Poster: Toward a Better Monitoring of Air Pollution using Mobile Wireless Sensor Networks

Ahmed Boubrima, Walid Bechkit, Hervé Rivano, Lionel Soulhac

Poster: Toward a Better Monitoring of Air Pollution using Mobile Wireless Sensor Networks

Ahmed Boubrima*, Walid Bechkit*, Hervé Rivano* and Lionel Soulhac †
*Univ Lyon, Inria, INSA Lyon, CITI, F-69621 Villeurbanne, France
†LMFA, Univ Lyon, CNRS UMR 5509 ECL, INSA Lyon, Univ Claude Bernard, 69134 Ecully, France

***Abstract*—Mobile wireless sensor networks (MWSN) are widely used for monitoring physical phenomena such as air pollution where the aim is usually to generate accurate pollution maps in real time. The generation of pollution maps can be performed using either sensor measurements or physical models which simulate the phenomenon of pollution dispersion. The combination of these two information sources, known as data assimilation, makes it possible to better monitor air pollution by correcting the simulations of physical models while relying on sensor measurements. The quality of data assimilation mainly depends on the number of measurements and their locations. A careful deployment of nodes is therefore necessary in order to get better pollution maps. In this ongoing work, we tackle the placement problem of pollution sensors and design a mixed integer programming model allowing to maximize the assimilation quality while ensuring the connectivity of the network. We perform some simulations on a dataset of the Lyon city, France in order to show the effectiveness of our model regarding the quality of pollution coverage.**

***Keywords*—** Mobile wireless sensor networks, placement, air pollution monitoring, data assimilation.

I. Introduction and related work

Mobile wireless sensor networks (MWSN) are widely used in environmental applications where the aim is to sense a physical phenomenon such as temperature, humidity, air pollution, etc. Air pollution is one of the main physical phenomena that still need to be studied and characterized. According to the World Health Organization (WHO), exposure to air pollution is accountable to seven million casualties in 2012. Air pollution is therefore considered as a major issue of modern megalopolis, where the majority of world population lives. As a consequence, the effective monitoring of pollutant emissions is at the heart of many sustainable development efforts. The progress of electrochemical sensors, that are smaller and cheaper while keeping a reasonable measurement quality, makes the use of MWSN for air pollution monitoring viable (1). The aim of using MWSN for air pollution monitoring is usually to generate accurate pollution maps in real time (2). Unlike our work presented in (3) where we consider only sensor measurements in the generation of pollution maps, we propose in this abstract paper to perform pollution mapping based on physical models which simulate the phenomenon of pollution dispersion. Our aim is to focus on data assimilation in order to correct the simulations of the physical models based on sensor measurements, which would lead to better performances. In this context, we tackle in this paper the optimal deployment of mobile sensor nodes for effective data assimilation.

The deployment issue of wireless sensor networks has been extensively studied in the literature where several mathematical models, optimal algorithms and near-optimal heuristics have been proposed. The problem has been defined in multiple ways depending on the context of the deployment. The main issues targeted in the literature are coverage, connectivity, network lifetime and the network deployment cost. To the best of our knowledge, existing deployment approaches are either event-aware (like the work of (4)) or correlation-aware (like the work of (5)). In the first case, a sensor is assumed to have a detection range, usually circular, within which the sensor is capable of detecting any event that may happen. This cannot be applied to air pollution monitoring because a pollution sensor needs to take an air sample in order to evaluate the pollutant concentration. The second class of deployment approaches is based on the correlation that sensor measurements may present in order to select the minimum number of sensing nodes. In addition, non realistic properties of some phenomena like pollution being Gaussian are assumed in these works.

Novel application-aware deployment methods have been recently proposed to consider the characteristics of the application case in the design of the deployment approach; one example is the work of (6) on wind monitoring. Following the same direction, we propose in this abstract paper a mixed integer programming model (MILP) allowing to maximize the assimilation quality and hence minimize the error of assimilation while ensuring the connectivity of the mobile sensor network. Unlike most of the existing deployment approaches, we base on data assimilation to define an appropriate mathematical formulation of coverage quality in the context of air pollution mapping. We perform some simulations on a dataset of the Lyon city, France in order to show the effectiveness of our model.

II. Deployment model

We consider three types of nodes: 1) mobile sensors that are used to measure pollution concentrations and which are deployed on top of electric buses and trams; 2) relay nodes which are used for connectivity; and 3) sinks which are used to gather data from all the other nodes. We also consider as input the map of a given urban area that we call the deployment region. Let $\mathcal{P}$ be a set of discrete points approximating the deployment region at a high-scale ($|\mathcal{P}| = \mathcal{N}$). The set $\mathcal{P}$ can be obtained using a 2D or 3D discretization. Our goal in this paper is to be able to determine with a high precision the concentration value of pollution at each point $p \in P$. We ensure that for each point $p \in P$, either there is a mobile sensor which passes through p; or the pollution concentration can be estimated with a high precision based on the physical model simulations and the data gathered by the neighboring mobile sensors. In addition to pollution coverage, we also ensure that all the deployed sensors can send their data to at least one sink node while optimizing the positions of sinks and relay nodes.

We use decision variables r_p (respectively y_p) to specify if a relay node (respectively a sink) is deployed at point p or not. As for mobile sensor nodes, let first $\mathcal{I}$ be the set of lines of buses and trams. We consider each trajectory $i \in \mathcal{I}$ as a subset of $\mathcal{P}$. We use binary variables t_i to denote the fact that a mobile sensor is deployed on the trajectory i or not. In addition, let the binary variables x_p denote the fact that there is at least one mobile sensor which passes through the point p or not. A variable x_p is necessarily equal to 0 if no mobile sensor passes through it as formulated in constraint 1 where $\mathcal{I}_p$ denotes the set of trajectories that include the point p.

$$x_p \leq \sum_{i \in \mathcal{I}_p} t_i, \quad p \in \mathcal{P} \tag{1}$$

A. Deployment budget

We first denote by θ_i (respectively δ_p and ψ_p) the deployment cost of a sensor on trajectory i (respectively a relay node or a sink at point p). The network deployment cost should not exceed the deployment budget J as follows:

$$\sum_{i \in \mathcal{I}} \theta_i \cdot t_i + \sum_{p \in \mathcal{P}} \delta_p \cdot r_p + \sum_{p \in \mathcal{P}} \psi_p \cdot y_p <= J \tag{2}$$

B. Air pollution coverage

Using data assimilation, the estimated concentration $\widehat{\mathcal{Z}}_p$ at a given location $p \in \mathcal{P}$ where no mobile sensor passes through is formulated as the sum of $\mathcal{M}_p$, which is the physical model simulation value at p, and a weighted combination of the errors of the physical model at neighboring sensor nodes $m_q,\ q \in \mathcal{P}\ where\ x_q = 1$. The m_q values are determined with respect to the ground truth (or real) concentrations and can be estimated using sensors with high accuracy. The weights $\mathcal{W}_{pq}$ are called correlation coefficients and can be evaluated in a deterministic way based on the distance between the location of the measured concentration and the location of the estimated concentration. In this case, $\widehat{\mathcal{Z}}_p$ is evaluated using formula 3. The data assimilation equation in formula 3 is constrained by formula 4, which ensures that the denominator is never equal to 0. $\mathcal{B}_{pq}$ parameters define whether there is a correlation between points p and q or not. Given the formula of the assimilation estimated concentration $\widehat{\mathcal{Z}}_p$, the assimilation error with respect to the ground truth value can be derived as in formula 5 with m_p being the physical model simulation error at point p.

$$\widehat{\mathcal{Z}}_p = \mathcal{M}_p + \frac{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot m_q}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \tag{3}$$

$$\sum_{q \in \mathcal{P}} \mathcal{B}_{pq} \cdot x_q \geq 1 \tag{4}$$

$$\mathcal{E}_p = m_p + \frac{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot m_q}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \tag{5}$$

Using the already deployed pollution monitoring stations, the simulation errors can be estimated through rough comparison between the currently available real data and the outputs of the simulation model. In this paper, we consider the main cause of the simulation errors, which is in fact the inputs of the physical model. It is worth noting that simulated concentrations that are provided by the physical models are based on weather conditions and pollution emissions. Therefore, we consider multiple possibilities of simulation errors depending on the values of the inputs of the physical model. We first define a set of possible snapshots of simulation errors that we denote by $\mathcal{L}$. For a given $l \in \mathcal{L}$, let m_p^l (respectively $\mathcal{E}_p^l$) be the simulation error (respectively the assimilation error) at point p in the case of l. We formulate in formula 6 the coverage quality at point p as the maximum value among all possible errors. Then, we denote by the function $f(\{\mathcal{E}_p^*,\ p \in \mathcal{P}\})$, the overall coverage quality of a given sensor network topology. Without loss of generality, we consider that f is the mean function. The function f is defined using constraint 7 where $\mathcal{N} = |\mathcal{P}|$.

$$\mathcal{E}_p^* = \ Max_{l \in \mathcal{L}}\{|\mathcal{E}_p^l|\} = \ Max_{l \in \mathcal{L}}\{|m_p^l + \frac{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot m_q^l}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q}|\} \tag{6}$$

$$f = \sum_{p \in \mathcal{P}} \mathcal{E}_p^* / \mathcal{N} \tag{7}$$

Formula 7 together with constraint 6 define the two main constraints of our pollution coverage model. However, in order to get a linear model, we need to linearize constraint 6 by eliminating the fraction and the absolute and maximum functions. Using state-of-the-are linearization techniques, we get the linear form of constraint 6 in constraints 8, 9, 10 and 11 where h_{pq} is an auxiliary variable and H is a big number used to ensure that when $x_q = 0$, h_{pq} is also equal to 0.

$$\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot h_{pq} \geq \ \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (m_q^l + m_p^l), \quad l \in \mathcal{L} \tag{8}$$

$$\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot h_{pq} \geq \ -\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (m_q^l + m_p^l), \quad l \in \mathcal{L} \tag{9}$$

$$-H \cdot x_q \leq h_{pq} \leq H \cdot x_q \tag{10}$$

$$-H \cdot (1 - x_q) + \mathcal{E}_p^* \leq h_{pq} \leq H \cdot (1 - x_q) + \mathcal{E}_p^* \tag{11}$$

C. Network connectivity

We formulate the connectivity constraint as a network flow problem. We first denote by $\Gamma(p),\ p \in \mathcal{P}$, the set of neighbors of a node located at point p. This set can be determined using sophisticated path loss models. It can also be determined using the binary disc model, in which case

$\Gamma(p) = \{q \in P \ where \ q \in Disc(p, R)\}$ where R is the communication range of sensors. Then, we define the decision variables g_{pq} as the flow quantity transmitted from a node located at point p to another node located at point q. In addition, let f_{ip} be the flow quantity transmitted from a mobile node deployed on trajectory i to a relay or sink node located at point p. We suppose that each mobile sensor of the resulting MWSN generates a flow unit in the network, and verify if these units can be recovered by sinks. We formulate constraint 12 to ensure the generation of the flow units by mobile sensors and constraints 13, 14 and 15 to ensure that relay nodes are flow conservative. As a result, sinks recover all the generated units and the MWSN is connected.

$$\sum_{p \in \Gamma(i)} f_{ip} = t_i, i \in \mathcal{I} \quad (12)$$

$$\sum_{i \in \mathcal{I}_p} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} - \sum_{q \in \Gamma(p)} g_{pq} \leq (|\mathcal{I}| + 1) \cdot y_p, p \in \mathcal{P} \quad (13)$$

$$\sum_{i \in \mathcal{I}_p} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} - \sum_{q \in \Gamma(p)} g_{pq} \geq 0, p \in \mathcal{P} \quad (14)$$

$$\sum_{q \in \Gamma(p)} g_{pq} \leq |\mathcal{I}| \cdot r_p, \quad p \in \mathcal{P} \quad (15)$$

III. Simulation results

We perform our simulations on a pollution map corresponding to the 2008 Nitrogen Dioxide (NO_2) concentrations in the Lyon district of La-Part-Dieu, which is the heart of the Lyon City. This Pollution map was generated by an enhanced physical model called SIRANE, which is designed for urban areas and takes into account the impact of street canyons on pollution dispersion. The dataset has been provided by LMFA, which is a research lab specialized in fluid mechanics in the Lyon city, France. We discretize the deployment region with a spatial resolution of 50 meters. We consider as potential positions of nodes all the grid points and as trajectories of mobile sensors all the lines and columns in the deployment grid. The physical model maximum errors have been generated through a comparison between the physical model results and the data of the currently deployed pollution monitoring stations. As a proof of concept, we consider that the error snapshots of the inputs of our model correspond to 10 different levels varying from 10% to 100% of the maximum errors of the physical model.

We evaluate the quality of air pollution monitoring depending on the deployment budget. We depict in Fig. 1 the assimilation error depending on the number of mobile nodes that can be deployed. The assimilation error of the obtained networks is evaluated based on a set of 100 randomly generated simulation error maps. We consider two cases: using uniform and Gaussian random distributions. Results show that the higher the number of mobile nodes, the less the assimilation error in both cases, which is reasonable since reducing the estimation error requires more information and hence more sensors. Also, we notice that the assimilation errors obtained using the Gaussian function are less than those obtained using the uniform distribution. Indeed, with the Gaussian distribution, most errors are located near the mean of the simulation error (which is equal to 0) whereas in the other case, the errors are distributed uniformly.

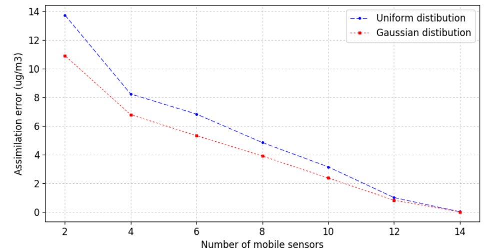

Fig. 1: Coverage quality depending on the deployment budget (number of mobile nodes).

IV. Conclusion and ongoing work

In this abstract paper, we target the deployment issue of mobile sensor networks and propose a mixed integer programming model allowing to improve air pollution monitoring through the correction of physical models. Our main idea is to define an appropriate coverage formulation for pollution data assimilation and then derive an optimal deployment model using integer linear programming. We applied our model on a dataset of the Lyon City, France in order to assess the impact of the deployment budget on the quality of pollution monitoring. We are currently working to extend our model while taking into account more coverage constraints like the calibration of sensor nodes and the drift in sensor measurements.

References

[1] M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks *et al.*, "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks," *Atmospheric Environment*, vol. 70, pp. 186–203, 2013.

[2] A. Boubrima, W. Bechkit, and H. Rivano, "Optimal wsn deployment models for air pollution monitoring," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2723–2735, 2017.

[3] ——, "Error-bounded air quality mapping using wireless sensor networks," in *Local Computer Networks (LCN), 2016 IEEE 41st Conference on*. IEEE, 2016, pp. 380–388.

[4] M. Rebai, H. Snoussi, F. Hnaien, L. Khoukhi *et al.*, "Sensor deployment optimization methods to achieve both coverage and connectivity in wireless sensor networks," *Computers & Operations Research*, vol. 59, pp. 11–21, 2015.

[5] V. Roy, A. Simonetto, and G. Leus, "Spatio-temporal sensor management for environmental field estimation," *Signal Processing*, vol. 128, pp. 369–381, 2016.

[6] W. Du, Z. Xing, M. Li, B. He, L. H. C. Chua, and H. Miao, "Sensor placement and measurement of wind for water quality studies in urban reservoirs," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 3, p. 41, 2015.