

Water networks security: A two-stage mixed-integer stochastic program for sensor placement under uncertainty[☆]

Vicente Rico-Ramirez^{a,*}, Sergio Frausto-Hernandez^a,
Urmila M. Diwekar^b, Salvador Hernandez-Castro^c

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

^b Vishwamitra Research Institute, 34 N. Cass Avenue, Westmont, IL 60559, USA

^c Universidad de Guanajuato, Facultad de Química, Col. Noria Alta S/N, Guanajuato, Gto., C.P. 36050, Mexico

Received 15 November 2005; received in revised form 7 August 2006; accepted 17 August 2006

Available online 27 September 2006

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 and the cost of the sensors. Our work explicitly includes uncertainties in the attack risk and population density, so that the resulting problem involves optimization under uncertainty. In our formulation, we include the location of a number of sensors as first stage decision variables of a two-stage mixed-integer stochastic linear problem; the second stage evaluates the population at risk for the scenario obtained in the first stage and that information is then used to modify the first stage decisions for the next iteration. 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 1.0; four acyclic water flow patterns are considered. Results show a significant effect of uncertainty in sensor placement and total cost. © 2006 Elsevier Ltd. All rights reserved.

Keywords: Mixed-integer stochastic programming; Stochastic decomposition; Water networks security

1. Introduction

Stochastic programming has become a common tool for decision making on process design, production planning, location, scheduling and finance. Hence, literature in the last decade reports many papers that have addressed the issue of solving optimization problems under uncertainty; Sahinidis (2004) provided an excellent review on the various applications and opportunities of this area.

Several modeling approaches based on a variety of philosophies have been proposed. Among the applications on process systems engineering, most of the methods can be classified into one of the following two categories: (1) integration methods and (2) sampling methods. Acevedo and Pistikopoulos (1998) compared the Gaussian Quadrature formula (integration method)

with Monte Carlo sampling for sample average approximation. Wei and Realf (2004) also presented a review on this issue and developed two sample average approximation methods (optimality gap and confidence level) for solving convex stochastic MINLPs. Similarly, Novak Pintaric and Kravanja (2004) presented a sequential two-stage strategy for the stochastic synthesis of chemical processes, in which flexibility and static operability are simultaneously taken into account.

In addition, a number of contributions have recently been reported to approach the solution of deterministic and stochastic problems of optimal sensor placement. For instance, Bagajewicz, Fuxman, and Uribe (2004) developed a mixed-integer linear programming formulation for the design of sensor networks for simultaneous process monitoring and fault detection/resolution.

This paper describes a stochastic approach for the optimal placement of sensors in municipal water networks to detect maliciously injected contaminants. This approach represents yet another application of stochastic programming, and it is relevant in the context of homeland security efforts.

[☆] A preliminary version of this paper was presented at the ESCAPE-15 Conference.

* Corresponding author. Tel.: +52 461 6117575; fax: +52 461 6117744.
E-mail address: vicente@iqcelaya.itc.mx (V. Rico-Ramirez).

1.1. Optimal placement of sensors in municipal water networks: a stochastic approach

Recently, Berry, Fletcher, Hart, and Philips (2003) presented a mixed-integer linear programming (MIP) 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 modeled as the release of a large volume of harmful contaminant at a single point in the network with a single 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 (population density) potentially consuming water at that point. The resulting MIP formulation assumes constant probabilities for attack risks and population densities, making the problem a deterministic MIP. Nevertheless, the results of the analysis show that the optimal configuration is a strong function of the fixed number of sensors used in the formulation. Also, 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 (stochastic programming problem). In our formulation, we include the number and location of the 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. Hence, the stochastic mixed-integer linear program is solved by applying the stochastic decomposition (SD) algorithm of Higle and Sen (1991).

1.2. Two-stage stochastic programming and the SD algorithm

The basic SD algorithm (Higle & Sen, 1991) was developed to solve stochastic linear problems with recourse (*SLPwR*). The main class of *SLPwR* problems involves two stages. In the first stage, the choice of the (first stage) decision variables x is made. In the second stage, following the observation of the values of the uncertain parameters, ω , and the evaluation of the objective function, a corrective action (represented by the second stage decisions), y , is suggested. The standard mathematical form of a *SLPwR* problem is given by Eqs. (1) and (2). Eq. (1) is referred as the first stage problem:

$$\begin{aligned} & \text{Min } c^T x + Q(x) \\ & \text{Subject to :} \\ & Ax = b \\ & x \geq 0 \end{aligned} \quad (1)$$

where A is a coefficient matrix, c the coefficient vector and $Q(x)$ is the recourse function defined by $Q(x) = E[Q(x, \omega)]$. E is the

expectation operator and $Q(x, \omega)$ is obtained from the second stage problem, Eq. (2):

$$\begin{aligned} & Q(x, \omega) = \text{Min } q^T y \\ & \text{Subject to :} \\ & W(\omega)y = h(\omega) - T(\omega)x \\ & y \geq 0 \end{aligned} \quad (2)$$

In Eq. (2), q is a coefficient vector and W , h and T are coefficient matrices which in principle might depend on the random variables ω . The theory and steps of the basic SD algorithm can be found in Higle and Sen (1996) and Ponce-Ortega, Rico-Ramírez, Hernández-Castro, and Diwekar (2004); a step-by-step illustrative example can also be reviewed in Rico-Ramírez (2002). Seeking completeness, a summary is provided here.

- Step 0: Set the iteration counter $v=0$, $V_0 = \{\emptyset\}$ and $\theta^v = -\infty$. The initial values of the first decision variables, x^1 , are assumed as given.
- Step 1: Set $v = v + 1$ and sample to generate an observation ω^v independent of any previous observation.
- Step 2: Determine the coefficients of a piecewise linear approximation to the recourse function, $Q(x)$:

- (a) Solve the dual program of the second stage problem

$$\text{Max } \pi^T (h_v - T_v x^v)$$

$$\text{subject to : } \pi^T W \leq q$$

$$\pi \geq 0$$

to find the values of the vector of the second stage dual variables (Lagrange multipliers of primal restrictions) π , π_v^v , and make $V_v = V_{v-1} \cup \pi_v^v$.

- (b) Get the coefficients of the optimality cut:

$$\alpha_v^v = \frac{1}{v} \sum_{k=1}^v (\pi_k^v)^T h_k$$

$$\beta_v^v = \frac{1}{v} \sum_{k=1}^v (\pi_k^v)^T T_k$$

where π_k^v is the solution to the problem (for all $k|k \neq v$):

$$\text{Max } \pi^T (h_k - T_k x^v)$$

$$\text{subject to : } \pi \in V_v$$

Observe that the solution vector to this problem can only be one of the vectors already included in the set of solutions, V_v (the solution set is restricted to decrease effort).

- (c) Update the coefficients of previous cuts:

$$\alpha_k^v = \frac{v-1}{v} \alpha_k^{v-1}, \quad k = 1, \dots, v-1$$

$$\beta_k^v = \frac{v-1}{v} \beta_k^{v-1}, \quad k = 1, \dots, v-1.$$

- Step 3: Solve the first stage problem after the addition of the optimality cut:

$$\text{Min } c^T x + \theta^v$$

$$\text{subject to : } Ax = b$$

$$\theta^v + \beta_k^v x \geq \alpha_k^v, \quad k = 1, \dots, v$$

$$x \geq 0, \quad \theta^v \in \Re$$

to obtain x^{v+1} . Go to Step 1.

- There are several stopping criteria for the algorithm; in the most simple case, the algorithm stops if:
 - (a) change in the objective function is small;
 - (b) no new dual vectors are added to the set of dual solutions, V .

One can observe that the first and the second stages of the problem are both linear programs. Also, as can be reviewed in the text by Birge and Louveaux (1997), if the first stage becomes a mixed-integer linear program ($x \in \{0,1\}$) as opposed to $x \geq 0$), the basic steps of the algorithm hold. However, if the second stage becomes a mixed-integer program ($y \in \{0,1\}$) as opposed to $y \geq 0$), then the derivation of the optimality cut (Step 2) involves a branch and bound strategy where, besides seeking feasibility and optimality, enforcing integrality of the solution greatly complicates the algorithm. Fortunately, the resulting stochastic program of this work is first stage integer.

2. Water networks security model: MIP formulation versus SMILPwR formulation

This section describes the MIP approach to represent the water network security problem (Berry et al., 2003) and the reformulation of such model as a stochastic mixed-integer linear programming problem with recourse (SMILPwR). The computational implementation of the solution algorithm is also provided.

2.1. MIP formulation of Berry et al. (2003)

The municipal water network is represented as a directed graph $G = (V, E)$, where E is the set of edges or pipes ($E = e_1, \dots, e_m$) and V is a set of vertices or nodes ($V = v_1, \dots, v_n$) where the pipes meet. Vertices represent sources (reservoir, tanks) where water is introduced or sinks (demand points) where water is consumed. Fig. 1 shows a typical municipal water network configuration. Each pipe connects two nodes and is usually denoted as (v_i, v_j) . 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 1 if there is a positive flow along the directed pipe $e = (v_i, v_j)$ during flow pattern p , and it is equal to 0 otherwise. The probability of attack for node v_i during flow pattern p is represented as η_{ip} ; such that the summation over i of all of the values of η_{ip} is equal to 1. δ_{ip} is the population density at node v_i while flow p is active; understanding the value of the population density as the number of individuals consuming water at each particular node. The decision variables, x_{ij} , define the sensor placement, and the derived variables, 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 1 if node v_j is contaminated by an attack at node v_i during pattern p , and 0 otherwise. As mentioned before, an attack is the single injection of a large volume of harmful contaminant at a single point, but then all points downstream can be contaminated. Propagation from v_k to v_j occurs if v_k is contaminated, there is positive flow from

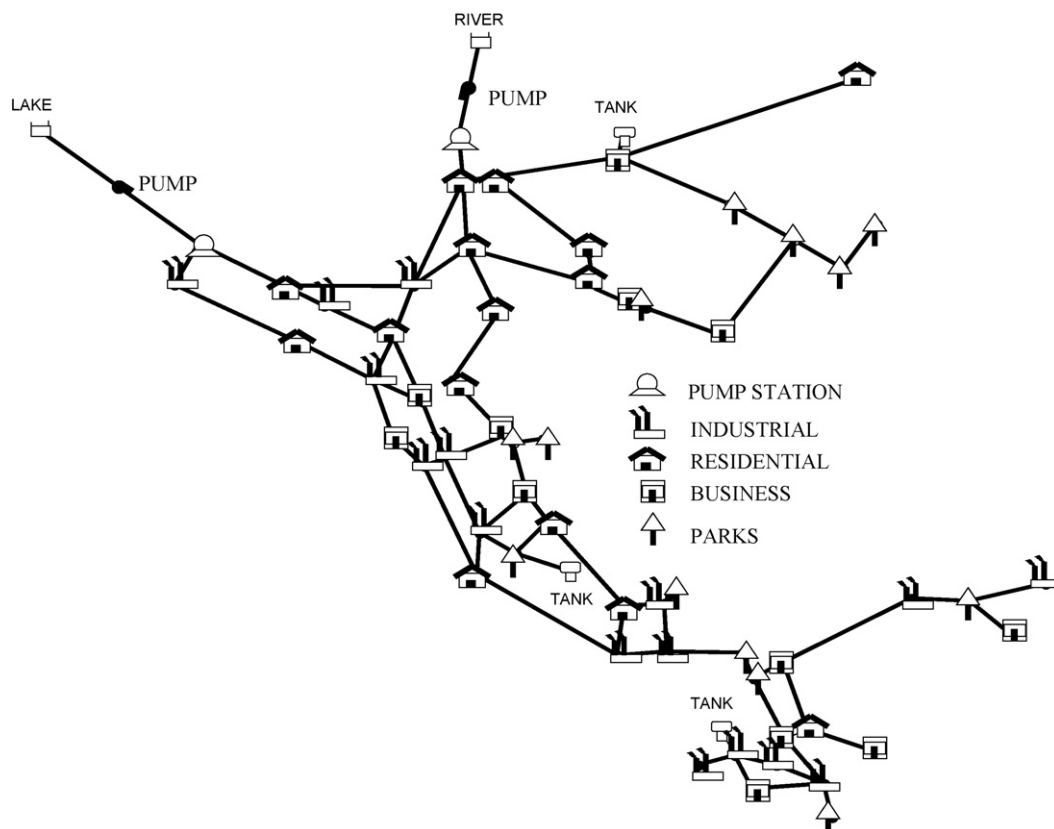


Fig. 1. Municipal water network.

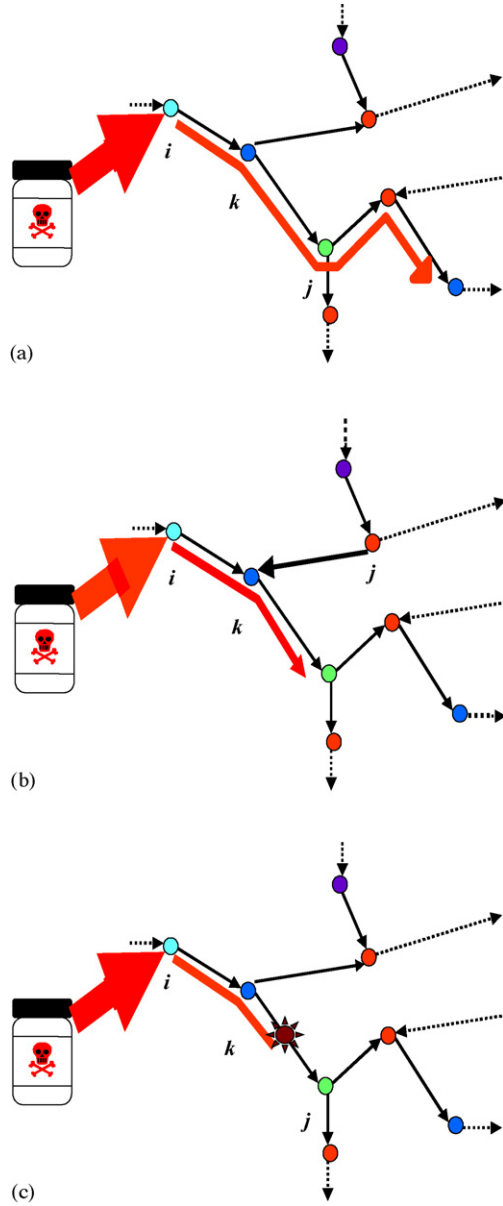


Fig. 2. Modeling the propagation of an injected contaminant. (a) Propagation from node k to node j occurs (positive flow). (b) Propagation from node k to node j does not occur. (c) Sensor prevents propagation from node k to node j .

v_k to v_j and there is no sensor in between. Fig. 2 illustrates the model of the contaminant propagation. Finally, X_{\max} is the maximum number of sensors. The MIP formulation of Berry et al. is given by Eq. (3):

$$\begin{aligned} & \text{Min} \sum_{i=1}^n \sum_{p=1}^P \sum_{j=1}^n \eta_{ip} y_{ipj} \delta_{jp} \\ & \text{s.t.} \\ & y_{ipi} = 1 \quad \forall i = 1, \dots, n, \quad p = 1 \dots P \\ & x_{ij} = x_{ji} \quad \forall i = 1, \dots, n-1, \quad i < j \\ & y_{ipj} \geq y_{ipk} - x_{kj} \quad \forall (k, j) \in E, \quad \text{s.t.} \quad f_{kjp} = 1 \\ & \sum_{(i,j) \in E, i < j} x_{ij} - X_{\max} \leq 0 \\ & x_{ij} \in \{0, 1\} \quad \forall (i, j) \in E \end{aligned} \quad (3)$$

The first constraint of the model ensures contamination of a node when it is directly attacked. The second constraint implies that a sensor on a pipe will detect the contaminant despite the flow direction. The third one is the contaminant propagation constraint. Finally, the last constraint enforces the total number of sensors to be less or equal to the maximum number X_{\max} . In Eq. (3), η_{ip} and δ_{ip} are fixed given parameters on a particular scenario; therefore, the model is a deterministic MIP.

2.2. Reformulation as a SMILPwR

Although the fundamentals of our model are the same as 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 algorithm (Higle & Sen, 1991) can be used to provide the optimal solution to the problem.

Our work explicitly includes uncertainties in risk of attack, η_{ip} , and population density, δ_{ip} . An assumed probability distribution function is assigned to each of those parameters. That consideration makes the problem a stochastic program. Notice that the symbol ω was used in Section 1 to represent the uncertain parameters. Hence, in our SMILPwR formulation, the vector ω consists of all of the values of both parameters, risk of attack, η , and population density, δ , and such values change for each node i and flow pattern p . Uncertainty in population density is assumed because the model consider a 24 h cycle; although during that period the total population remains basically the same, the transit of people among different nodes makes it difficult to predict the exact number of individuals at each particular node.

Also, we proposed the separation of the model constraints into two stages, following the concept of recourse. The first stage includes the decisions about the placement of the sensor through the binary variables x_{ij} . Also, the costs of the sensors (and sensor placement), C_{ij} , are included in the first stage objective function. Hence, by including such costs, the optimal solution may not use the total number of sensors and, therefore, the number of sensors in the configuration also becomes an optimization variable.

The derived variables (which represent the propagation of contaminants), y_{ipj} , are then considered as the second stage decisions. The recourse function (expected value of the second stage objective function) is almost identical to the objective function of the original MIP model. It includes the calculation of the population at risk under an attack. However, a factor S has been added to the calculation. S is a cost associated per each member of the population which suffers from the attack. So, the first stage tries to minimize the cost of the sensors as well as the expected value (E_{ω}) of the cost associated to the population suffering the attack (computed through the objective of stage two); notice then that both terms are calculated in a monetary basis. Nevertheless, C_{ij} and S might simply be interpreted as

weighting (or scaling) factors from the numerical point of view.

In summary, the two-stage approach works as follows: (i) the first stage decides about the number and the location of the sensors; (ii) given the first stage decisions, for some sampled values of the uncertain parameters, the second stage evaluates the population at risk for that scenario; (iii) the second stage objective function calculation is then used (through an optimality cut) to modify the first stage decisions for the next iteration; that is precisely the idea of recourse.

Eqs. (4) and (5) show the constraints of the two stages. The constraints of the first stage impose a number of sensors (less or equal to the maximum) to be placed in the network. The constraints in the second stage propagate contamination if there is no sensor that prevents it and evaluate the population at risk under an attack. Notice that the first stage is an integer programming problem.

First stage (integer):

$$\begin{aligned} \text{Min } & \sum_{i=1}^n \sum_{j=1}^n C_{ij} x_{ij} + E_{\omega}[Q(x, \omega)] \\ \text{s.t. } & x_{ij} - x_{ji} = 0 \quad \forall i = 1, \dots, n-1, \quad i < j \\ & \sum_{(i,j) \in E, i < j} x_{ij} - X_{\max} \leq 0 \\ & x_{ij} \in \{0, 1\} \quad \forall i, j \in E \end{aligned} \tag{4}$$

Second stage:

$$\begin{aligned} Q(x, \omega) = & \min \sum_{i=1}^n \sum_{p=1}^P \sum_{j=1}^n S y_{ipj} \eta_{ip} \delta_{jp} \\ \text{s.t. } & y_{ipi} = 1 \quad \forall i = 1, \dots, n \\ & \forall p = 1, \dots, P \\ & y_{ipk} - y_{ipj} \leq x_{kj} \quad \forall (k, j) \in E \quad \text{s.t. } f_{kjp} = 1 \\ & y_{ipj} \geq 0, \quad \forall i = 1, \dots, n, \quad p = 1, \dots, P, \quad j = 1, \dots, n \end{aligned} \tag{5}$$

2.3. The solution approach for the SMILPwR model

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 an integer linear programming problem has to be solved in order to find the first stage decisions in each of the iterations. Fig. 3 shows a flowsheet of the solution steps. Our computational implementation of the algorithm involves a framework that integrates the Hammersley sequence sampling technique (Kalagnanam & Diwekar, 1997) coded in FORTRAN, the GAMS modeling environment (Brooke, Kendrick, Meeraus, & Raman, 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. Fig. 4 represents the computational implementation of the algorithm.

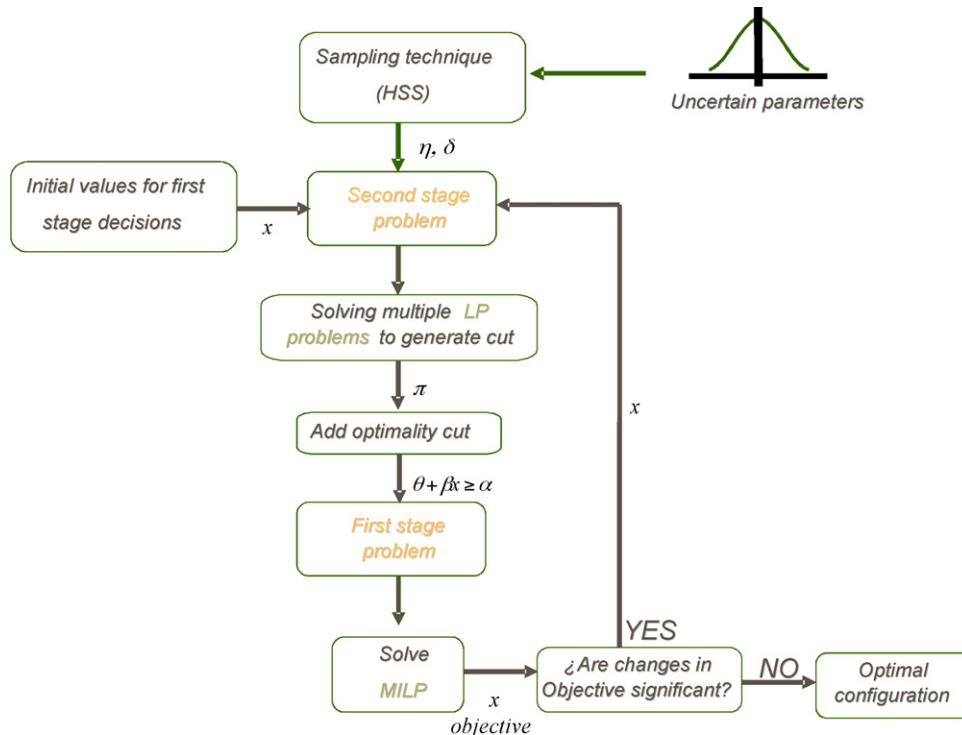


Fig. 3. The solution algorithm.

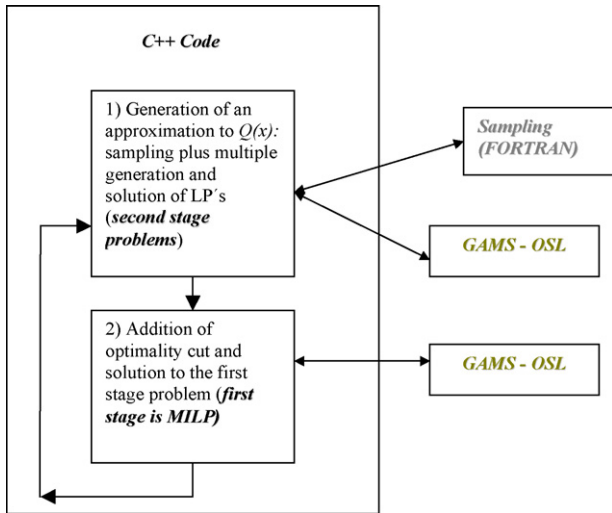


Fig. 4. Computational implementation.

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 Fig. 5a. It has 36 nodes and 40 pipes with 1 pump station and 1 storage tank. The EPANET example assumes as given 4 time periods (4 water flow patterns, i.e., different directions of flow) of 6 h in a 24 h total period. Fig. 5b shows one of the flow patterns used. Also, the nodes are distributed in 4 zones: pump station, residential neighborhood, business district and industrial district; see Table 1 for the node distribution. The maximum number of sensors is seven. The definition of the four zones, the node distribution and the value of the maximum number of sensors are the same as those provided by the EPANET case-study and used by Berry et al. (2003) while solving the deterministic case. For testing purposes, we assign

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 neighborhood	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

normal and triangular distributions to the population density. For the normal distribution, we assume a mean equal to 500 and a standard deviation of 63 (about 95% probability of having values between 375 and 625). For the triangular distribution, we consider a minimum of 375, a maximum of 625 and a mode of 500. The same probability distribution was assumed for each node and an independent sampling was performed for each of them (that is, 36 values are sampled). After sampling, the values are normalized so that the total population is always 18,000. Hence, the normalized values $\bar{\delta}_{ip}$ for the population density can be calculated from the sampled values δ_{ip} by using Eq. (6):

$$\bar{\delta}_{ip} = 18,000 \frac{\delta_{ip}}{\sum_i^{36} \delta_{ip}}. \tag{6}$$

A normal distribution is assumed for the probability of attack. For this case, however, all of the nodes in the same zone are assumed to have the same probability of being attacked (only four values of probability of attack are sampled). Four possible scenarios are considered. Table 2 shows the probability data for each of the four scenarios. Values in Table 2 do not necessarily correspond to realistic expectations of the probability of attack; they represent instances of situations in which the probability of attacking a particular zone is much higher than those of the other zones. With such selections, we intend to find how significant is the location of the attack in the optimal sensor configuration. The sampled values of the probability of attack are also normal-

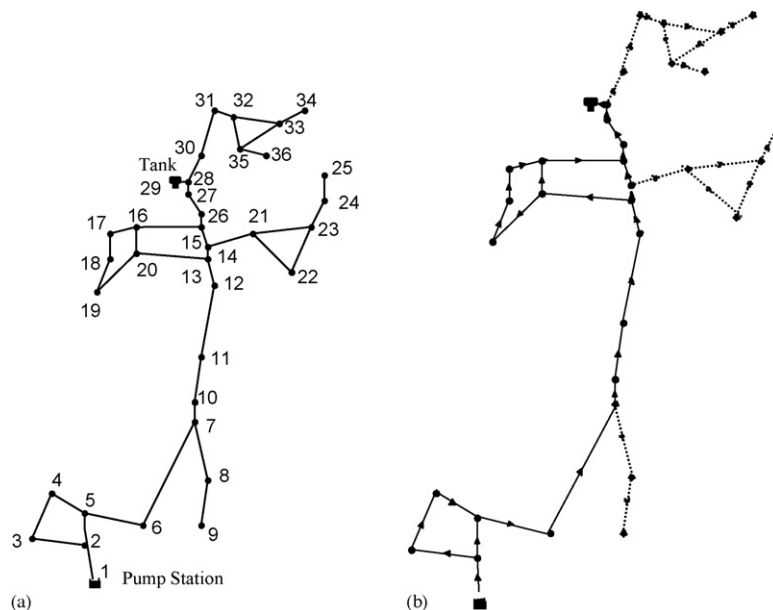


Fig. 5. Case study. (a) Nodes of the network. (b) Example of a flow pattern: (—) 100 gpm; (---) 25 gpm.

Table 2
Mean values of the 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

ized so that their sum is 1. Therefore, information of Table 2 implies for instance that, for the scenario 3, the probability of attack of all of the nodes of the residential neighborhood is the same for each of them, and their sum is approximately 0.76 (0.76 is the mean of the normal distribution for such a zone). Standard deviations for the probability of attack are chosen so that there is a 95% probability that the sampled values are found between $\pm 25\%$ of their mean value (25% of the mean value is approximately equal to 2S.D.). In addition, two costs for sensors are used: low (15,000,000) and high (45,000,000). Finally, the cost associated (S) per each affected individual is 30,000. Such values are assumed as constant and were selected in such a way that both of the terms of the objective function of Eq. (4) have the same order of magnitude. That tried to avoid

trivial solutions in which the optimal configurations suggest either no installation of sensors or installation of one sensor at each node. As mentioned before, one can also interpret those constant values as weights (or scaling factors) in the objective function. The number of samples and iterations of the SD algorithm was 400 with a computational time of 28 min and 30 s (Intel Pentium processor, 1.73 GHz). The stopping criterion used corresponds to that of the basic SD algorithm; the algorithm stops when the relative change of the objective function from one iteration to the next one is less than a given tolerance (1×10^{-3} in this case). Table 3 shows the optimal stochastic solutions for the scenario where the residential neighborhood is under risk. The deterministic solution is also presented. The deterministic case was solved by using the mean values of the stochastic parameters; for population density, the mean value is 500. The solutions for the other three scenarios present similar behaviors. Such results are shown in Tables 4–6.

4. Analysis of results

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

Table 3
Results for the residential neighborhood

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

Table 4
Results for the industrial district

Population density	Sensor cost	Variable	Stochastic solution	Deterministic solution
Triangular	Low	Sensors allocated	$X_{4-5}, X_{5-6}, X_{11-12}, X_{12-13}, X_{26-27}, X_{31-32}, X_{33-34}$	$X_{5-6}, X_{12-13}, X_{26-27}, X_{30-31}$
		Population affected (%)	15.33	30.52
		Total cost	1.8783120E+8	2.2485000E+8
	High	Sensors allocated	$X_{5-6}, X_{12-13}, X_{26-27}, X_{28-30}$	X_{12-13}, X_{26-27}
		Population affected (%)	27.56	44.12
		Total cost	3.2886790E+8	3.2825000E+8
Normal	Low	Sensors allocated	$X_{5-6}, X_{11-12}, X_{13-14}, X_{15-26}, X_{30-31}, X_{31-32}, X_{35-36}$	$X_{5-6}, X_{12-13}, X_{26-27}, X_{30-31}$
		Population affected (%)	15.97	30.52
		Total cost	1.9124361E+8	2.2485000E+8
	High	Sensors allocated	X_{12-13}, X_{26-27}	X_{12-13}, X_{26-27}
		Population affected (%)	44.70	44.12
		Total cost	3.3139140E+8	3.2825000E+8

Table 5
Results for the business district

Population density	Sensor cost	Variable	Stochastic solution	Deterministic solution
Triangular	Low	Sensors allocated	$X_{6-7}, X_{7-10}, X_{13-20}, X_{14-15}, X_{14-21}, X_{15-26}, X_{26-27}$	$X_{7-8}, X_{7-10}, X_{13-20}, X_{14-15}, X_{14-21}, X_{26-27}$
		Population affected (%)	21.02	26.10
		Total cost	2.1854458E+8	2.3097294E+8
High	High	Sensors allocated	$X_{11-12}, X_{15-26}, X_{27-28}, X_{30-31}, X_{32-33}$	$X_{7-10}, X_{14-21}, X_{15-26}$
		Population affected (%)	22.35	38.09
		Total cost	3.4571300E+8	3.4070487E+8
Normal	Low	Sensors allocated	$X_{7-10}, X_{11-12}, X_{13-20}, X_{14-15}, X_{14-21}, X_{27-28}, X_{28-30}$	$X_{7-8}, X_{7-10}, X_{13-20}, X_{14-15}, X_{14-21}, X_{26-27}$
		Population affected (%)	20.28	26.10
		Total cost	2.1456336E+8	2.3097294E+8
High	High	Sensors allocated	$X_{11-12}, X_{14-21}, X_{15-26}$	$X_{7-10}, X_{14-21}, X_{15-26}$
		Population affected (%)	38.71	38.09
		Total cost	3.4403738E+8	3.4070487E+8

Table 6
Results for the pump station

Population density	Sensor cost	Variable	Stochastic solution	Deterministic solution
Triangular	Low	Sensors allocated	$X_{1-2}, X_{10-11}, X_{12-13}, X_{19-18}, X_{30-31}, X_{32-33}$	$X_{1-2}, X_{6-7}, X_{12-13}, X_{27-28}$
		Population affected (%)	0.29	21.30
		Total cost	9.1588195E+7	1.7505000E+8
High	High	Sensors allocated	$X_{1-2}, X_{7-10}, X_{15-26}, X_{23-24}, X_{31-32}$	X_{1-2}, X_{12-13}
		Population affected (%)	3.54	30.47
		Total cost	2.4416009E+8	2.5455000E+8
Normal	Low	Sensors allocated	$X_{1-2}, X_{11-12}, X_{15-26}, X_{23-24}, X_{28-30}, X_{32-35}$	$X_{1-2}, X_{6-7}, X_{12-13}, X_{27-28}$
		Population affected (%)	0.14	21.30
		Total cost	8.9214319E+7	1.7505000E+8
High	High	Sensors allocated	$X_{1-2}, X_{7-8}, X_{11-12}, X_{26-27}, X_{30-31}$	X_{1-2}, X_{12-13}
		Population affected (%)	3.63	30.47
		Total cost	2.464705E+8	2.5455000E+8

the distribution used for the population density seems not to be important:

- The number of sensors used and the optimal sensor configuration change with respect to the cost of the sensors. In general, one could expect that the fraction of the population at risk decreases as one increases the number of sensors; hence, when using cheap sensors, the number of sensors placed tends to the maximum. However, because of the compromise between the cost of the sensors and the associated cost per each affected person, the optimal configuration does not always include the maximum number allowed. That is particularly true for the case of expensive sensors.
- The probability of attack has a much bigger impact than population density.
- Stochastic results are significantly different from deterministic results when using cheap sensors. In most cases, the percentage of population under risk is smaller in the stochastic case. There are only two cases where the opposite situation happens (business and industrial districts with normal population density distribution and high sensor cost). We believe, however, that determining which one of two approaches provides lower values for the objective

function is not the main result, but the difference between such values. Notice that a deterministic solution can both overestimate or underestimate the solution because of lack of information; our opinion is then that the stochastic solution, which uses multiples scenarios to generate the solution, will in general provide a better estimate of expected value of the objective function. Table 7 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.

Table 7
Value of the 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

- (d) The effect of uncertainty on the sensor configuration decreases when expensive sensors are used: sensors tend to be placed in nodes with the highest population density.

In a realistic situation, one would know the cost of the sensors (and the maximum number of them) and would expect to have a good idea of the distribution function for the population density; hence, the probability of attack would be the most significant of the parameters of the proposed formulation, but it would also be the most difficult to model. The optimal sensor placement suggested by our approach would have to be the one determined through the best possible approximation for the probability of attack that the available information would allow us to obtain.

On the other hand, since the first stage of the problem is a MIP model and this stage has to be solved for every sample (iteration), an increment in the number of nodes (and, therefore, in the number of edges) largely increases the combinatorial complexity of a problem. This is significant, since literature reports realistic networks with as many as 400 nodes. Hence, the proposed formulation would still be applicable, but the basic SD algorithm would have to be modified; problems of such size require a more efficient stopping criterion. The use of incumbent solutions (Higle & Sen, 1996) can then be thought as a feasible alternative.

5. Conclusions

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, probabilities of attack and sensor cost seem to have a more important effect than that of the population density; that applies for both the stochastic and the deterministic case.

Acknowledgement

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

References

- Acevedo, J., & Pistikopoulos, E. N. (1998). Stochastic optimization based algorithms for process synthesis under uncertainty. *Computers and Chemical Engineering*, 22, 647–671.
- Bagajewicz, M., Fuxman, A., & Uribe, A. (2004). Instrumentation network design and upgrade for process monitoring and fault detection. *AIChE Journal*, 50, 1870–1880.
- Berry, J., Fletcher, L., Hart, W., & Philips, C. (2003). Optimal sensor placement in municipal water networks. In *Proceedings of the world water and environmental resources conference*.
- Birge, J. R., & Louveaux, F. (1997). *Introduction to stochastic programming*. New York, NY, USA: Springer-Verlag.
- Brooke, A., Kendrick, D., Meeraus, A., & Raman, R. (1998). *GAMS-A user guide*. Washington, DC, USA: GAMS Development Corporation.
- Higle, J. L., & Sen, S. (1991). Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16, 650–669.
- Higle, J. L., & Sen, S. (1996). *Stochastic decomposition*. New York, NY, USA: Kluwer Academic Publisher.
- Kalagnanam, J., & Diwekar, U. M. (1997). An efficient sampling technique for off-line quality control. *Technometrics*, 39(3), 308.
- Novak Pintaric, Z., & Kravanja, Z. (2004). A strategy for MINLP synthesis of flexible and operable processes. *Computers and Chemical Engineering*, 28, 1105–1119.
- Ponce-Ortega, J. M., Rico-Ramirez, V., Hernandez-Castro, S., & Diwekar, U. M. (2004). Improving convergence of the stochastic decomposition algorithm by using an efficient sampling technique. *Computers and Chemical Engineering*, 28, 767–773.
- Rico-Ramirez, V. (2002). Two-stage stochastic linear programming: A tutorial. *SIAG/OPT Views and News*, 13(1), 8–14.
- Rossman, L. A. (1999). The EPANET programmer's toolkit for analysis of water distribution systems. In *Proceedings of the annual water resources planning and management conference*.
- Sahinidis, N. V. (2004). Optimization under uncertainty: State-of-the-art and opportunities. *Computers and Chemical Engineering*, 28, 971–983.
- Wei, J., & Realf, M. (2004). Sample average approximation methods for stochastic MINLPs. *Computers and Chemical Engineering*, 28, 333.