution is the reformulation of the problem as a first-stage integer, two-stage stochastic mixed-integer linear problem. That is significant for two reasons: *i)* first, the uncertainties are formally incorporated into the model, and *ii)* the basic stochastic decomposition (SD) algorithm (Higle and Sen, 1991) can be used to provide the optimal solution to the problem.

The network is represented as a directed graph G = (V,E), where E is the set of pipes (E=$e_1,\ldots,e_m$) and V is a set of nodes (V=$v_1,\ldots,v_n$) where the pipes meet. Each pipe connects two nodes and is usually denoted as (v_i, v_j). Also, it is assumed that the network is planar. The analysis is performed under a number of water flow patterns, where the direction of the flow in each pipe is known. The parameters f_{ijp} describe the pattern; f_{ijp} is equal to one if there is a positive flow along the directed pipe e=(v_i, v_j) during flow pattern p, and is equal to zero otherwise. The probability of attack for node v_i during flow pattern p is represented as ω_{ip}; we have used $\sum \omega_{ip} = 1$. δ_{ip} is the population density at node i while flow p is active. Finally, X_{max} is the maximum number of sensors, Cij is the sensor cost, and S is a cost associated to each member of the population which suffers from the attack. Cij and S can both be interpreted as weighting factors in the objective function.

The first stage (integer stage) decisions, x_{ij}, define the sensor placement, and the second stage decisions, y_{ipj}, define the propagation of an injected contaminant. A sensor on pipe (v_i, v_j), x_{ij}, detects contaminants moving on either direction. y_{ipj} is equal to one if node v_j is contaminated by an attack at node v_i during pattern p, and zero otherwise. Equation (1) and Equation (2) provide the structure of the stages. The first stage tries to minimize the cost of the sensors as well as the expected value ($E\omega$) of the cost associated to the population suffering the attack (computed through the objective of stage two); the constraints of the first stage impose the maximum number of sensors to be placed in the network. The constraints in the second stage propagate contamination if there is no sensor which prevents it.

First stage (integer):

$$Min \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} x_{ij} + E_{\omega}\left[Q(x,\omega)\right]$$

$$\begin{aligned} & s.t. \\ & x_{ij} - x_{ji} = 0 \quad \forall \quad i = 1,\ldots,n-1 \quad i < j \\ & \sum\nolimits_{(i,j)\ j \in E,\, i<j} x_{ij} - X_{\max} \leq 0 \\ & \quad x_{ij} \in \{0,1\} \quad \forall \quad i,j \in E \end{aligned} \tag{1}$$

Second stage:

$$Q(x,\omega) = \min \sum_{i=1}^{n} \sum_{p=1}^{P} \sum_{j=1}^{n} S y_{ipj} \omega_{ip} \delta_{jp}$$

$$\begin{aligned} & s.t. \\ & y_{ipi} = 1 \quad \forall \quad i = 1,\ldots,n \\ & \qquad\qquad \forall \quad p = 1,\ldots,P \\ & y_{ipk} - y_{ipj} \leq x_{kj} \quad \forall \quad (k,j) \in E \quad s.t. \quad f_{kjp} = 1 \end{aligned} \tag{2}$$

2.1 The solution approach

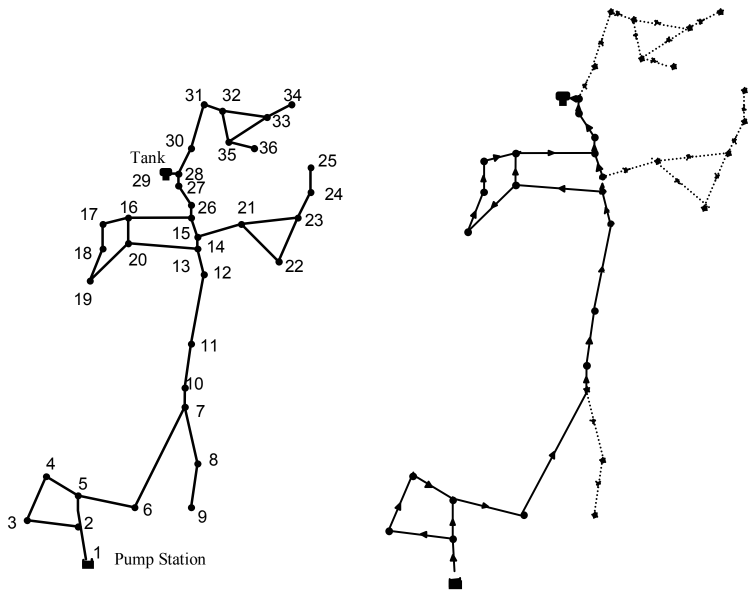
Because the problem is integer in the first stage, the solution can be achieved with the same basic steps of the SD algorithm developed for linear problems; the only difference is that, instead of a linear problem, now a mixed-integer linear programming problem has to be solved in order to find the first stage decisions in each of the iterations. Our computational implementation of the algorithm involves a framework that integrates the Hammersley sequence sampling technique (Kalagnanam and Diwekar, 1997) coded in FORTRAN, the GAMS modelling environment (Brooke et al., 1998), and a C++ manager program. The manager program generates the appropriate MILP and LP problems for each of the SD iterations, transfers the control of execution and verifies the convergence of the algorithm. The steps of the SD algorithm can be found in Ponce-Ortega *et al.* (2004); a brief summary of such steps is given as follows:

- Assume the initial values of the first stage decision variables, set $\nu=0$.
- Set the iteration counter to $\nu = \nu + 1$ and sample to generate an observation u^{ν} independent of any previous observation.
- Determine the coefficients of a piecewise linear approximation to $Q(x)$: Solve the dual program of the second stage problem to get the coefficients of the optimality cut and update the coefficients of previous cuts.
- Solve the first stage problem (MILP) after the addition of the optimality cut to obtain $x^{\nu+1}$. Go to the sampling step and repeat the procedure.

The algorithm stops if the change in the objective function is sufficiently small.

3. Illustrative Example and Results

This section presents the solution of example network number 2 of EPANET 1.0 (Rossman, 1999). The water network is shown in Figure 1a. It has 36 nodes and 40 pipes with 1 pump station and 1 storage tank. 4 time periods (four water flow patterns) of 6 hours are considered in a 24 hour total period. Figure 1b shows one of the flow patterns used. Also, the nodes are distributed in 4 fictitious zones: Pump station, residential neighbourhood, business district and industrial district (see Table 1). The maximum number of sensors is 7. All nodes in the same zone have the same probability of being attacked. We use normal distribution for the probability of attack. For the density population we use normal and triangular distributions with mean equal to 500. After sampling, the density population and probability of attack are normalized so that the total population is always 18000 and the sum of the probabilities of attack is 1. Four possible scenarios are considered, assuming different zone with higher probability of attack in each of them.



(a) *(b)*

Figure 1. Case study. (a) Municipal water network. (b) Example of a flow pattern, —— 100 gpm .... 25 gpm

Table 1. Distribution of nodes of the network

Zone	Nodes	Total
Pump Station	1	1
Industrial District	3,10,11,12,16,17,18,20,21,22,26,30	12
Residential Neighbourhood	2,4,5,6,8,9,19,23,27,28,31,32	12
Business District	7,13,14,15,24,25,29,33,34,35,36	11

Table 2 shows the probability data for each of the four scenarios. In addition, two costs for sensors are used: low (15 000 000) and high (45 000 000). Finally, the cost assumed (S) for each affected person is 30000. Such values are assumed as constant. You can also interpret those constant values as weights in the objective functions. The number of samples and iterations of the SD algorithm was 400. Table 3 shows the optimal stochastic solutions for the scenario where the residential neighbourhood is under risk. The deterministic solution (mean values used for stochastic parameters; for population density the mean value is 500) is also presented.

The solutions for the other 3 zones present similar behaviours. There are meaningful differences among the cases of high and low sensor costs, and even more significant between the stochastic and deterministic cases. On the contrary, the effect of the distribution used for the population density seems to be less important. In general, one could expect that the fraction of the population at risk decreases with the number of sensors; however, because of the compromise between the cost of the sensors and the associated cost to each affected person, the optimal solution does not always include the maximum number of sensors. Specifically, that is true for the cases of high sensor cost.

Table 2. Probability of attack used for each zone and scenario

Scenario	Probability of Attack (%)			
	Pump Station	Industrial	Residential	Business
1	65	12	12	11
2	1	76	12	11
3	1	12	76	11
4	1	12	12	75

Table 3. Results for the residential neighbourhood

Population Density	Sensor Cost	Variable	Stochastic Solution	Deterministic Solution
Triangular	Low	Sensors allocated	X_{6-7},X_{12-13},X_{14-15}, X_{28-30}, X_{31-32}, X_{32-33}, X_{35-36}	X_{6-7},X_{12-13},X_{27-28},X_{31-32}
		Population affected (%)	13.64	30.71
		Total Cost	1.7866887E+8	2.2585000E+8
	High	Sensors allocated	X_{6-7},X_{12-13},X_{27-28},X_{28-30}	X_{6-7},X_{28-30}
		Population affected (%)	26.19	42.61
		Total Cost	3.2145500E+8	3.2010000E+8
Normal	Low	Sensors allocated	X_{5-6},X_{6-7},X_{7-8},X_{11-12},X_{19-18},X_{31-32}, X_{35-36}	X_{6-7},X_{12-13},X_{27-28},X_{31-32}
		Population affected (%)	13.48	30.71
		Total Cost	1.7781497E+8	2.2585000E+8
	High	Sensors allocated	X_{6-7},X_{28-30},X_{31-32}	X_{6-7},X_{28-30}
		Population affected (%)	35.19	42.61
		Total Cost	3.2507400E+8	3.2010000E+8

4. Analysis of Results and Conclusions

Table 4 shows the values of the stochastic solution (VSS) for each of the cases considered. VSS represents the difference between the objective functions of the stochastic and deterministic optimal solutions. We can observe values as high as 96% as a consequence of the major impact of uncertainty in the solution. To conclude, this paper proposes a two-stage mixed-integer stochastic linear programming approach for the optimal placement of sensors in a municipal water network. Three main results can be summarized: *i)* First, this approach not only allows the incorporation of uncertainties to the problem but also provides an efficient solution strategy through the SD algorithm, *ii)* The results of the illustrative example (VSS, optimal placement and effect to population) reveal the significant effect of the uncertainties in the optimal solution of the problem and *iii)* With respect to the parameters under investigation, the sensor cost seems to have a more important effect than the population density; that applies for both the stochastic and the deterministic case.

Table 4. Value of stochastic solution for each zone and scenario

Population Density	Sensor Cost	Value of Stochastic Solution (%)			
		Pump Station	Industrial	Residential	Business
Triangular	Low	91.12%	19.70%	26.40%	5.68%
	High	4.25%	0.18%	0.42%	1.44%
Normal	Low	96.21%	17.57%	27.01%	7.64%
	High	4.04%	0.94%	1.53%	0.96%

References

Berry J., L. Fletcher, W. Hart, and C. Philips, 2003, Optimal Sensor Placement in Municipal Water Networks, in the Proceedings of the World Water and Environmental Resources Conference, Philadelphia PA, USA.

Brooke, A., D. Kendrick, A. Meeraus and R. Raman, 1998, GAMS-A User Guide. GAMS Development Corporation, Washington, DC, USA.

Higle, J. L. and S. Sen, 1991, Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse, Mathematics of Operations Research, 16, 650-669.

Kalagnanam J. and U. M. Diwekar, 1997, An Efficient Sampling Technique for Off-line Quality Control, Technometrics, 39(3), 308.

Ponce-Ortega, J. M., V. Rico-Ramírez, S. Hernandez-Castro and U. M. Diwekar, 2004, Improving Convergence of the Stochastic Decomposition Algorithm by Using an Efficient Sampling Technique, Comp. Chem. Eng., 28, 767-773.

Rossman L. A., 1999, The EPANET Programmer’s Toolkit for Analysis of Water Distribution Systems, in the proceedings of the Annual Water Resources Planning and Management Conference.

Acknowledgements

Vicente Rico-Ramírez would like to thank the financial support received trough the collaborative research grant NSF-CONACYT 35974-A (Mexico).