Wireless Sensor Networks in Smart Cities:
The Monitoring of Water Distribution Networks Case

RONG DU

Licentiate Thesis
Stockholm, Sweden 2016
Abstract

The development of sensing and communication technologies such as wireless sensor networks (WSNs) are making it possible to monitor our cities and living environments. Due to the small size of the sensor nodes, and their capabilities of transmitting data remotely, they can be deployed at locations that are not easy or impossible to access. For example, WSNs can be used to monitor water distribution networks (WDNs), the structural health of buildings, or the traffic of urban roads. Among the applications of WSNs for smart cities, monitoring WDNs is one of the most important for our health and well being. The accurate and real-time monitoring of WDNs can also reduce the waste of water from pipeline leakage. This thesis studies WSNs for smart cities with specific reference to one of the most vital infrastructures, WDNs.

The design of WSNs for monitoring WDNs faces major challenges. Generally, WSNs are resource-limited because most of the sensor nodes are battery powered. Thus, their resource allocation/arrangements has to be carefully controlled. The pipelines of WDNs are mostly buried underground and thus it is expensive to replace the sensor nodes once their energy expires. The thesis considers two prominent problems that occur when WSNs are deployed to monitor WDNs: scheduling the sensing of the nodes of static WSNs, and sensor placement for mobile WSNs. These studies are reported in the thesis from three published or submitted papers.

In the first paper, it is considered how to schedule the sleep/sensing for each sensor node to maximize the whole WSNs lifetime while guaranteeing a monitoring performance constraint. The scheduling problem is formulated as a network lifetime maximization with cardinality constraint. The decision variables are the sleep/sensing states of each sensor node. Given that they are binary variables, the problem is a binary optimization for which finding the optimal solution is hard. Thus, the problem is transformed into a more tractable energy balancing problem. A novel solution algorithm based on dynamic programming is proposed for such a transformed problem. It is proved that this algorithm finds one of the optimal solutions for the energy balancing problem by a low complexity procedure.

In the second paper of the thesis, the fundamental question of how the energy balancing problem approximates the original scheduling problem is addressed. It is proved that the original scheduling problem is equivalent to a maximum flow problem with a cardinality constraint, which allows analyzing the relation of energy balancing with lifetime maximization from the maximum flow perspective. It is shown that even though these two problems are not equivalent, the gap of the two problems is small enough. Thus, the proposed algorithm that solves the energy balancing problem can find a good approximation solution for the original scheduling problem.

Whereas the first two papers were concerned with static WSNs to monitor WSNs, the second part of the thesis considers the use of mobile sensor nodes. Here, the limited resource is the number of available such mobile nodes due to their cost. The mobility of the sensor nodes allows them providing more information than the static ones. Thus, they are released into the WDN once pollution events are detected to perform a more accurate monitoring and avoid excessive use of anti-pollutants. Due to that the mobile
sensor nodes can only move along the pipeline with the water, a stochastic mobility model
for the mobile sensor nodes is formulated. To maximize the monitoring coverage and
secure the largest part of the population served by the water networks, an optimization
problem for determining the releasing locations for the mobile sensor nodes is formulated.
Due to the integer decision variables, the problem is combinatorially hard. An approximate
solution algorithm based on submodular maximization is proposed. The performance of the
algorithm in terms of optimality is investigated.

The investigations of this thesis show that the energy balancing approach is appealing
to prolong network lifetime. Regarding the case with mobile sensor nodes, the thesis
shows that the releasing location of the sensor nodes plays an important role in the
monitoring performance. Although there are various WSN applications for smart cities,
a common characteristic of such applications is that the area to be monitored usually has
a network structure. For example, the edges and vertices of WDNs could be the pipelines
and junctions; the edges and vertices of buildings could be the corridors and rooms; the
dges and vertices of urban traffic networks could be the roads and junctions. Therefore,
the studies of this thesis have the potential to be generalized for several IoT scenarios in
smart cities.

**Keywords**: Integer Programming, Nonconvex Optimization, Network Lifetime, Dy-
amic Programming, Submodular Maximization, Resource Allocation, IoT, Water Distri-
bution Networks, Smart Cities.
Acknowledgments

First of all, I would like to express my sincere appreciation towards my supervisor Associate Professor Carlo Fischione and Associate Professor Ming Xiao, for their constructive supports and guidance in the last two years. I would also like to thank Dr. Lazaros Gkatzikis for his patient guidance and fruitful discussions during my first year in KTH. Working with them has allowed me to develop me knowledge substantially.

I would offer my thanks to the people of Automatic Control department for building a harmonic and funny environment. Especially, I thank Riccardo Sven Risuleo, Miguel Ramos Galrinho, and Sebastian Hendrik van de Hoef for providing help in courses, especially in the first few months when I start my PhD in KTH; Alexandros Nikou, Robert Mattila, Pedro Miguel Otao Pereira, Xinlei Yi, Christos Verginis and Manne Henriksson for interesting discussions; Yuzhe Xu, Hossin Shokri Ghadikolaei, Sindri Magnusson, José Mairton Barros da Silva Jr., and Xiaolin Jiang for supportive comments, suggestions, and helps. I am also grateful for the assistance and support from the administrators at Automatic Control Department: Anneli Ström, Hanna Holmqvist, Karin K. Eklund, Gerd Franzon, Silvia Cardenas Svensson.

Finally, I also want to thank my parents and grandparents for their love and encouragements. I am deeply grateful to the most important one in my life, Yuanying, for her understanding, support and love. Thank you for making my life a sunny day.

Rong Du
Stockholm, May 2016
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xv</td>
</tr>
<tr>
<td>List of Acronyms</td>
<td>xvii</td>
</tr>
</tbody>
</table>

## I Thesis Overview

### 1 Introduction

1.1 Background

1.1.1 Motivations

1.1.2 Static WSN for water distribution network monitoring

1.1.3 Mobile sensor nodes for water distribution network monitoring

1.2 Problem formulation

1.2.1 Example 1: Lifetime maximization

1.2.2 Example 2: Coverage maximization

1.3 Outline and contribution of the thesis

1.3.1 Water distribution network monitoring with static WSNs

1.3.2 Water distribution network monitoring with mobile sensor networks

1.3.3 Contributions not covered in the thesis

1.4 Conclusions and future works

1.4.1 Conclusions

1.4.2 Future works

### 2 Preliminaries

2.1 Graph representation of WSNs

2.2 Compressive sensing

2.3 Submodular maximization
List of Figures

1.1 Pipeline monitoring by wireless sensor networks .................................. 5
1.2 A prototype of mobile sensor for monitoring water distribution network ...... 7
1.3 Optimization problems considered in the thesis ..................................... 10
1.4 Network lifetime against transmission range ......................................... 12
1.5 Network lifetime against transmission range ......................................... 13
1.6 Monitoring performance against the number of static sensor nodes .......... 14

A.1 Wireless sensor networks in water distribution network systems ............... 26
A.2 Data gathering schemes based on compressive sensing ......................... 29
A.3 Sensor node activation procedure ....................................................... 38
A.4 The case when sensor nodes are not strictly in a line ............................ 40
A.5 Monitoring performance against $M_{cs}/N$-ratio .................................. 41
A.6 Network lifetime against transmission range ....................................... 42
A.7 Network lifetime against transmission range ....................................... 43
A.8 Network lifetime against probability of sensor node failure ................... 45

B.1 An example to transform a network with vertex capacity to a network with edge capacity ................................................................. 66
B.2 A subnetwork that contains a backward arc ........................................ 68
B.3 Network lifetime against initial nodal energy ...................................... 74
B.4 Network lifetime against transmission range ....................................... 76
C.1 Mobile sensor network for monitoring water distribution network ........... 81
C.2 Water distribution networks for simulations ........................................ 96
C.3 Monitoring performance against the number of static sensor nodes .......... 97
C.4 Monitoring performance against the number of mobile sensor nodes ......... 98
C.5 Approximation ratio of different greedy algorithms ............................... 99
C.6 Monitoring performance and computational time of different greedy algo-
  rithms ................................................................................................. 100
List of Tables

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The Application scenario of the optimization problem considered in the included papers of the thesis</td>
<td>11</td>
</tr>
<tr>
<td>B.1</td>
<td>Major notations used in the paper</td>
<td>58</td>
</tr>
<tr>
<td>B.2</td>
<td>Comparison of different methods for multi-dimensional knapsack problem</td>
<td>62</td>
</tr>
<tr>
<td>B.3</td>
<td>Gap between lifetime achieved by energy balancing (Algorithm 4) and maximum lifetime (exhaustive search)</td>
<td>74</td>
</tr>
</tbody>
</table>
List of Acronyms

CDC Compressive Data Collection  
CDG Compressed Data Gathering  
CS Compressive Sensing  
CSF Compressed Sparse Function  
DCT Discrete Cosine Transform  
EDAL Energy-efficient Delay-aware Algorithm  
MECDA Minimum Energy Compressed Data Aggregation  
MDK Multi-dimensional Knapsack  
WDN Water Distribution Network  
WSN Wireless Sensor Network
Part I

Thesis Overview
Chapter 1

Introduction

The ever-reducing cost of Wireless Sensor Networks (WSNs) is allowing to embed them everywhere to monitor and control virtually any space and environment and to form the so-called Internet of Things or Internet of Everything. For example, temperature and humidity of rooms in smart buildings can be easily monitored and controlled by WSNs [1, 2] to provide a comfortable and environmental-friendly living and working condition. WSNs monitor road traffic [3, 4] to provide information for drivers for a better route planning, congestion avoidance, and safer driving. Vibrations in bridges and towers can be monitored to ensure the structural health of the building [5–7]. Water qualities and pipeline leakages in water distribution networks (WDNs) can be better monitored by WSNs, to ensure the drinking water of citizens be clean, and to avoid wasting water due to the leakages. All these technological systems are enabled by the communication infrastructure of WSNs and together form smart cities and environments [8].

In WSN monitoring systems, resources such as the number of available sensor nodes and the energy (battery) of sensor nodes, are limited. Thus, they must be carefully allocated for the efficient operation of WSN systems. For example, we can arrange the placement of the sensor nodes so to have a desired area adequately covered; we can schedule the message transmission and transmit power of sensor nodes to reduce message collisions and save energy; we can arrange the sleeping of sensor nodes to save energy. Thus, the placement and scheduling problems in WSNs are particular important. Moreover, the applications in smart cities with large and complex monitoring areas introduce constraints, such as connectivity constraints and integer constraints on decision variables. These constraints make the allocation problems interesting and challenging.

This thesis studies one of the monitoring system mentioned above, i.e., the monitoring of WDNs. An important characteristic of such system is that, the monitored area is usually underground; it is hard to get access in it to replace the sensor nodes, or recharge their batteries after they expire. Thus, the lifetime of the wireless sensor network (WSN) is an important metric for monitoring WDNs. To increase the WSN lifetime, a common way is the sleep/awake mechanism for sensor nodes. The problem of arrangement of node’s battery, i.e., the scheduling of sleep/awake of sensor nodes to prolong WSN lifetime is studied in this thesis. Another important metric is the monitoring performance, such as
the coverage of the WSN, the average time for events detection, and the accuracy of measurements. To maximize these metrics, a solution is to deploy the sensor nodes at strategic locations. However, benefits from the new prototype of mobile sensor nodes for water monitoring, using mobile sensor networks to monitor WDN becomes very appealing. The benefits of the mobility of the sensor nodes is that we may get a better coverage, or detection time, with less sensor nodes. Thus, the thesis also studies the allocation of mobile sensor nodes to different releasing points to improve the monitoring performance.

1.1 Background

In this section, we give an overview on WSNs for WDN monitoring.

1.1.1 Motivations

WDN is an important infrastructure in modern cities, since it relates the daily water usage of residents. However, it faces at least two major threats, i.e., contaminations and leakages.

Water in pipeline networks can be easily polluted by chemical or biological contaminants. Such contaminants can enter the WDN by accident or by malicious action. More importantly, once the contaminants enter the WDN, they can spread with the water flows and affect larger areas. For example, it was reported that, on April 11th, 2014, the tap water of Lanzhou city, China\(^1\), was found polluted by toxic chemical, which affected several million residents. Besides, water terrorist activities may also happen. Once pollution is detected in WDN, some pipelines must be isolated from the unpolluted area, and anti-pollutants need to be dosed into these pipelines. Therefore, sensors that measure flow rate, oxygen level, pH level, etc., and actuators such as pumps, valves, have to be deployed in WDNs for active realtime monitoring and dynamic control.

Water waste due to pipeline leakages is another important issue in WDNs. Drinkable water is a precious resource. The Food and Agriculture Organization of the United Nations has reported that 1.2 billion people live in areas of water scarcity, and the number will increase to 1.5 billion by 2025 \[^9\]. However, water is severely wasted due to pipeline leakages. Just in London, UK, 589 million litres water are lost due to leakage per day, which is about 1/4 of its overall daily water supply \[^10\]. To fight against water scarcity, some monitoring systems have deployed sensors that monitor flow rate, vibrations, acoustic, etc., on the pipelines to detect and locate the leakages. Projects, such as Hydrobionets\(^2\) and TEVA-Spot \[^21\], have studied the monitoring of water distributions network with WSNs.

1.1.2 Static WSN for water distribution network monitoring

WDN monitoring is important to avoid contaminations and leakages. For a better understanding on the WSN design, the Battle of the Water Sensor Networks (BWSN) \[^11\]

\(^1\)http://www.bbc.com/news/world-asia-27002602
\(^2\)http://www.hydrobionets.eu/
Figure 1.1: Pipeline monitoring by wireless sensor networks. The water distribution network consists of pipelines and junctions. The sensor nodes are placed along the pipelines to monitor the parameters such as flow rate, pipeline vibrations, oxygen level, and pH level of the water, among others. The measurements are sent to the sink nodes, which are located at the junctions, and further transmitted to the public network, such that city managers, water supplier companies, and citizens could access the data.

was held in 2006 among several research groups. An important studied problem in WDN monitoring is the sensor placement problem. The location of sensor nodes could be optimally designed depending on different objectives, such as the monitored areas, the detection time, the populations of the monitored areas, etc. [12–14]. Since the sensor nodes can only be deployed at some candidate locations, such as the junctions of pipelines due to inaccessibility of the pipelines, the optimization problems are usually formulated as integer optimization. The optimal solution of such problems is difficult to achieve, especially due to the large size WDN. The result is that the optimal solutions cannot be achieved and are often approximated. The sensor placement problems are therefore solved by some heuristic algorithms, such as genetic algorithms [15–17], or greedy algorithms, such as the approach based on submodular maximization [18]. Moreover, since the sensor nodes must be connected such that their measurements can reach the monitoring center, the problem of sensor placement with connectivity constraint has been considered in [19].
One way to connect the sensors along the pipelines are by wires. However, such systems suffer from high cost of deployment. The communications of the nodes will break down if the wires are damaged [20]. On the other hand, connecting sensor nodes wirelessly is more robust and affordable. Thus, there have been an increasing interests on WSN for WDN monitoring for the last decade. Benefiting from the study of WSN for WDN, monitoring systems have been already implemented in different cities. In Ann Arbor, a distribution system is built for online contamination monitoring has been built [21]. The system applies probabilistic analysis to assess the WDN. In Singapore, a system called WaterWiSe [22] has been deployed to monitor pipeline leakages. Some sensor nodes are mounted on poles, whereas some other sensor nodes are inserted into the pipes. The data of the sensor nodes are transmitted by 3G modems. In Boston, a system called PipeNet [23] has been developed to detect and localizing leakages based on pressure, acoustic and vibration data. The sensor nodes which are powered by batteries need to transmit there data by short range communication to the gateways nodes, which are powered by the grid. Then the data are transmitted, by long range communication, to the monitoring center. The data transmission of the above mentioned systems are based on electromagnetic waves. On the other hand, a system called MISE-PIPE [24] uses magnetic induction based wireless communications. Coils are winded on the pipelines, and form a magnetic induction waveguide to relay the data. However, the data rate of the system is limited due to the small bandwidth.

### 1.1.3 Mobile sensor nodes for water distribution network monitoring

As opposed to static WSNs, mobile sensor networks (MSNs) [25, 26] consist of sensor nodes that have the capability to move in the monitored area. The mobile sensor nodes can react dynamically based on the measurement of the WDNs. Thus, the monitoring performance may be improved by using fewer sensor nodes [27]. This makes it appealing to monitor WDNs by MSNs.

Mobile sensor nodes for WDN monitoring are working inside pipelines. This remands that they must be water-proof to protect the circuits, and small enough to move inside the pipelines. Therefore, the prototypes of the mobile sensor nodes for WDN monitoring are specially designed. In [28], a prototype called Triopus is designed similarly to an octopus, i.e., each node is equipped with a motor that drives three arms that can attach the sensor node onto the inner surface of pipelines. In [29], a mobile sensor node, PipeProbe, is designed to detect the spatial topology of hidden water pipelines. The node has no motor unit, and its movement is determined by the water flows inside the pipelines. In [30], a mobile sensor node is introduced to conduct continuous measurement as it moves in the flow stream to monitor water quality inside pipelines. The sensor node has no motor unit for space and power saving. Commercial mobile sensor nodes are also available, such as SmartBall [31].

The prototypes of mobile sensor nodes mentioned above greatly support the study of WDN monitoring by MSNs. Due to that the mobile sensor nodes are moving inside pipelines, where GPS is not available, the localization problem is not trivial. A solution is to localize them based on RFID tags [32, 33]. The localization problem is difficult also
1.2 Problem formulation

Figure 1.2: A prototype of mobile sensor for monitoring water distribution network [29]. The sensors are inside a waterproof plastic capsule, whose diameter is approximately 3 cm. The node has no mobile unit and so its physical movement basically follows the water flows. It collects data at different points of the pipeline as it travels in it.

considering that some prototypes of mobile sensor node do not have motor unit, their trajectories are neither available to be scheduled as the case for open-field monitoring [34], nor fixed as the case of using public transport for urban monitoring [35]. Thus, the releasing locations and time of the mobile sensor nodes are usually arranged to improve the monitoring performance [36–38].

1.2 Problem Formulation

In this thesis, we consider WSNs monitoring problems for WDNs that can be posed as instances of the following optimization problem

\[
\begin{align*}
\max_x & \quad f(x_1, x_2, \ldots, x_N) \\
\text{s.t.} & \quad x_i \in \mathbb{Z}^+ \cup \{0\}, \\
& \quad \sum_{i \in M} x_i \geq k, M \subseteq \{1, 2, \ldots, N\}, \\
& \quad (x_1, x_2, \ldots, x_N) \in \mathcal{X},
\end{align*}
\]

where \( x \) here represents the resources, such as available sensor nodes, and their energies. The objective of the problem is to maximize the WSN performance, which could be WSN lifetime, coverage, etc, which depends on the allocation of the WSN resource. The non-negative integers represent the allocation of the WSN resources, such as the number of available sensor nodes. Constraint (1.1c) represents the requirement on allocating the WSN resources, e.g., at least or at most a certain number of nodes or energy must be used. Constraint (1.1d) represents the other communication requirements on the allocation, such as the connectivity of the chosen nodes. The problem is difficult due to the integer constraint, and achieving the optimal solution may be impossible in reasonable time. Thus, in this thesis, we establish some greedy-based algorithm to find the approximate solutions specifically for the scenarios that are briefly described next.
1.2.1 Example 1: Lifetime Maximization

Energy of sensor nodes is an important resource for guaranteeing a long lifetime of WSNs. For the case that node replacement or recharging is impossible, it is desired to have the WSN lifetime as long as possible. An idea is to apply the awake/sleep mechanism, where the sleeping sensor nodes could save energy and the awake sensor nodes could provide good monitoring performance. Thus, the awake/sleep scheduling of the WSNs must be determined carefully.

Consider a WSN $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of sensor nodes, and $\mathcal{E}$ is the set of communication link between sensor nodes. Suppose the energy consumption of an awake node in a timeslot is normalized to be 1, and the consumption of a sleeping node is 0. Let $x_i(t) = 1$ represent sensor node $v_i$ is awake during timeslot $t$, otherwise, $x_i(t) = 0$. The awake/sleep schedule at timeslot $t$ is then represented by $x(t)$. Given the initial normalized energy of the sensor nodes to be $E_i$, the lifetime maximization considered in the thesis can be formulated as follows:

$$\max_{x(t), t=1, \ldots, T} \quad T$$

s.t. \hspace{1cm} \sum_{t=1}^{T} x_i(t) \leq E_i, \forall i, \quad (1.2a)$$

$$\sum_{v_i \in \mathcal{V}} x_i(t) \geq M, \forall i, 1 \leq t \leq T, \quad (1.2b)$$

$$G(x(t)) \text{ is connected}, \forall t, \quad (1.2c)$$

$$x_i(t) \in \{0, 1\}, \forall i, t, \quad (1.2d)$$

where the objective is to maximize the WSN lifetime, Constraint (1.2b) represents the limitation of the energy resource of sensor nodes; Constraint (1.2c) represents at least $M$ number of sensor nodes must be activated in every timeslot to guarantee the monitoring performance; Constraint (1.2d) represents the activated sensor nodes must be connected such that the measurements can reach the sink nodes. In such problem, the WSN resources are the energy of the sensor nodes (Constraint (1.2b)), and the number of sensor nodes (Constraint (1.2c)). Note that, if Constraint (1.2c) is relaxed, the problem can be turned to a maximum flow problem, which can be solved efficiently. However, with the Constraint (1.2c), Problem (1.2) becomes non-trivial and challenging. In the thesis, an energy balancing approach is used to determine the awake/sleep scheduling of the WSNs, which will be described later.

1.2.2 Example 2: Coverage Maximization

Coverage is another important metric for WSNs. In coverage maximization problems, the location of the sensor nodes are determined such that the sensor nodes can cover an area as large as possible. In the thesis, we consider the scenario of water distribution network monitoring by mobile sensor nodes. Static sensor nodes are assumed to be already deployed in the water distribution network. Since the mobile sensor nodes are carried by the water, a
random mobility model is used to characterise the moving path of the mobile sensor nodes. Thus, the coverage of the mobile sensor nodes can be considered as a function depending on the flow rates and the location of the sensor nodes. Due to physical constraints, the mobile sensor nodes can only be released from the junctions of the water distribution network, we denote $x_i$ the number of mobile sensor nodes to be released at junction $v_i$, and $x = [x_1, \ldots, x_K]$ the releasing of the mobile sensor nodes. Thus, the coverage maximization problem can be formulated as follows:

$$
\max_x \sum_i f_i(x_i) \\
\text{s.t. } \sum_i x_i \leq N, \\
x_i \in \mathbb{Z}^+ \cup \{0\},
$$

where $f_i(x_i)$ is the coverage of the mobile sensor nodes that are released at junction $v_i$, and $N$ is the number of available mobile sensor nodes. We will show that the objective function has the submodular property, which allows us to apply a greedy-based algorithm to achieve an approximate solution with performance bound.

### 1.3 Outline and Contribution of the Thesis

This thesis is based and the following publications/submissions:


The thesis considers the WSNs communication resource allocation problems for monitoring WDN. Figure 1.3 shows the types of considered optimizations. In papers [C1], [J1], and [J2], the problems are formulated as binary optimization problems, where the variables represents whether a sensor node is awake or sleep in a timeslot. In paper [C2], the problem is formulated as an integer optimization problem. The variables represent the number of mobile sensor nodes to be released at each junction. Moreover, its objective
Figure 1.3: The nature of the optimization problems included in the thesis have general as well as sub-modularity nature. In the first three papers, [C1], [J1], [J2], the problems can be cast as binary optimizations (BP). In paper [C2], the considered problem is an integer optimization problem (IP), and its objective functions has a submodularity property.

Table 1.1 shows the application scenarios of the optimization problems. In the first part ([C1], [J1], [J2]), we considered a static WSN, of which we wish to maximize the lifetime. The resource here is the batteries of the sensor nodes, and we need to determine which sensor nodes should be working in each timeslot to achieve long lifetime. Since the WSN is static, the connectivity constraints on the working sensor nodes must be satisfied. The cardinality constraints are in two domains: time domain and spatial domain. In time domain, it means that the number of timeslot a sensor node would be working can not exceed its battery capacity; whereas in spatial domain, it means that the number of working sensor nodes must be large enough, such that a good monitoring performance is achieved. In the second part of the thesis, we considered a mobile WSN. The resource here is the number of the available mobile sensor nodes to be used, which is represented by the cardinality constraint.
Table 1.1: The Application scenario of the optimization problem considered in the included papers of the thesis. The first three papers, [C1], [J1] and [J2], considered both connectivity constraint and cardinality constraint of the sensor nodes in static WSNs, whereas [C2] considered only cardinality constraints of the sensor nodes in mobile WSNs.

<table>
<thead>
<tr>
<th>connectivity</th>
<th>cardinality</th>
<th>objective</th>
<th>WSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>lifetime</td>
<td>static WSN</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>lifetime</td>
<td>static WSN</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>lifetime</td>
<td>static WSN</td>
</tr>
<tr>
<td>-</td>
<td>✓</td>
<td>coverage (population)</td>
<td>mobile WSN</td>
</tr>
</tbody>
</table>

1.3.1 Water distribution network monitoring with static WSNs

This part is based on [C1], [J] (Appendix A), and [J2] (Appendix B). We consider a lifetime maximization problem of WDN monitoring with dense sensor networks, where the communication range of each sensor node could be small to save energy. To further save energy, part of the sensor nodes could be in a sleep mode without sacrificing the monitoring performance. This introduces a connectivity constraint and a cardinality constraint on the working sensor nodes, such that the measurements of these nodes could reach the destination, and the monitoring performance is met. Notice that the decision variables, i.e., the awake/sleep of each sensor node in each timeslot, is binary, and therefore the original problem is not trivial. Thus, the original lifetime maximization problem is turned to an energy balancing problem, i.e., in each timeslot, we need to find the group of sensor nodes that satisfy the two constraints and have the largest residual normalized energy. We first propose a dynamic programming based algorithm, called SACC algorithm, with low-complexity to solve the energy balancing problem (in [C1]) under the assumption that the sensor nodes are deployed in a line and have the same transmission range. The assumptions are later relaxed in [J1], such that the optimality of the solution still holds. Figure 1.4 shows some insights on the performance of the proposed SACC algorithm in terms of WSN lifetime. In each time slot, 1/5 of the sensor nodes should be activated for monitoring to satisfy the cardinality constraint. The WSN lifetime achieved by the SACC algorithm is much longer than the ones achieved by the other algorithms.

Further, we study the relationship of the energy balancing problem and the lifetime maximization problem in [J2]. Even though the two problems under consideration are not equivalent. By studying the upper bound of the solution of the lifetime maximization problem, we showed that the lifetime of the solution achieved by the energy balancing problem is very close to the maximum lifetime problem. Thus, the energy balancing approach is a good approximation for the lifetime maximization problem.

Figure 1.5 shows the WSN lifetime achieved by different algorithms. When the transmission ranges of the sensor nodes are the same (see (a)), we can see that the lifetime achieved by the proposed algorithm is close to the upper bound lifetime achieved by a relaxed maximum flow problem, which means that the lifetime achieved by energy
Figure 1.4: Comparison of the WSN average lifetime achieved by the proposed SACC algorithms with other important algorithms, such as CDG [39], CSF [59], CDC [63], and MECDA [61]. The normalized transmission range is the ratio of the nodal transmission range to the length of pipeline, and the larger the ratio is, the denser the WSN is. The sensor nodes initially have 100 unit energy. A sensor node consumes 1 unit energy in a timeslot if it is awake, otherwise it consumes 0 unit energy. In each timeslot, at least one fifth of the total sensor nodes should be awake to guarantee the monitoring performance. The achieved WSN lifetime is much longer than the other algorithms. It is close to 500 timeslots when the normalized transmission range is large enough.

balancing is close to the maximum lifetime. In the case where nodal transmission ranges are different (see (b)), the gap is small when the transmission ranges are large enough, i.e., the WSN is dense. Therefore, the proposed algorithm is a good approximation approach to prolong WSN lifetime.

1.3.2 Water distribution network monitoring with mobile sensor networks

This part is based on [C2] (Appendix C). We consider mobile sensor nodes to improve the WDN contamination detection performance of static WSNs. Since the mobile sensor nodes have no motor units, their movement is randomly determined by the water flow in the pipelines. Therefore, we consider the problem of allocating mobile sensor nodes to selected releasing locations to cover as more population as possible, which is an integer optimization problem, i.e., how many mobile sensor nodes should be released at each junction. The problem can be considered as a sensor placement problem, where the coverage area of each sensor node is stochastic, due to the movement model of the mobile sensor nodes. To solve the problem, we analyze the submodular properties of the sensor releasing problem. Benefiting from these properties, we proposed a greedy based algorithm
1.3. Outline and contribution of the thesis

Figure 1.5: Comparison of the network average lifetime achieved by the proposed SACC algorithms with the other algorithms and an upper bound found by solving a maximum flow problem. The normalized transmission range is the ratio of the nodal transmission range to the length of pipeline, and the larger the ratio is, the denser the wireless sensor network is. The sensor nodes initially have 50 unit energy. A sensor node consumes 1 unit energy in a timeslot if they are awake; otherwise it consumes 0 unit energy. In each timeslot, at least one fifth of the total sensor nodes should be awake to guarantee the monitoring performance. It shows that, if the sensor nodes have the same transmission range, (see (a)), the network lifetime achieved by the proposed algorithm (blue line) is close to the upper bound lifetime (yellow line). The energy balancing is a good approach for lifetime maximization. When the transmission ranges of the sensor nodes are different, (see (b)), the gap of the lifetime achieved by the proposed algorithm is large when the transmission range is comparatively small. See Figure B.4 for more details.

to achieve an approximate solution of which we analyzed the worst case performance. The simulation results also show that the benefits of having additional mobile sensor nodes than using additional static sensor nodes in the monitoring performance, if the releasing location of the mobile sensor nodes are determined carefully, as shown in Figure. 1.6.

1.3.3 Contributions not covered in the thesis

The following publication is not covered in the thesis, but contain related materials and applications:

Figure 1.6: The performance of using only one mobile sensor node compared to that only static sensor nodes are used. 4 static sensor nodes have been deployed in a WDN with 95 vertices and 119 edges. Each vertex supplies the water for a population uniform randomly selected from 50 to 150. It shows the benefit of using mobile sensor nodes over using additional static sensor nodes. Also, it shows that the releasing location of the mobile sensor nodes plays an important role in the monitoring performance.

1.4 Conclusions and Future Works

1.4.1 Conclusions

WDN is a vital infrastructure to monitor in smart cities. In this thesis, we studied the monitoring of WDN by both WSN and MSN. For the WSN part, we considered the WSN lifetime maximization problem for the long-term monitoring to reduce the operation/maintenance costs. A connectivity constraint was introduced to guarantee that the measurements of the sensor nodes could reach the monitoring center, and a cardinality constraint was introduced to guarantee the monitoring performance of the sensor nodes. Then, the lifetime maximization problem arranges the sleep/awake schedules of the sensor nodes while ensuring the two constraints be always satisfied. We proposed to solve the lifetime maximization problem by energy balancing sub-problems. We showed that, if the sensor nodes are deployed on a line and have the same transmission range, the lifetime achieved by energy balancing is close to maximum lifetime, based on the study of the upper bound of the maximum lifetime. The simulation results also shows that the lifetime achieved by the proposed algorithm can extend approximately 20% of the lifetime achieved by the other state-of-art algorithms. Thus, the energy balancing approach is efficient and effective for the lifetime maximization problem in our special case.

For the MSN part, we considered using existing static WSN to monitor water contaminations. Once a contamination event is detected, the mobile sensor nodes are
1.4. Conclusions and future works

released into the WDN to have a more accurate monitoring of the pollutant distribution. Since the mobile sensor nodes have no motor units, their mobility is characterized by a stochastic model. The submodular property of the sensor releasing problem was studied, based on which two greedy algorithms were proposed to achieve an approximate solution for the problem. One algorithm has a better monitoring performance, whereas its computational complexity is high. Thus, it suits for WDNs with small size. The other one has low complexity with an acceptable performance. Therefore, it suits for WDNs with larger size. The simulation results showed the importance of release the mobile sensor nodes at the right place, since the improvement is negligible if they are released randomly. Moreover, the results showed the benefits of using mobile sensor nodes against using additional static sensor nodes in terms of monitoring performance.

1.4.2 Future works

There are several interesting research directions based on the work of this thesis. Some of them are listed as follows:

- **Generalize the equal-transmission range assumption for the energy balancing problem of static WSN:** In [J1], we showed that when the equal-transmission range assumption holds, the property on optimality of the proposed SACC algorithm holds. However, in the general case where the sensor nodes have different transmission range, this conclusion may not hold. In this general case, one may still use a dynamic programming approach to solve the energy balancing problem; however, the time-complexity may be high. Another way is to consider some greedy approaches. It is interesting to study the optimality or the performance bound of these greedy approaches. The performance of using such greedy approaches to find the solution of the network lifetime problem is also important. Besides, we showed that the network lifetime based on solving the energy balancing problem is very close to the maximum network lifetime, when the sensor nodes have the same transmission range. However, if this equal-transmission range assumption does not hold, the gap might be large, as could be seen in Figure B.4 (c). Thus, it is also interesting to study the relationship of the energy balancing problem and the lifetime maximization problem under the case where nodes have different transmission range.

- **Network lifetime maximization with adjustable transmission range:** In the network lifetime maximization problem considered in the thesis, the transmission range of all the sensor nodes are considered fixed to a minimum value for energy saving. However, it is also interesting to consider the case where the transmission range of the nodes is adjustable, with the same cardinality constraint and the connectivity constraint. The adjustable transmission range to be considered could be continuous values, or could be discrete values, representing some predefined power levels. In this case, the network lifetime maximization problem becomes more complicated. Its optimal solution may be hard to achieve. Then, one could study finding the upper bound or lower bound of the network lifetime. One could also work on the greedy algorithms for the problem, and analyze its performance. The problem on whether
energy balancing is a good approach for lifetime maximization problem is still of relevant interest.

- **Dynamic mobile sensor releasing:** In the WDN monitoring with MSN part, we only consider the releasing locations of the mobile sensor nodes. However, the releasing time of them could also be scheduled. One may release all the mobile sensor nodes at the same time to have the fastest measurements, as what we did in the thesis. It is also possible to release the mobile sensor nodes at different times, and the releasing is determined by the measurement of the previous released nodes. For this problem, the idea of adaptive submodular maximization might be useful for an online algorithm. One may also consider proposing an offline algorithm based on dynamic programming. The performance of the algorithms should also be studied.

- **Applying the results on other monitoring applications:** The thesis considers the application of WDN monitoring; however, it is interesting to see whether the results could be used in other monitoring applications. For example, the results of the MSN part could be useful in the case of monitoring rivers. The releasing locations of the sensor nodes could be any vertex and edge of the water network. The problem could be more interesting if the mobile sensor nodes have motor units. In this case, one could either turn off the motor and let the mobile sensor node flow with the water to save energy, or he could control the motor to let the node move upstream.

- **Other issues:** Some other interesting topics include the following: generalizing the network lifetime maximization problem for the monitoring of a 2-dimensional free space, or an area that could be characterized by a grid network; consider the optimal releasing of the mobile sensor nodes under the case where the water flows are time-varying.
This chapter gives some essential elements of the background theory used in the thesis. Section 2.1 briefly describes how a WSN is represented by a graph. Section 2.2 summarises the basics of compressive sensing that are used in this thesis. Section 2.3 recalls the submodular maximization problem and a common greedy algorithm to solve it.

2.1 Graph Representation of WSNs

WSNs are usually modelled as a graph, where the sensor nodes are represented by the vertices in the graph. If the data of a sensor node \( v_i \) can directly reach a sensor node \( v_j \), then there is a directed edge \( \langle v_i, v_j \rangle \) in the graph. In a data gathering process, the measurements of the sensor nodes must reach the sink node at the end. Thus, the routings of the sensor nodes form a tree, of which the sink node is the root. In the routing tree, the sensor nodes transmit their measurements to the sink node in a multi-hop way. If node \( v_i \) is the next hop of \( v_j \) in the routing tree, then \( v_i \) is the parent node of \( v_j \), and \( v_j \) is a child node of \( v_i \). Then, the sensor nodes that help other nodes in relaying data are also called relay nodes, and the sensor nodes have no child nodes are called leaves nodes in the routing tree.

2.2 Compressive Sensing

Consider a WSN with \( N \) sensor nodes, whose measured data is represented by a vector \( d = [d_1, d_2, \ldots, d_N]^T \). Consider the following compressed data gathering (CDG) scheme: 1) every node \( i \) multiplies its data \( d_i \) with a random vector of size \( M \ll N \), \( \phi_i \), to be its local coded data \( \phi_i d_i \). 2) For a leave node \( i \) in the routing tree, it transmits this local coded data \( d^\text{send}_i = \phi_i d_i \) to its parent node; For a parent node \( i \), it receives all the coded data of his children nodes, \( d^\text{send}_j, \forall v_j \in N_+(v_i) \), where \( N_+(v_i) \) is the set of children nodes of node \( v_i \). Then it transmits the vector sum of its local coded data and the received coded data from its children, to be \( d^\text{send}_i = \phi_i d_i + \sum_{v_j \in N_+(v_i)} d^\text{send}_j \).

In so doing, all the sensor nodes transmit data of size \( M \), which balances the energy consumptions of the sensor nodes in data gathering. Suppose the sink node receives a
vector $y$, then it can be represented by

$$y = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_N \end{bmatrix} \begin{bmatrix} d \end{bmatrix} = \Phi d.$$  

To recover $d$ from $y$, since $M < N$, the problem is ill-posed. However, in dense sensor networks, where the measurements of the sensor nodes have strong spatial correlations, the data could be considered as sparse [39], which means there is a basis $\Psi$, such that $d = \Psi x$, where $x$ is the corresponding coefficients of $d$ in the basis $\Psi$, and $x$ has $K \ll N$ non-zero coefficients (or the other coefficients are close to zero). Then, the sink node could reconstruct $x$ through solving the following minimization problem [39, 40]:

$$\min_x \|x\|_{l_1} \quad \text{s.t.} \quad y = \Phi \Psi x.$$  

Given the reconstructed $\hat{x}$, the sink node can recover the original data $d$ by $\hat{d} = \Psi \hat{x}$.

In practice, $\Psi$ could be a discrete cosine transform or wavelet basis, and $\Phi$ could be a matrix randomly generated at the sink node. By applying compressive sensing in data gathering, the energy consumptions of the sensor nodes are reduced and balanced, and the sink node could recover within a desired accuracy the measurements of the sensor nodes. Therefore, compressive sensing often considered in many monitoring applications [4, 39].

### 2.3 Submodular Maximization

**Definition 1** (Submodular set function). Consider a set function $f : 2^U \rightarrow \mathbb{R}$, where $U$ is an infinite nonempty set. The set function $f$ is called submodular on the ground set $U$ [41] if, whenever $A \subseteq B \subseteq U$ and $\alpha \in U \setminus B$, it holds that $f(A \cup \{\alpha\}) - f(A) \geq f(B \cup \{\alpha\}) - f(B)$.

Submodular set functions have the following property:

**Proposition 1.** If $f$ and $g$ are submodular set functions on the same ground set $U$, then $\alpha f + \beta g$ is submodular for any $\alpha, \beta \geq 0$.

A submodular maximization problem is formulated as follow:

$$\max_{X \subseteq U} f(X),$$

$$\text{s.t.} \quad |X| \leq M,$$

which aims at finding a subset of $U$ with at most $M$ elements, to maximize the submodular function. Generally, the problem (2.1) is NP-hard [41, 42]. A common greedy algorithm to give an approximate solution to the problem is given in [41]. The algorithm starts with setting $X_0 = \emptyset$, and it iteratively updates as follows:

$$X_{t+1} = X_t \cup \{\arg \max_{\alpha} f(X_t \cup \{\alpha\})\},$$

until $|X_t| = M$. The greedy algorithm has the following property:
2.3. Submodular maximization

Theorem 1 ([43]). Consider feasible submodular maximization problem (2.1). If $f$ is submodular and monotone, the greedy algorithm gives a $(1 - 1/e)$-approximation for the problem. In another words, if $f^*$ is the optimal solution of the problem, and $f^G$ the greedy solution by the algorithm, then

$$\frac{f^* - f^G}{f^* - f(\emptyset)} \leq e^{-1}.$$  

Moreover, in [43] it is shown that the approximation factor $1 - 1/e$ is the best approximation ratio we can possibly achieved in polynomial time.
Part II

Included Papers