Optimal Deployment of Wireless Sensor Networks for Air Pollution Monitoring

Ahmed Boubrima*†, Frédéric Matigot*, Walid Bechkit*, Hervé Rivano* and Anne Ruas‡

*Université de Lyon, INRIA, INSA-Lyon, CITI-INRIA, F-69621, Villeurbanne, France

†École nationale Supérieure d'Informatique, LMCS Lab, Algiers, Algeria

‡IFSTTAR, Paris, France

Abstract—Recently, air pollution monitoring emerges as a main service of smart cities because of the increasing industrialization and the massive urbanization. Wireless sensor networks (WSN) are a suitable technology for this purpose thanks to their substantial benefits including low cost and autonomy. Minimizing the deployment cost is one of the major challenges in WSN design, therefore sensors positions have to be carefully determined. In this paper, we propose two integer linear programming formulations based on real pollutants dispersion modeling to deal with the minimum cost WSN deployment for air pollution monitoring. We illustrate the concept by applying our models on real world data, namely the Nottingham City street lights. We compare the two models in terms of execution time and show that the second flow-based formulation is much better. We finally conduct extensive simulations to study the impact of some parameters and derive some guidelines for efficient WSN deployment for air pollution monitoring.

Keywords— Air pollution, Detection of crossing pollution thresholds, Wireless sensor networks, Deployment, Connectivity, Coverage.

I. INTRODUCTION

The air pollution effects on human health have been extensively studied and a lot of works have shown the dramatic impact of the air pollution on humans. Regarding to the World Health Organization, around seven million people died in 2012 as a result of air pollution exposure [1]. In 2013, the International Agency for Research on Cancer (IARC) classified in the particulate matter, the main component of outdoor pollution, as carcinogenic for humans [2]. The effects of this alarming phenomenon would be more and more disturbing in smart cities of the future because of the increasing industrialization, economic development and massive urbanization. For all these reasons, we believe that the reduction of pollutant emissions as well as the detection and monitoring of pollution should be at the heart of sustainable development of smart cities.

The air pollution is traditionally monitored thanks to conventional measuring stations equipped with multiple sensors. These systems are usually inflexible, expensive and time consuming. An alternative solution would be to use wireless sensor networks which consist of a set of nodes that can measure information from the environment, process and relay them to the base station and that, through particular nodes called sinks [3]. The use of WSN for air pollution monitoring may have a great interest which can be mainly ascribed to the cost reduction, the node autonomy, the fine spatial and temporal granularity and to the self organization and self healing of the network without heavy infrastructure [4]. Wireless sensor networks used for pollution monitoring may target two objectives: i) the periodic air quality sampling and ii) the detection of threshold crossings in order to trigger adequate alerts [5]. In this work, we focus mainly on the second application where sensors are deployed to control concentrations of pollutants released by pollution sources like factories and sewage treatment plants [6][7], and we target a minimum cost optimal deployment which ensures pollution coverage and network connectivity while considering the phenomenon dispersion.

A major challenge in wireless sensor networks design is to minimize the deployment cost. The problem consists in determining the optimal positions of sensors and sinks so as to cover the environment and ensure network connectivity while minimizing the deployment cost [8], the deployment cost may include the financial nodes cost, the energy cost, etc. Coverage issue commonly known as k-coverage problem, requires that at least k sensors monitor each interest point [9]. The network is said connected if each sensor can communicate information to at least one sink [8]. For simplicity's sake, most papers on the deployment issue assume that two nodes are able to communicate with each other if the distance between them is less than a radius called the communication range [10]. Most research work on coverage use a simple detection model which assumes that a sensor is able to cover a point in the environment if the distance between them is less than a radius called the detection range [9], [11]. This can be true for some applications like presence sensors but is not suitable for pollution monitoring. Indeed, a pollution sensor can only detect pollutants that come into contact with it, and thus such a sensor does not have a detection zone like presence sensors.

We propose in this paper two optimization models for the deployment of WSN for air pollution monitoring. The expected deployment should ensure the coverage of pollution and the network connectivity while minimizing the deployment cost. Unlike existing research work, we do not base in this work on a pre-defined sensing range around a sensor node. Instead, we consider the real pollution propagation model which measures the pollutants concentration in the air. Based on this pollution dispersion modeling and the related work on ILP formulations of WSN coverage and connectivity, we first propose an optimization model of the minimum cost WSN deployment for air pollution monitoring. In this model, pollution coverage is formulated by analogy to the Set Covering Problem and connectivity modeling is based on the flow concept. Then, we propose a second model more effective in which coverage and connectivity are jointly modeled using only the flow concept. To the best of our knowledge, this is the first work that

targets wireless sensor networks deployment for air pollution monitoring based on real pollutants dispersion modeling.

The rest of this paper is organized as follows. In section II, we present air pollution modeling. Then, we review some related works on ILP formulations of WSN coverage and connectivity in section III. Section IV details our two proposed optimization models while section V shows the simulation parameters and the obtained results. Finally, we conclude and propose some perspectives in section VI.

II. Air pollution modeling

In order to define an adequate modeling of air pollution coverage, our approach was to consider the real pollution propagation models. The atmospheric dispersion of pollutants was extensively studied in the literature and a lot of models were proposed and validated. These models are of major interest for many applications such as weather forecasting, assessment of contamination, poisoning, etc. The theoretical study of pollutant atmospheric dispersion is mainly based on fluid mechanics theory [12]. Without loss of generality, we focus in this work on steady state dispersion models and in particular on the Gaussian dispersion model. Assuming that the wind direction is along the x axis and the pollutant is released at the point $(0, 0, h_s)$ in free space, this model allows the calculation of the pollutant concentration at a point (x, y, z) using the following equation [13]:

$$C(x,y,z) = \frac{Q}{2\pi u \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y}} \left(e^{-\frac{(z-H)^2}{2\sigma_y}} + e^{-\frac{(z+H)^2}{2\sigma_y}} \right) \quad (1)$$

With:

- C : Pollutant concentration (g/m^3)
- σ_y: Horizontal dispersion coefficient ($\sigma_y = a_y.|x|^{b_y}$)
- σ_z: Vertical dispersion coefficient ($\sigma_z = a_z.|x|^{b_z}$)
- Q : Emission rate at the source (g/s)
- u : Wind speed (m/s)
- H : Pollutant effective release height (m)

The parameter H, which represents the pollutant effective release height, is equal to the sum of the pollutant source height h_s, and the plume rise Δh. Briggs formulas are commonly used for the calculation of the Δh parameter. To simplify the analysis, we only consider the case where the temperature of the pollutant T_s is greater than that of the ambient air T_a. In this case, the formula giving the value of Δh is as follows:

$$\Delta h = \frac{1,6 \cdot F^{1/3} \cdot x^{2/3}}{u} \quad (2)$$

Where:

$$F = \frac{g}{\pi} \cdot V \cdot \left(\frac{T_s - T_a}{T_s} \right) \quad (3)$$

With:

- Δh : Plume rise (m)
- F : Buoyancy (m^4/s^3)
- u : Average wind speed (m/s)
- x : Distance from the source (wind direction) (m)
- g : Gravity constant (9,8 m/s^2)
- V : Volumetric flow (m^3/s)
- T_s : Pollutant temperature at its emission point (K)

- T_a : Ambient air temperature (K)

Based on this Gaussian model, many enhanced systems were developed especially those recommended by the EPA (Environmental Protection Agency) such as ISC-PRIM and AERMOD [14]. These models take into account meteorological data, buildings, ...etc. As proof of concept and without loss of generality, we will consider in what follows the basic Gaussian pollution dispersion model.

III. Related work

Several integer linear programming formulations have been proposed in the literature to model wireless sensor networks coverage and connectivity issues [15]. Some of these models have as objective to optimize sensors and sinks deployment [16] while the others aim to ensure a topology control like the adaptation of the transmission range of nodes in order to minimize the energy consumption [10]. The common property of all these models is that they do not take into account the characteristics of the environment and the nature of the phenomenon. Instead, they assume that each node can transmit information to neighbors which are in its communication range and cover points within its detection range. Coverage and connectivity constraints were been sometimes modeled separately. In some of these works, authors assume that coverage involves connectivity while considering that the communication range is a multiple of the detection range [15]. In other works, authors assume that coverage is already ensured with some existing sensors, and thus formulate only the connectivity constraint [10]. Some authors formulated coverage and connectivity together in the same ILP model. Unfortunately, these models formulate the two constraints independently, coverage with the Set Covering Problem based on the sensors detection areas and connectivity based on the flow concept.

Chakrabarty et al. [11] represent the environment as a two or three dimensional grid of points which form the sensor field to be covered, and propose first a non linear formulation for minimizing the cost of sensors deployment while ensuring complete coverage of the sensor field. Then, they apply some transformations to linearize the first model and obtain an ILP formulation. The authors assume that the points to be covered are the same on which sensors should be placed, and model coverage based on the distance between the different points of the environment. Therefore, each sensor has a circular detection area which defines the points that it can cover. Different types of sensors are considered with different detection ranges and different costs. Clearly, sensors which have larger detection ranges have higher cost, and thus using them may be prohibitive in terms of the total placement cost.

Chakrabarty et al.'s model suffers from the intractability since it is based on a non linear formulation. In addition, sensors may have a non circular detection area in some applications. Meguerdichian and Potkonjak [17] deal with all these drawbacks and propose an ILP formulation of coverage based on the Set Covering Problem which is a well known optimization problem. These authors consider a set of positions where sensors can be placed and a set of discrete points approximating the sensor field. A sensor can have a detection area with a shape which is not necessarily circular. This ILP model does not take into account sensors with different costs

and treat only the 1-Coverage where a point should be covered by only one sensor.

The integer programming formulations proposed in [11] and [17] do not take into account the different coverage requirements of the environment points, instead the authors assume uniform coverage where all the points in the sensor field have to be covered by the same number of sensors. This may be unrealistic in some applications where some zones in the environment are more critical than others. Altinel et al [15] focused on this issue and proposed an integer linear programming formulation that considers different coverage requirements among the sensor field points. They also show that their model can deal with connectivity under the assumption that the transmission range is as least equal to the double of the detection range. This cannot be true in most of the applications since sensing and transmission are two independent functionalities of a sensor node.

Connectivity constraint has been studied in different contexts including topology control and deployment issues. Most of the existing models are based on the flow concept [10], [16], [18]. Theses formulations assume that each sensor generates a flow unit and check whether the generated units can be recovered at the sinks. In some models [10], sinks were considered already located and deployment modeling treats only sensor nodes. The other models [16], [18] consider the two types of nodes, sensors and sinks. However, all these models assume that the potential positions of sinks are different from those of sensors. This cannot be applied in some applications where a potential position can correspond to both sensors and sinks.

Unfortunately, all the works presented in this section do not base on the nature of the phenomenon and suppose that a sensor is able to cover points within its detection range, this cannot be directly applied to air pollution monitoring since a pollution sensor can only detect pollutants that come into contact with it. Moreover, the coverage and connectivity constraints are modeled independently in the sense that coverage is formulated by analogy to the Set Covering Problem and connectivity formulation is based on the flow concept. In the next section, we address these issues while proposing two ILP formulations based on real pollutants dispersion modeling, and treat in the second model the joint formulation of coverage and connectivity using only the flow concept.

IV. OPTIMIZATION MODELS

In this paper, we propose two integer programming formulations for optimal deployment of wireless sensor networks to monitor efficiently air pollution. We consider in the two formulations a real pollutants dispersion model which can be the Gaussian model or any other sophisticated one. Thus, our formulations can be used with any type of environment just by adapting the atmospheric dispersion model. We focus mainly on pollutants released by static sources like factories, sewage treatment plants, etc. In the first basic formulation, we model coverage by analogy to the *Set Covering Problem* and connectivity using the *flow concept*. In the enhanced model, we propose a novel formulation in which we treat coverage and connectivity in a joint way using only the *flow concept*.

In smart cities applications, some restrictions on node positions may apply because of authorization or practical issues. For instance, in order to alleviate the energy constraints, we may place sensors on lamp posts and traffic lights as experimented in CitySense [19]. We denote the set of potential positions $\mathcal{P}$. Each potential position $p \in \mathcal{P}$ may correspond after solving the problem to the position of a sink or a sensor. The set $\mathcal{P}$ can also be a grid of points obtained using a discretization process especially in free space environments where there are no deployment restrictions. Moreover, we consider a set $\mathcal{I}$ that consists of positions of static pollution sources like factories, sewage treatment plants, etc. Let $\mathcal{N}$ denote the number of potential positions (i.e. $|\mathcal{P}| = \mathcal{N}$) and $\mathcal{M}$ denote the number of pollution sources (i.e. $|\mathcal{I}| = \mathcal{M}$). We use binary decision variables x_p resp. y_p to define if a sensor resp. a sink should be placed at position p. Sensors and sinks have different costs, thus we denote by c_p^{sensor} resp. c_p^{sink} the sensor resp. the sink deployment cost at position p. Our optimization models aim to minimize the sensors and sinks overall deployment cost. Thus, the objective function to minimize is as follows :

$$\mathcal{F} = \sum_{p \in \mathcal{P}} c_p^{sensor} * x_p + \sum_{p \in \mathcal{P}} c_p^{sink} * y_p \qquad (4)$$

Where a potential position p cannot belong to a sensor and a sink at the same time as formulated in constraint 5.

$$x_p + y_p \leq 1, \quad p \in \mathcal{P} \qquad (5)$$

A. Basic Model

We formulate the coverage constraint based on pollution dispersion modeling. We assume that only one source in the environment can release pollutants at the same time, in which case the emission rate is equal to Q. Using an atmospheric dispersion model, we determine for each pollution source i the zone $\mathcal{Z}_i$ which will be polluted in the case where the source starts to emit these pollutants :

$$\mathcal{Z}_i = \{p \in \mathcal{P} \ where \ C_{ip} \geq C_0\} \qquad (6)$$

Where C_{ip} is the pollutant concentration measured at point p when the source i starts to emit pollutants and C_0 is the threshold of pollutant concentration from which a point of the environment is considered as polluted. We assume that sinks are also equipped with pollution sensors. We define then the binary parameter $\mathcal{W}_{ip}$ which is set to 1 if p belongs to $\mathcal{Z}_i$; otherwise, $\mathcal{W}_{ip} = 0$. When using the Gaussian dispersion model, the value of $\mathcal{W}_{ip}$ is calculated as follows :

$$\mathcal{W}_{ip} = \left(\frac{Q}{2\pi u \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y}} \left(e^{-\frac{(z-H)^2}{2\sigma_y}} + e^{-\frac{(z+H)^2}{2\sigma_y}}\right) \geq C_0\right) \quad (7)$$

Where σ_y, σ_z, Q, u and H are the same parameters presented in section II. At this point, we formulate the K-Coverage constraint based on the ILP formulation of the *Set Covering Problem* :

$$\sum_{p \in \mathcal{P}} \mathcal{W}_{ip} * (x_p + y_p) \geq K, \quad i \in \mathcal{I} \qquad (8)$$

This formulation assumes that the coverage requirements among all the pollution sources are the same, i.e. each source should be covered by at least K sensors. To model a differentiated coverage where the coverage requirements of the sources may be different, index $i \in \mathcal{I}$ is added to K to obtain K_i.

As in [10], [16] and [18], we model connectivity using the *flow concept*. However, our formulation is better in the sense that we consider the same potential positions set $\mathcal{P}$ for sensors and sinks and we do not assume that potential positions of sinks are different from those of sensors. We first denote by $\Gamma(p)$, $p \in \mathcal{P}$ the neighbors set of a node at the potential position p. This set can be determined using path loss models that are adequate for the nature of the environment. Then, we define the decision variables g_{pq} as the flow quantity transmitted from node located at potential position p to node located at potential position q. The following constraints ensure that present sensors and sinks form a connected wireless sensor network; i.e. each sensor can communicate with at least one sink.

$$\sum_{q \in \Gamma(p)} g_{pq} - \sum_{q \in \Gamma(p)} g_{qp} \geq x_p - \mathcal{N} * y_p, \quad p \in \mathcal{P} \quad (9)$$

$$\sum_{q \in \Gamma(p)} g_{pq} - \sum_{q \in \Gamma(p)} g_{qp} \leq x_p, \quad p \in \mathcal{P} \quad (10)$$

$$\sum_{q \in \Gamma(p)} g_{pq} \leq N * x_p, \quad p \in \mathcal{P} \quad (11)$$

$$\sum_{p \in \mathcal{P}} \sum_{q \in \Gamma(p)} g_{pq} = \sum_{p \in \mathcal{P}} \sum_{q \in \Gamma(p)} g_{qp} \quad (12)$$

Constraints 9 and 10 are designed to ensure that each present sensor, i.e. $x_p = 1$, generates a flow unit in the network. These constraints are equivalent to what follows :

$$\sum_{q \in \Gamma(p)} g_{pq} - \sum_{q \in \Gamma(p)} g_{qp} \begin{cases} = 1 & if \ x_p = 1, y_p = 0 \\ = 0 & if \ x_p = y_p = 0 \\ \leq 0, \geq -\mathcal{N} & if \ x_p = 0, y_p = 1 \end{cases}$$

The first case corresponds to present sensors which should generate, each one of them, a flow unit. The second case, combined with constraint 11, ensures that absent nodes, i.e. $x_p = y_p = 0$, do not participate in the communication. The third case concerns present sinks, and has no effect since there are at most $\mathcal{N}$ units in the network, and thus a sink cannot receive more than $\mathcal{N}$ units. However, this case avoids that a contradiction occurs as a consequence of the formulation of the two first cases by the same constraints 9 and 10. In addition, constraint 11 ensures that present sinks ($y_p = 1$) cannot transmit flow units, and thus they are considered as receivers. Constraint 12 means that the overall flow is conservative and thus the units generated by present sensors will be recovered by present sinks. Our basic optimization model can then be written as follows :

$$\begin{aligned} Minimize \quad & (4) \\ Subject \ to. \quad & (5), (8), (9), (10), (11) \ and \ (12) \\ x_p \quad & \in \{0, 1\}, \ p \in \mathcal{P} \\ y_p \quad & \in \{0, 1\}, \ p \in \mathcal{P} \\ g_{pq} \quad & \in \{0, ..., \mathcal{N}\}, \ p \in \mathcal{P}, \ q \in \Gamma(p) \end{aligned}$$

B. Enhanced Model

In the basic model, connectivity and coverage are formulated as disjoint constraints. For a more homogeneous modeling, we propose a joint coverage/connectivity ILP formulation using only the flow concept, and this by considering pollution sources as a part of the network. In the new model, each pollution source i should transmit some flow units to potential nodes p which are located in its impacted zone i.e. $p \in \mathcal{Z}_i$. In addition, sensors are flow conservative so as that sinks receive the flow units generated by pollution sources. Therefore, The definition of the joint coverage/connectivity is to ensure that sinks will be informed each time that pollution sources emit pollutants.

In this enhanced formulation, a sensor cannot receive more than one unit from the same pollution source, hence we define the binary decision variable f_{ip} as the flow quantity from the pollution source i to the potential node p. The following constraints ensure coverage and connectivity for pollution monitoring in a joint way based only on the flow concept.

$$\sum_{p \in \mathcal{Z}_i} f_{ip} = K, \quad i \in \mathcal{I} \quad (13)$$

$$\begin{aligned} \sum_{i \in \mathcal{I} \ : \ p \in \mathcal{Z}_i} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} - \sum_{q \in \Gamma(p)} g_{pq} & \\ \leq K * \mathcal{M} * y_p, \quad p \in \mathcal{P} & \end{aligned} \quad (14)$$

$$\begin{aligned} \sum_{i \in \mathcal{I} \ : \ p \in \mathcal{Z}_i} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} - \sum_{q \in \Gamma(p)} g_{pq} & \\ \geq 0, \quad p \in \mathcal{P} & \end{aligned} \quad (15)$$

$$\begin{aligned} \sum_{p \in \mathcal{P}} \Big(\sum_{i \in \mathcal{I} \ : \ p \in \mathcal{Z}_i} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} \Big) & \\ = \sum_{p \in \mathcal{P}} \sum_{q \in \Gamma(p)} g_{pq} + K * \mathcal{M} & \end{aligned} \quad (16)$$

$$\sum_{q \in \Gamma(p)} g_{pq} <= K * \mathcal{M} * x_p, \quad p \in \mathcal{P} \quad (17)$$

$$\sum_{i \in \mathcal{I} \ : \ p \in \mathcal{Z}_i} f_{ip} <= \mathcal{M} * (x_p + y_p), \quad p \in \mathcal{P} \quad (18)$$

Constraint 13 ensures that each pollution source i generates K flow units in the network, each unit is sent to a node located in the impacted zone of i. All of these nodes should be present, i.e. $x_p + y_p = 1$, thanks to constraint 18, otherwise f_{ip} values are set to 0. The flow is conservative on present sensors, i.e. $x_p = 1$, as formulated in constraints 14 and 15, this means that the inflow of each present sensor is equal to its outflow. These two constraints take into account three cases where a

sensor, a sink or nothing is placed at a position p, and ensure when combined with constraints 17 and 18 that absent nodes, i.e. $x_p + y_p = 0$, do not participate in the communication. Constraint 16 ensures the overall flow conservation in the network, and thus implies that present sinks will recover the $K * \mathcal{M}$ flow units generated by pollution sources. The enhanced optimization model can then be written as follows :

[Enhanced Model]

$$
\begin{aligned}
Minimize \quad & (4) \\
Subject\ to. \quad & (5), (13), (14), (15), (16), \\
& (17)\ and\ (18) \\
x_p \quad & \in \{0,1\},\ p \in \mathcal{P} \\
y_p \quad & \in \{0,1\},\ p \in \mathcal{P} \\
g_{pq} \quad & \in \{0,...,K * \mathcal{M}\},\ p \in \mathcal{P}, \\
& q \in \Gamma(p) \\
f_{ip} \quad & \in \{0,1\},\ i \in \mathcal{I},\ p \in \mathcal{Z}_i
\end{aligned}
$$

V. SIMULATION RESULTS

We have performed extensive simulations to evaluate our two optimization models. For all the tests, we used the Gaussian dispersion model to simulate pollutant propagation. ILP formulations are implemented using the IBM ILOG CPLEX Optimization Studio and executed on a PC with Intel Xeon E5649 processor under Linux. The common tests parameters are summarized in table I. We consider that nodes have circular transmission range, all the environmental conditions (temperature, pressure, etc.) are homogeneous and pollutants have the same characteristics.

| Parameter	value
Nodes transmission range	$60m$
Pollution sources height	$25m$
Pollution sources emission rate (Q)	$1.59g/s$
Pollutant concentration threshold (C_0)	$0.034mg/m3$
Wind speed	$5m/s$
Pollutant temperature	30^oC
Ambient air temperature	15^oC
Volumetric flow	$0.603mm^3$
Horizontal dispersion coefficients	$a_y = 0.34, b_y = 0.82$
Vertical dispersion coefficients	$a_z = 0.275, b_z = 0.69$

TABLE I: Summarize of common simulation parameters.

As a proof of concept, we first execute our models on a real data set, namely the Nottingham City Open Data [20]. We choose 717 street lights in a $700m\ X\ 700m$ area as shown in figure (1a). Due to the lack of factories data on factory locations in the city, we select some positions for pollution sources that we represent on the same figure. Figure (1b) shows the selected street lights on which should be placed sensors and the sink. We recall that the Gaussian dispersion model assumes that wind direction is from left to right along the x axis. As shown in figure(1b), only 3 sensors and one sink are sufficient to monitor pollutants emitted by the seven factories.

For all the remaining simulations, we consider a free space environment in which we generate randomly a number of pollution sources. Then, we define the set of potential nodes as the grid points obtained by 2D discretization of the

(a) Initial configuration. Blue points correspond to factories positions and red points correspond to street lights positions.
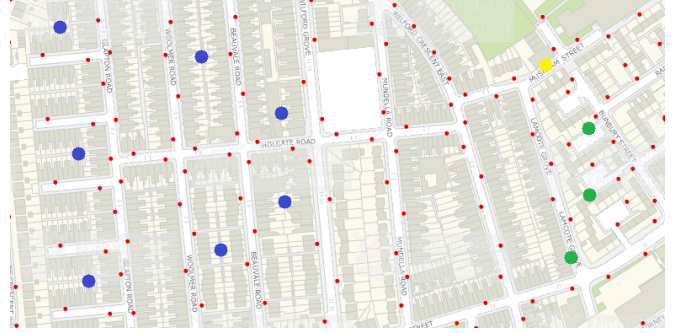
(b) Final configuration. Green points correspond to sensors selected positions and the yellow point correspond the sink selected position.

Fig. 1: Optimal deployment of WSN for pollution monitoring in Nottingham City. Coverage requirement is set to 1 and a sink is considered ten times much expensive than a sensor.

initial environment with a resolution of $50m$; i.e. the distance between each pair of potential nodes is equal to $50m$.

A. Evaluation of execution time

Before to study the impact of different parameters on deployment cost, we evaluate the execution time of the two proposed models. Table II shows averaged CPU times of both basic and enhanced models over ten executions performed depending on coverage requirements and sensors height; i.e. height at which sensors are placed. For all these executions, we considered 400 potential nodes and 8 pollution sources randomly generated in a $1000X1000m$ area. We report in each cell of the table the average of the execution time.

We checked that the results obtained with the two models are similar in terms of deployment cost. However, the enhanced model is much faster than the basic one as shown in table II. This is due to the fact that the enhanced model treats coverage and connectivity in a joint way using only the flow concept. We also notice that the execution time for 2-coverage is higher than the execution time for 1-coverage. In what follows, we use the enhanced model to evaluate the impact of some parameters on the deployment choices.

Height	Basic Model		Enhanced Model	
	K = 1	K = 2	K = 1	K = 2
10	128.798013	158.470817	2.094243	2.218378
15	123.622492	158.109400	2.084798	2.280900
20	125.220984	186.392799	2.097111	2.346862
25	120.252016	222.966718	2.105083	2.280286
30	133.821470	280.666131	2.102622	2.200396
35	128.629180	287.696362	2.123649	2.210283
40	120.523160	270.322285	2.138433	2.202760

TABLE II: CPU times in seconds of our two optimization models depending on different sensors height and coverage requirements values.

B. Impact of the nodes height

In this scenario, we study the impact of the height at which are placed sensors and sinks on the deployment cost. We assume that all the sensors and sinks are deployed at the same height which is considered in the range from 0 to 45 meters. We recall that pollution sources height is equal to $25m$. With the simulation parameters presented in table I, the value of ΔH is nearly equal to $0.1m$ and thus H, the pollutant effective release height, is nearly equal to $25m$. We plot in figure 2 the sensors and sinks overall deployment cost depending on their height for 1-Coverage and 2-Coverage. We notice that the deployment cost is minimal when the nodes height is close to the height of pollution sources. This is explained by the fact that the pollutant concentration reaches its maximum value at the pollution effective release height H. Moreover, pollutants are more likely to drop than to increase which is due to gravitation. Indeed, the deployment cost at $45m$ is much greater than the deployment cost at $0m$. The figure also shows that 2-Coverage requires more nodes than 1-Coverage to ensure fault tolerance. Indeed, 2-Coverage requires that two sensors monitor each pollution source. However, the results show that the cost factor between the two levels of requirements ranges from 1.30 to 1.36, but not the twice, and reaches the minimum value at height close to effective release height. This is explained by the intersection existence between the different polluted zones $\mathcal{Z}_i$ which means that in some cases a sensor can monitor more than one pollution source.

C. Impact of the cost ratio between sinks and sensors

We have also evaluated the impact of the ratio between sinks cost and sensors cost on their required number for minimizing the deployment cost. We plot in figure 3 the obtained results when the cost ratio ranges from 1 to 10. On one hand, We notice that sensors are less used when their cost is close the sinks cost. For instance, only sinks are used when the cost ratio is equal to 1. This is expected since sinks are equipped with pollution sensors and do not require other nodes to forward their packets unlike sensor nodes. On the other hand, when the cost ratio increases, more sensors are used and the number of required sinks tends to one, in which case the network is formed by only one sink. This is explained by the fact that adding some relay sensor nodes to ensure connectivity costs less than using a lot of sinks which are equipped with pollution sensors.
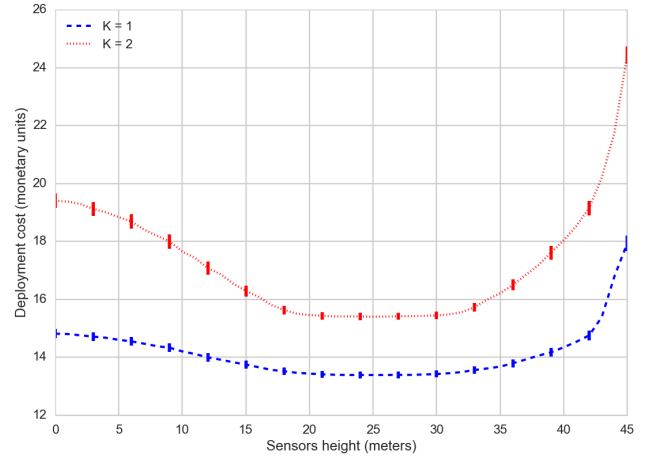

Fig. 2: Deployment cost average and the 98% confidence intervals depending on the nodes height. A $1000m X 1000m$ area is considered and 8 pollution sources are randomly generated. A sink is considered ten times much expensive than a sensor.
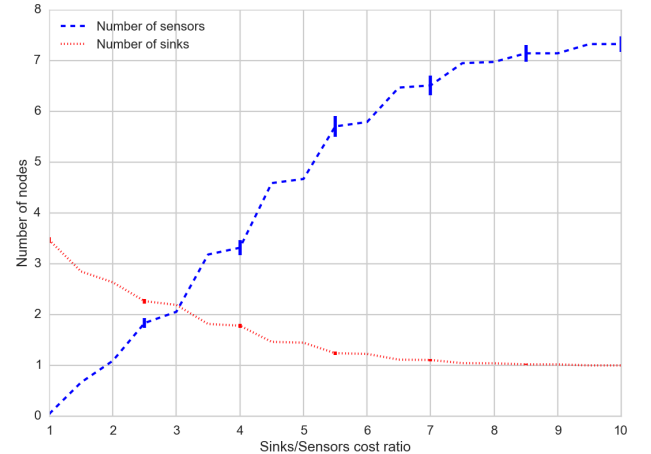

Fig. 3: Number of sinks and sensors average and the 98% confidence intervals depending on their cost ratio in the case of 1-Coverage. A $400m X 1000m$ area is considered and 5 pollution sources are randomly generated. Nodes are placed at a height of $20m$.

D. Impact of the pollution sources density

In this last scenario, we study the impact of pollution sources density on deployment cost. We plot in figure 4 the deployment cost variations depending on the number of sources for 1-Coverage and 2-Coverage requirements. Figure 4 shows that more there are pollution sources in the environment, more there are sensors required and thus higher is the deployment cost. This is can be explained by the number of pollution zones $\mathcal{Z}_i$ that increases with the number of pollution sources, and thus requires much sensors to ensure the coverage requirements. However, the obtained curves show

that the cost remains the same when the environment becomes dense of pollution sources. This is due to the fact that in this case too much sensors and sinks are placed so that a new pollution source will be necessarily covered. Moreover, the deployment cost for 2-Coverage is close to the deployment cost for 1-Coverage when there are few pollution sources in the environment, but is too much higher when the environment is dense of pollution sources. Indeed, the ratio between the two coverage requirements in the deployment cost ranges from 1.1 to 1.3.
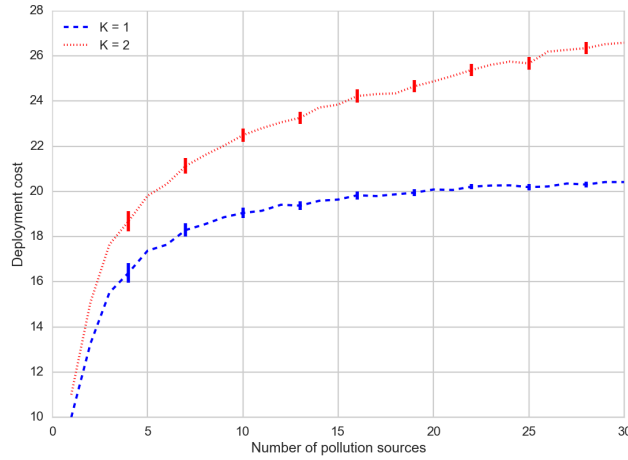
Fig. 4: Deployment cost average and the 98% confidence intervals depending on the number of pollution sources. A $400m X 1000m$ area is considered, nodes are placed at a height of $20m$ and a sink is considered ten times much expensive than a sensor.

VI. CONCLUSION AND FUTURE WORK

We addressed in this work the optimal deployment of wireless sensor networks for air pollution monitoring. To this end, the careful study of the dispersion modeling is something that should not be overlooked. We therefore studied a known model of diffusion of atmospheric pollution. We proposed two ILP optimization models that ensures both pollution coverage and network connectivity. We implemented and compared the two models in terms of execution time, the results showed that the second flow-based formulation is much better. We also conducted extensive simulations and derived some results to guide the choice towards an optimal deployment. As a future work, we plan to consider the impact of other parameters such as wind direction, the nature of pollutants, urban topography, etc. Moreover, we are also working on the design of specific heuristics to solve the addressed problem.

ACKNOWLEDGMENT

REFERENCES

[1] World Health Organization, "Burden of disease from household air pollution for 2012," available on http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf, last access: 2015-03-27.

[2] International Agency for Research on Cancer, "Iarc: Outdoor air pollution a leading environmental cause of cancer deaths," available on http://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf, last access: 2015-04-01.

[3] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[5] ETSI, "Electromagnetic compatibility and radio spectrum matters (erm); system reference document (srdoc): Spectrum requirements for short range device, metropolitan mesh machine networks (m3n) and smart metering (sm) applications," available on http://www.etsi.org/deliver/etsi_tr/103000_103099/103055/01.01.01_60/tr_103055v010101p.pdf, last access: 2015-04-01.

[6] Y. J. Jung, Y. K. Lee, D. G. Lee, K. H. Ryu, and S. Nittel, "Air pollution monitoring system based on geosensor network," in *IGARSS 2008. IEEE International*, vol. 3. IEEE, 2008, pp. III–1370.

[7] M. Gao, F. Zhang, and J. Tian, "Environmental monitoring system with wireless mesh network based on embedded system," in *Proceedings of the 2008 Fifth IEEE International Symposium on Embedded Computing*. IEEE Computer Society, 2008, pp. 174–179.

[8] M. Younis and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 6, no. 4, pp. 621–655, 2008.

[9] A. Ghosh and S. K. Das, "Coverage and connectivity issues in wireless sensor networks," *Mobile, wireless, and sensor networks: Technology, applications, and future directions*, pp. 221–256, 2006.

[10] M. Cardei, M. O. Pervaiz, and I. Cardei, "Energy-efficient range assignment in heterogeneous wireless sensor networks," in *ICWMC'06. International Conference on*. IEEE, 2006, pp. 11–11.

[11] K. Chakrabarty, S. S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks," *Computers, IEEE Transactions on*, vol. 51, no. 12, pp. 1448–1453, 2002.

[12] A. Daly and P. Zannetti, "Air pollution modeling–an overview," *Ambient air pollution*, 2007.

[13] J. M. Stockie, "The mathematics of atmospheric dispersion modeling," *Siam Review*, vol. 53, no. 2, pp. 349–372, 2011.

[14] Environmental_Protection_Agency, "Epa preferred/recommended models," available on http://www.epa.gov/scram001/dispersion_prefrec.htm, last access: 2015-04-13.

[15] İ. K. Altınel, N. Aras, E. Güney, and C. Ersoy, "Binary integer programming formulation and heuristics for differentiated coverage in heterogeneous sensor networks," *Computer Networks*, vol. 52, no. 12, pp. 2419–2431, 2008.

[16] M. E. Keskin, İ. K. Altınel, N. Aras, and C. Ersoy, "Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility," *Ad Hoc Networks*, vol. 17, pp. 18–36, 2014.

[17] S. Meguerdichian and M. Potkonjak, "Low power 0/1 coverage and scheduling techniques in sensor networks," Citeseer, Tech. Rep., 2003.

[18] M. Patel, R. Chandrasekaran, and S. Venkatesan, "Energy efficient sensor, relay and base station placements for coverage, connectivity and routing," in *IPCCC 2005. 24th IEEE International*. IEEE, 2005, pp. 581–586.

[19] R. N. Murty, G. Mainland, I. Rose, A. R. Chowdhury, A. Gosain, J. Bers, and M. Welsh, "Citysense: An urban-scale wireless sensor network and testbed," in *Technologies for Homeland Security, 2008 IEEE Conference on*. IEEE, 2008, pp. 583–588.

[20] Nottingham_OpenData, "Streetlights," available on www.opendatanottingham.org.uk/dataset.aspx?id=35, last access: 2015-04-10.