HOSTED BY

Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 1 (2014) 83–94

journal homepage: www.elsevier.com/locate/inpa

Remote monitoring cost minimization for an unreliable sensor network with guaranteed network throughput

Xu Xu, Weifa Liang *, *Zichuan Xu*

Research School of Computer Science, The Australian National University, Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 16 June 2014
Received in revised form
16 October 2014
Accepted 20 October 2014
Available online 5 November 2014

Keywords:
Remote sensing monitoring
Wireless sensor networks
Service cost minimization
Constrained optimization
Algorithm design and analysis

ABSTRACT

In this paper we consider a link-unreliable remote monitoring scenario where the monitoring center is geographically located far away from the region of the deployed sensor network, and sensing data by the sensors in the network will be transferred to the remote monitoring center through a third party telecommunication service. A cost associated with this service will be incurred, which will be determined by the number of gateways employed and the cumulative volume of data successfully received within a specified monitoring period. For this scenario, we first formulate a novel constrained optimization problem with an objective to minimize the service cost while a pre-defined network throughput is guaranteed. We refer to this problem as the throughput guaranteed service cost minimization problem and prove that it is NP-complete. We then propose a heuristic for it. The key ingredients of the heuristic include identifying gateways and finding an energy-efficient forest of routing trees rooted at the gateways. We also perform theoretical analysis on the solution obtained. Finally, we conduct experiments by simulations to evaluate the performance of the proposed algorithm. Experimental results demonstrate the proposed algorithm outperforms other algorithms in terms of both the service cost and the network lifetime.

1. Introduction

Wireless sensor networks (WSNs) have been used in many application domains, including industrial control, home automation, military sensing, asset tracking, habitat monitoring, soil analysis, and so on [19,14]. Traditionally, the base station is deployed at a pre-defined strategic location in the network and data generated from sensors is transmitted to the base station through multi-hop relays using low-power radios such as IEEE 802.15.4. In this paper we consider a remote monitoring scenario where a homogeneous sensor network with unreliable wireless communications is deployed in a region that is geographically different from the one of the monitoring center. Such a scenario is driven by many real applications. For example, a farmer who lives in Canberra deploys sensor networks on his remote farms to monitor crops growth, where the farms are geographically located in other states such as Queensland, Tasmania, and Western Australia. The sensing data generated by the sensors from the sensor networks is to be sent back to the monitoring center in Canberra in real-time for further processing and decision-making. Clearly, the limited transmission range of low-power radio

* Corresponding author. Tel.: +61 2 61253019; fax: +61 2 61250010.

E-mail addresses: grace.xu@anu.edu.au (X. Xu), wliang@cs.anu.edu.au (W. Liang), edward.xu@anu.edu.au (Z. Xu).

Peer review under the responsibility of China Agricultural University.

http://dx.doi.org/10.1016/j.inpa.2014.10.001

by the sensor and the long distance between the monitoring center and the deployed sensor networks (at least several hundreds of kilometers away) make the traditional multi-hop data transfer paradigm inapplicable. Instead, the farmer must employ the long-distance data transfer service provided by a third party telecommunication company. As illustrated in Fig. 1, some sensor nodes in the sensor network must register themselves to the service so that they are able to access to the third party network and all sensing data must be relayed through them to the remote monitoring center. We refer to such sensor nodes as the *gateways*.

It requires high-bandwidth radios on the gateways to communicate with the facilities of the third party network, such as 3G or 4G radio communications. Due to the high energy consumption of 3G/4G radios [2], gateways usually work in the following two modes. One is to utilize low-power radios, e.g., IEEE 802.15.4, for data communication within the sensors in the sensor network. Data transmission over low-power radios is unreliable and thus causes data loss. The other is to adopt high-bandwidth radios to transmit the collected data to a third party network. We assume that such communication is reliable as it is beyond the control of the sensor network owner. However, this does incur a service cost that must be paid by the sensor network owner [23]. We thus assume that every sensor node is equipped with dual-radios (a high-bandwidth radio and a low-power radio) and can be chosen to act as a gateway. Once the gateways are chosen, the other sensors will forward their sensing data to the gateways for further relay.

The service cost incurred for transmitting the collected data over the third party network is usually comprised of a *fixed cost* for a data *quota* as well as a *penalty cost* for any exceeding data usage beyond the data quota, and charged on a fixed period basis (e.g., monthly for mobile plans), referred to as the *charging period*. The volume of data successfully received by the monitoring center within a charging period is defined as the *network throughput*. In most sensing monitoring applications, different users usually have different monitoring quality requirements. For example, the monitoring center must receive at least a certain percentage of all generated data in a given period. In this paper we refer to the given percentage of all generated data within a charging period as the *network throughput requirement*. If higher throughput is required, a large volume of sensing data has to be sent via the third party network and a more expensive service cost will be incurred, and vice versa. Our objective in this paper thus is to minimize the service cost, subject to a specified network throughput requirement. In addition, due

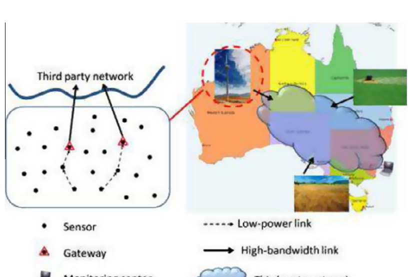


Fig. 1 – An overview of the remote monitoring.

to relaying data and the large energy consumption on high-bandwidth radio communications, gateways consume their batteries faster than that of other nodes. To balance the energy consumption among sensor nodes, gateways are required to be periodically rotated in order to prolong the network lifetime, where the network lifetime is defined as the first node failure due to its battery expiration [6].

To solve this constrained optimization problem, we need to jointly determine the number of gateways employed and the data routing structure. This is because (i) the number of gateways plays an important role in the service cost. If the number of gateways is small, quotas might be severely overused at gateways and expensive penalties will be applied. On the other hand, if the number of gateways is large, the fixed cost would be high, and a large fraction of data quotas at most gateways would certainly be under-utilized and will be wasted. A fine tradeoff must be explored to make the best use of data quota and to avoid or minimize the penalties. (ii) The volume of data relayed by each individual gateway depends on not only the number of sensors forwarding their data to the gateway but also the end-to-end reliability between each of these sensors and the gateway. Sensors are to be allocated to gateways and a set of routing trees rooted at the gateways are to be built to span all sensors such that the sum of the service cost of all gateways is minimized, while the expected volume of data relayed by the gateways meets the network throughput requirement.

Since we consider the link-unreliable wireless sensor networks, the volume of received data through routing trees and the resultant cost are the expected results. That is, the actual received data volume may not meet the pre-defined throughput requirement, and also the actual service cost could exceed the expected one. The probability analysis on these cases will be conducted to show the quality of the solution.

Our main contributions in this paper are as follows. We first formulate a novel constrained optimization problem – the throughput guaranteed service cost minimization problem and show its NP-completeness. We then propose a heuristic, which includes identifying gateways dynamically and finding an energy-efficient forest of routing trees rooted at the gateways. We also conduct theoretical analysis on the performance of the obtained solution. We finally perform experiments by simulations to evaluate the performance of the proposed algorithm and study the impact of different constraint parameters on its performance. Experimental results demonstrate that the proposed algorithm outperforms other algorithms in terms of the service cost and the network lifetime. To the best of our knowledge, this is the first time that the problem of minimizing the service cost for remote monitoring scenarios is considered, and a feasible solution is provided.

The remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 introduces the system model, notions, the problem definition, and proves that the problem is NP-complete. Section 4 proposes a heuristic and Section 5 provides theoretical analysis on the quality of the solution. Section 6 conducts extensive experiments to evaluate the performance of the proposed algorithm, and Section 7 concludes the paper.

2. Related work

Data gathering is one primary function of wireless sensor networks, which has been extensively explored in the past decade [17,8,10,11,22,25]. Previous studies on this problem can be classified into three categories based on different routing protocols adopted [4]: flat-based routing, hierarchical-based routing, and location-based routing. (i) In the flat-based routing, all sensors are assigned with equal capabilities and play the same roles. This type of routing protocols includes directed diffusion [9], rumor routing [5], random-walks-based routing [16], and so on. (ii) In the hierarchical-based routing, sensors are clustering into different clusters, the network has intra-cluster and inter-cluster layers, and sensors serve as different roles (cluster heads and cluster members) [8,22,21]. Cluster heads gather data from their members and relay data to the base station, while cluster members only communicate with each other in the same cluster. (iii) In the location-based routing, the locations of sensors are not given but can be obtained either through information exchanges between neighboring nodes or GPS [26]. Our paper falls into category (ii). The idea of periodic gateway identification in this paper is similar to the cluster head selection and rotation scheme in LEACH [8]. However, LEACH selects cluster heads by assigning each sensor a random probability while our proposed algorithm chooses sensors with relatively high residual energy as gateways. Besides, in our algorithm, the routing trees rooted at gateways are built to guarantee that the expected volume of data collected from these trees meets the network throughput requirement, based on the assumption that wireless links are not reliable in the sensor network. Furthermore, unlike most existing works that focus on the network lifetime, this paper focuses on minimizing the service cost by transmitting data to the remote monitoring center.

The dual-radio wireless sensor network has been studied recently. Most studies in literature focus on energy conservation [18,12,7]. This is directly driven by the fact that high-bandwidth radios (e.g., IEEE 802.11b/g) are more energy efficient in data transmission yet costly in idling energy consumption and start-up overhead [15], compared to low-bandwidth radios (e.g., IEEE 802.15.4). The main challenge of these works is to find a fine tradeoff by minimizing the amount of time spent by the high-bandwidth radio in idle status while using the lower-power radio as a paging and control channel for resource discovery and mobility support [15]. Stathopoulos et al. [18] considered dual-radio, dual-processor nodes in WSNs which provide both low-energy operations as well as increased computational performance and communication bandwidth. In such systems, low-power radios always remain vigilant while high-bandwidth radios are to be triggered by the applications. Because the high-bandwidth radio works at a low operating duty cycle to conserve energy, end-to-end paths do not always exist. They proposed a topology control mechanism which uses vigilant low-power radios to selectively wake up the mostly-off high-bandwidth radios for bulk traffic. The mechanism reduces energy consumption while incurring only a moderate increase in application latency. Lymberopoulos et al. [12] considered to opportunistically use two or more types of radios to achieve energy efficient design of a sensor platform. They concluded that high bandwidth radios are energy efficient only when the amount of data to be sent is large. Different from these studies, this paper adopts the dual-radio model not for the purpose of saving energy, but for remote data transmission. In the above works, two types of radios are both working for data transmission within the network. Whereas in our paper, only the low-bandwidth radio is used within the sensor network for data transmission, while the high-bandwidth radio is used to communicate with the third party facilities outside the sensor network. We investigate the cost incurred by data transmission over the high-bandwidth radios, which is a significant departure from existing studies.

The remote monitoring scenario with dual-radio platform was studied in our previous work [23]. Assuming that the number of gateways is given and not all generated data needs to be collected, Xu et al. studied the problem of throughput guaranteed network lifetime maximization. They analyzed the energy cost models of the two radios and proposed a heuristic to assign the gateways and build routing forest for energy-efficient data collection, which does not necessarily include all deployed nodes. This work differs from [23] in that we consider link unreliability that compromises the network throughput. The routing forest establishment strategy in [23] is not applicable to this scenario, because it does not take into account the data loss during data transmission. Rather, a new algorithm is to be developed so that the volume of data collected through the routing forest is no less than the required amount. The study of this paper is an extension of the work in paper [24], by providing theoretical analysis on the quality of the solution obtained and extensively experimental evaluations.

3. Preliminaries

We consider a dual-radio wireless sensor network $G = (V, E)$ deployed in a region that is geographically far away from the monitoring center of the network, where V is the set of sensor nodes and E is the set of links, $n = |V|$. Sensors have identical data generation rates r_g and their locations are stationary and known a *priori*. Each sensor is equipped with two radio interfaces: a low-power radio and a high-bandwidth radio, and can work on either type of the radios or both of them. The low-power radio is used for sensed data and communicating with other sensors within the sensor network. There is a link between two sensors if they are within the low-power radio's transmission range of each other. The *link reliability* of such a link $e \in E$, denoted by $p(e)$, however is determined by the path loss, concurrent transmission interferences, and ambient noises on wireless channels. We assume that the successful probabilities of any two data transmissions at different times over the same link e are independent, either of which only depends on $p(e)$. The high-bandwidth radios are employed to communicate with a third party network, and the data transmission over such radios is assumed reliable (beyond the control of the sensor network owner). The high-bandwidth radio at a sensor is only turned on when the sensor is a gateway and its buffered data needs to be sent immediately. We define the sensors that engage both of the two radios as *gateways*, Denote by GW the set of gateways and let $m = |GW|$ be the number of gateways which will be determined case by case. Sensors in $V \setminus GW$ only employ low-power radios to communicate with other sensors in the network.

To enable the sensing data from each source node to reach the remote monitoring center, the sensed data must go through the following three stages: it is first transmitted from its source node to a gateway (omitted if it is generated by the gateway itself) along a path in the routing tree rooted at the gateway, then relayed out of the sensor network by the gateway, and finally forwarded to the monitoring center by the third party network. Let T_i be the tree rooted at gateway $g_i \in GW$ and $V(T_i)$ be the set of nodes in $T_i, 1 \leqslant i \leqslant m$. For a given gateway g_i, the volume of data received at g_i via T_i within a period of t is denoted by $D^{(t)}(i)$. Let $e_1, e_2, \ldots, e_h$ be the link sequence in the path of T_i from a sensor node $v \in V(T_i)$ to gateway g_i. Denote by $p(v, g_i)$ the *end-to-end reliability* between v and g_i, then $p(v, g_i) = \prod_{i=1}^{h} p(e_i)$. We treat each attempt of v sending its data to g_i as *one trial* and each trial is an *i.i.d* event. Denote by $D(v, g_i)$ a 0–1 variable to represent whether one trial succeeds, i.e.,

$$D(v, g_i) = \begin{cases} 1 & \text{if the try is successful,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Then, $Pr[D(v, g_i) = 1] = p(v, g_i)$. The expectation of $D(v, g_i)$ is

$$E[D(v, g_i)] = Pr[D(v, g_i) = 1] \cdot 1 + Pr[D(v, g_i) = 0] \cdot 0 = p(v, g_i).$$

Following the definition of Poisson trials [13], the expected volume of data collected by gateway g_i in T_i within a period of t is

$$\begin{aligned} E[D^{(t)}(i)] &= E\left[\sum_{v \in V(T_i)} (r_g \cdot t \cdot D(v, g_i))\right] = r_g \cdot t \cdot \sum_{v \in V(T_i)} E[D(v, g_i)] \\ &= r_g \cdot t \cdot \sum_{v \in V(T_i)} p(v, g_i). \end{aligned} \tag{2}$$

3.1. *Service cost*

The service cost of remote monitoring is determined by the number of gateways and the extra volume of data beyond the data quota at each gateway, as telecommunication companies usually provide the data relay service for each gateway through optional data plans, each of which is with a *fixed cost* c_f for a data quota Q for a fixed period of τ (e.g. the mobile plan), and a *penalty* rate c_p for exceeding every MB data transfer during a charging period τ will be applied. Usually, the penalty rate is much more expensive than the data rate in the quota, i.e., $c_p > \frac{c_f}{Q}$. Denote by C_{ex} the service cost, then

$$C_{ex} = m \cdot c_f + \sum_{i=1}^{m} max\{0, (E[D^{(\tau)}(i)] - Q) \cdot c_p\}, \tag{3}$$

where $E[D^{(\tau)}(i)]$ is the expected volume of data received by gateway g_i within a charging period τ. The first term of the right hand side of Eq. (3) is the sum of the fixed cost of the m gateways, the second term is the total penalties incurred. The penalty to a gateway g_i is either 0 if the expected volume of transmitted data by the gateway does not exceed the quota Q, or $(E[D^{(\tau)}(i)] - Q) \cdot c_p$. Since the penalty depends on the *expected* volume of data exceeding the quota, the amount of the service cost itself is an expected value, too. For the sake of convenience, we still refer to it as service cost.

3.2. *Network throughput*

Recall that the volume of data received by the monitoring center within a charging period is the network throughput. To ensure the required data integrity and the monitoring quality by a specific application, we define the network throughput requirement as $D_{req}^{(\tau)} = \alpha \cdot n \cdot r_g \cdot \tau$, which is α percentage of the total volume of generated data during the charging period τ. And α is a pre-defined constant referred to as the *network throughput threshold* with $0 < \alpha \leqslant 1$. As we assume that data transmission within the sensor network is unreliable, this will result in data loss during its transmission. The network throughput, $\sum_{i=1}^{m} D^{(\tau)}(i)$, will not be deterministic, and its expectation is

$$\begin{aligned} E\left[\sum_{i=1}^{m} D^{(\tau)}(i)\right] &= \sum_{i=1}^{m} E(D^{(\tau)}(i)) = \sum_{i=1}^{m} r_g \cdot \tau \cdot \sum_{v \in V(T_i)} p(v, g_i) \\ &= r_g \cdot \tau \cdot \sum_{v \in V} p(v, g_i). \end{aligned} \tag{4}$$

To meet the specified throughput requirement, $E[\sum_{i=1}^{m} D^{(\tau)}(i)] \geqslant D_{req}^{(\tau)}$ should hold.

3.3. *Problem definition*

Given a dual-radio link-unreliable sensor network $G = (V, E)$ deployed for monitoring a region of interest, there is a monitoring center geographically located far away from the region of the sensor network G itself. Sensed data is transferred over high-bandwidth radios employed by some nodes to a third party network which will further forward the data to the monitoring center. Such data transfer by hiring service from a third party telecommunication company incurs cost and the amount of the cost is determined by the volume of data transferred and which data plan chosen. Data transmission over low-power radios within the sensor network causes data loss, yet it is required a certain percentage of sensing data generated by all sensors in a given charging period must be received by the monitoring center.

The *throughput guaranteed service cost minimization problem* in G thus is defined as follows. Given a network throughput threshold α and a specified data plan with a charging period of τ, the problem is to identify a set of nodes acting as gateways and find a forest of routing trees rooted at the gateways to transmit the sensing data generated within this period to the monitoring center such that the incurred service cost is minimized, subject to the throughput requirement.

To provide an efficient solution to the defined constrained optimization problem is challenging. The core difficulty lies in jointly determining the number of gateways and finding a routing tree rooted at each gateway. The volume of data received at each root is desirable to roughly equal the quota of the chosen data plan in order to avoid the penalty due to exceeding the quota, and the waste of money if the expected data volume is always below the quota. In each routing tree, nodes are to be spanned according to their remaining energy to balance the energy consumption in the network.

3.4. NP-completeness

Theorem 1. *The decision version of the throughput guaranteed service cost minimization problem is NP-complete.*

Proof. We show the claim by a reduction from the subset sum problem, which is NP-complete [20]. Given a set of positive integers $S = \{a_1, a_2, \ldots, a_n\}$, the subset sum problem is to find a partition of S which will result in two disjoint subsets S_1 and S_2 such that $\sum_{a_i \in S_1} a_i = \sum_{a_j \in S_2} a_j$. Given an integer $K = (\sum_{a_i \in S} a_i)/2$, the decision version of the instance of the subset sum problem is to determine whether there is a set partition S' and $S'' = S - S'$ such that $\sum_{a_i \in S'} a_i = \sum_{a_j \in S''} a_j = K$.

Having this instance of the subset sum problem, we construct an instance of the throughput guaranteed service cost minimization problem in a sensor network $G = (V \cup \{g_1, g_2\}, E)$ as follows. V is the set of sensors, there is a corresponding sensor $v_i \in V$ for each element $a_i \in S$. There are two gateway nodes g_1 and g_2, corresponding to sets S' and S'' respectively. There is an edge in E between each sensor node and either of the gateways, or two sensors if they are within the transmission range of each other. Assume that the reliability of a link between sensor v_i and either of these two gateways is $p_i = a_i/T$, where $T = max\{a_i \mid 1 \leqslant i \leqslant n\} + 1$ is the duration of the given monitoring period. We further assume that the reliability of each link between any two sensors is 1. Let $2K$ be the throughput requirement for the period of T. Assume that the data generation rate of sensors is $r_g = 1$, and the data plan is with a fixed cost c for the data quota of K within the charging period of T. The decision version of this special case of the throughput guaranteed service cost minimization problem is to ask: whether there are two routing trees rooted at gateways g_1 and g_2 such that the service cost is $2c$, subject to the throughput requirement being met. If there is a solution with the expected throughput $2K$, the expected volume of data received from either tree is K, and the total service cost is $2c$. That is, $\sum_{v_i \in V_1}(p_i \cdot T \cdot r_g) = \sum_{v_j \in V_2}(p_j \cdot T \cdot r_g) = K$, where V_i is the set of sensors in the tree rooted at g_i with $i = 1, 2$. Otherwise, if the expected volume of data received from one of the trees is less than K, to meet the $2K$ throughput requirement, the expected volume from another tree must be strictly larger than K, which incurs a cost more than c and the total service cost will be larger than $2c$. Clearly, if there is a solution to the above instance of the special throughput guaranteed service cost minimization problem, there is a solution to the instance of the subset sum problem. Since the subset sum problem is NP-complete and such a reduction is polynomial, the decision version of the throughput guaranteed service cost minimization problem thus is NP-hard. Meanwhile, it is easy to verify whether a given solution incurs a cost $2c$ with the expected throughput $2K$ in polynomial time. The problem of concern thus is in NP class. Therefore, the throughput guaranteed service cost minimization problem is NP-complete. □

4. Heuristic

Due to the NP-completeness of the problem, in this section we propose a heuristic. We start by giving a brief overview of the proposed algorithm, followed by providing the algorithm details.

4.1. Overview

Given a specific data plan for gateways, the service cost in a charging period is determined by the number of gateways, and the penalty for exceeding the quota at each individual gateway. Intuitively, if there is a solution to the problem, the minimum service cost can be achieved when neither the data quota is under-utilized nor any penalty is applied. That is, the amount of data relayed by each gateway within the charging period is exactly equal to its data quota. However, in reality, the volumes of data relayed by different gateways may not be balanced, which will result in money waste at some gateways which relay data less than the quota, or penalties at some others which relay data more than the quota.

To minimize the service cost while maintaining the throughput requirement, the proposed heuristic needs to identify an appropriate number of gateways, and design routing trees rooted at the gateways spanning the rest of sensors such that the service cost is minimized while the expected throughput requirement is maintained. As different number of gateways result in different service costs, we aim to find the one with the minimum service cost. A small number of gateways means the total fixed cost is relatively low. However, the quotas might be severely overused at some gateways and expensive penalties will be applied. If the total penalty is greater than the fixed cost of a gateway, it is worthwhile to employ one more gateway to share the work load among the gateways, the chance of exceeding the quota at each gateway becomes small, and the service cost can be reduced. On the other hand, a large number of gateways means that each gateway undertakes less data relay and small penalties or no penalty will be applied. However, the data quotas assigned to some gateways will be severely underutilized, and a large fraction of the fixed costs of these gateways will be wasted. If the volume of data relayed by a gateway can be redistributed to some other gateways without causing any quota exceeding among the involved gateways, its fixed cost will be saved by removing this gateway. Therefore, an appropriate number of gateways is to be found to fully utilize the data quota at each gateway and do not incur penalties.

Given the number of gateways m, we need to address which sensors should be identified as the gateways. The choice of gateways is guided by the following rationale. Gateways should have relatively high residual energy, this is because gateways relay data over the high-bandwidth radios and thus consume much more energy than other sensor nodes. If a sensor with low residual energy serves as a gateway, it may run out of energy before the end of the charging period, and the network lifetime will be significantly shortened. Once the gateways are identified, we construct the routing tree rooted at the gateways spanning all the other nodes in the network to balance the energy consumption overall the network.

4.2. *Algorithm with a fixed number of gateways*

For the sake of simplicity, in the following we first assume that the number of gateways m is given, and we aim to identify the m gateways and find m routing trees rooted at the gateways. We will remove this assumption later.

To select m gateways from n nodes, we first sort the nodes by their residual energy in non-increasing order. Denote by $er(v)$ the residual energy of node v at this moment. Let $v'_1, v'_2, \ldots v'_n$ be the sorted node sequence, where $er(v'_i) \geqslant er(v'_j), 1 \leqslant i < j \leqslant n$. Select the first $m' = \lceil n \cdot \beta \rceil > m$ nodes, where $0 < \beta \leqslant 1$ is a pre-defined parameter referred to as the *search space percentage*. A greater value of β indicates a larger search space for the m gateways, which may result in a better solution at the expense of a longer search time. We then randomly select m nodes from the m' nodes to be gateways.

What follows is to find a forest that consists of routing trees rooted at the m gateways in GW and spanning all the other nodes in $V \setminus GW$. Due to the energy imbalance among sensor nodes, the remaining energy at each node should be considered when it is added to a routing tree to prolong the network lifetime. To this end, we first construct a weighted, directed graph $G_d = (V', E', \omega)$, where $V' = V \cup \{s\}$ and node s is a "virtual sink", $E' = \{\langle v, u \rangle, \langle u, v \rangle | (v, u) \in E\} \cup \{\langle s, g \rangle \mid g \in GW\}$. That is, the virtual sink s is only connected to the m gateways and each of such links is assigned a weight of $\omega(\langle s, g \rangle) = 0$ for any $g \in GW$. For the weights of other edges in E', we incorporate the link reliability and the residual energy of sensor nodes into consideration. For example, for a directed edge $\langle v, u \rangle, \omega(\langle v, u \rangle) = IE \cdot \lambda^{1-er(v)/IE}/p(v, u)$ [10], where IE is the initial energy of each sensor, $p(u, v)$ is the link probability of its corresponding edge $(v, u) \in E$, and $\lambda > 1$ is a positive constant determining the impact of residual energy on the weight, referred to as the *weight adjustment parameter*. Note that all outgoing edges from a node v have identical weights, and for each edge $(v, u) \in E$, there are two directed edges $\langle v, u \rangle$ and $\langle u, v \rangle$ in E' with asymmetric weights. Initially, $er(v) = IE$ at each node $v \in V$. As nodes consume more and more energy as the network operates, the residual energy of each node becomes smaller and the weights of outgoing edges of the node will increase. The less the residual energy a node v has, the higher the weight of each of its outgoing edges. Having the auxiliary graph G_d, we now describe the construction of routing trees in G_d.

A single-source shortest path tree F in G_d rooted at s is constructed, using Dijkstra's algorithm [20]. Let $L(v, u)$ be the shortest path from node v to node u in graph G_d. In path $L(s, v)$ from the virtual sink s to any node $v \in V$, the first and second vertices in $L(s, v)$ are s and a gateway $g \in GW$, because the virtual sink can only connect to gateways. The removal of s and its incident edges from the shortest path tree F, a forest $\mathcal{F} = \{T_i \mid 1 \leqslant i \leqslant m\}$ consisting of routing trees rooted at the m gateways then follows. Note that the higher the link reliability $p(v, u)$ or the higher the residual energy at the node v, the lower the weight $\omega(\langle v, u \rangle)$, and the more likely the link is arranged in the higher level of a routing tree or vice versa.

For any node $v \in V(T_i)$, it sends its sensing data to gateway g_i along the reverse path $L(g_i, v)$ in T_i. The expected network throughput is calculated by Eq. (4) and the service cost is calculated by Eq. (3). The detailed algorithm is described in Algorithm 1.

Algorithm 1. Iden_GW

Input : $G(V, E)$, m, τ, β, r_g, c_f, c_p, Q
Output: The expected network throughput, the routing forest $\mathcal{F}$, and the service cost C_{ex}

/* Stage 1: identify m gateways */
Let $v'_1, v'_2, \ldots v'_n$ be the sorted node sequence in V in non-increasing order of residual energy;
$m' \leftarrow \lceil n \cdot \beta \rceil$ $(> m)$;
Select m nodes from the first m' nodes randomly as the gateways in GW;
/* Stage 2: build routing trees rooted at the m gateways */
Construct a weighted directed graph $G_d(V', E', \omega)$, with $V' = V \cup \{s\}$ and $E' = \{\langle v, u \rangle, \langle u, v \rangle | (v, u) \in E\} \cup \{\langle s, g \rangle \mid g \in GW\}$;
Find a single-source shortest path tree F in G_d rooted at s;
m routing trees rooted at the gateways in GW, T_i with $1 \leq i \leq m$ are obtained, by removing node s and all edges incident to it from F;
Calculate $\sum_{i=1}^{m} E[D^{(\tau)}(i)]$ according to Eq.(4);
Calculate C_{ex} according to Eq.(3);
return $\sum_{i=1}^{m} E[D^{(\tau)}(i)]$, $\mathcal{F} = \{T_i \mid 1 \leq i \leq m\}$, and C_{ex}.

4.3. *Algorithm without the given number of gateways*

In this subsection we propose an algorithm for the problem by removing the fixed number of gateways assumption as follows. We focus on finding an appropriate value of m to minimize the service cost. Let m_0 be such a value of m leading to the minimum service cost. In this ideal scenario, the expected volume of data collected by the m_0 gateways meets the throughput requirement, and the volume of generated data is evenly distributed (and relayed) to m_0 gateways, where $m_0 = \lfloor \frac{\alpha \cdot n \cdot \tau \cdot r_g}{Q} \rfloor$, the expected volume of data collected by the m_0 gateways will meet the throughput requirement. This will lead to the minimum service cost $C_{opt} = m_0 \cdot c_f$, because the data quota at every gateway is fully utilized and no data exceeding occurs. However, such a solution may never exist because its existence is fully determined by the network topology and link probabilities. In reality, the volume of data relayed by the m_0 gateways may not be balanced, either a larger or smaller number of gateways than m_0 may result in a lower service cost. In the following we develop a greedy heuristic to find a solution such that the service cost is the minimum among several candidate solutions.

The proposed heuristic is to identify the value of m first, then applies the algorithm in the previous section to find the solution to the problem. The rest is to find the value m, which proceeds iteratively in the two intervals $[1, m_0]$ and $[m_0 + 1, n]$ separately. We first find the appropriate value of m in the interval $[1, m_0]$ by setting $m = m_0$ and decreasing its value by one in each iteration. Within each iteration, it first calls algorithm `Iden_GW` with the current value of m as the input, and obtains a solution with a corresponding service cost. It then checks whether the service cost is the minimum one among all found candidate solutions so far. If not, the procedure continues until such a solution is found or the value of m is decreased to 1. Similarly, starting from $m = m_0 + 1$, we increase the value of m by one and compare the service cost with the minimum service cost found so far. In the end, a candidate solution with the minimum service cost will be chosen as the solution to the problem. The detailed description of algorithm `Min_Cost` is in Algorithm 2.

Algorithm 2. Min_Cost

```
Input  : G(V,E), τ, β, r_g, c_f, c_p, Q, D_req^(τ)
Output: A solution consisting of routing trees and
        the service cost
m_0 ← ⌊(α·n·τ·r_g)/Q⌋;
C_ex ← ∞ /* the initial service cost */ ;
/* search the interval [1, m_0] */;
m ← m_0; loop ←' true';
while (m ≥ 1 and loop) do
    call routine Iden_GW (G, m, τ, β, r_g, c_f, c_p, Q);
    E(D(m)) ← E(D^(τ)); F(m) ← F; C(m) ← C_ex;
    if E(D(m)) ≥ D_req^(τ) then
        if C(m) < C_ex then
            C_ex ← C(m); F ← F(m);
    else
        loop ←' false';
    m ← m − 1;
/* search the interval [m_0 + 1, n] */;
m ← m_0 + 1; loop ←' true';
while (m ≤ n and loop) do
    call routine Iden_GW (G, m, τ, β, r_g, c_f, c_p, Q);
    E(D(m)) ← E(D^(τ)); F(m) ← F; C(m) ← C_ex;
    if E(D(m)) ≥ D_req^(τ) then
        if C(m) < C_ex then
            C_ex ← C(m); F ← F(m);
    else
        loop ←' false';
    m ← m + 1;
return F and C_ex.
```

Notice that algorithm `Min_Cost` will be applied after every period of τ. We now analyze the computational complexity of algorithm `Min_Cost` by the following theorem.

Theorem 2. *Given a dual-radio sensor network $G(V, E)$ with unreliable links and a data plan, there is an algorithm for the throughput guaranteed service cost minimization problem in G, which takes $O(n^3)$ time, where $n = |V|$ is the number of sensors in the network.*

Proof. For a given m, identifying gateways takes $O(n \log n)$ time by sorting, while finding routing trees rooted at the gateways takes $O(|E| + |V| \log |V|) = O(n^2)$ time [20]. Finding an appropriate number of gateways takes at most n iterations, the computational complexity of algorithm `Min_Cost` thus is $O(n^3)$. □

5. Theoretical analysis

Since wireless communications in the sensor network are unreliable, the network throughput and service cost delivered by algorithm `Min_Cost` are the expected results. It is very likely that the actual volume of data received by the gateways may be below the network throughput requirement, or the actual service cost may be beyond the expected service cost. In the following we analyze the quality of the solution delivered by algorithm `Min_Cost` in this regard.

We first study the probability that the actual volume of data received by all gateways in $\mathcal{F}, \sum_{i=1}^{m} D^{(\tau)}(i)$ is less than the throughput requirement $D_{req}^{(\tau)} = \alpha \cdot r_g \cdot \tau \cdot n$. We refer to this probability as the *throughput failure probability*, denoted by $\Pr[\sum_{i=1}^{m} D^{(\tau)}(i) < D_{req}^{(\tau)}]$.

We assume that the expected volume of data received by all gateways in $\mathcal{F}$ is greater than the network throughput requirement., i.e., $E[\sum_{i=1}^{m} D^{(\tau)}(i)] > D_{req}^{(\tau)}$, which indicates that the network throughput requirement is met in most cases. We have the following lemma.

Lemma 1. *Given the network throughput requirement $D_{req}^{(\tau)}$ in any period of τ, the probability that the volume of data collected by all gateways within $\tau, \sum_{i=1}^{m} D^{(\tau)}(i)$, is less than $D_{req}^{(\tau)}$ is no more than $e^{-\mu_{\mathcal{F}}\left(1 - D_{req}^{(\tau)}/\mu_{\mathcal{F}}\right)^2/2}$, where $\mu_{\mathcal{F}} = E[\sum_{i=1}^{m} D^{(\tau)}(i)]$.*

Proof. Let $\delta = 1 - \frac{D_{req}^{(\tau)}}{\mu_{\mathcal{F}}}$. We have

$$\Pr\left[\sum_{i=1}^{m} D^{(\tau)}(i) < D_{req}^{(\tau)}\right] = \Pr\left[\sum_{i=1}^{m} D^{(\tau)}(i) < (1-\delta)\mu_{\mathcal{F}}\right] < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu_{\mathcal{F}}} \text{ by the Chenorff bound [13]} < e^{\frac{-\mu_{\mathcal{F}}\delta^2}{2}}. \tag{5}$$

By substituting δ in Eq. (5) with $1 - \frac{D_{req}^{\tau}}{\mu_{\mathcal{F}}}$, we have

$$\Pr\left[\sum_{i=1}^{m} D^{(\tau)}(i) < D_{req}^{(\tau)}\right] < e^{-\mu_{\mathcal{F}}\left(1 - D_{req}^{(\tau)}/\mu_{\mathcal{F}}\right)^2/2}.$$

□

Note that the throughput failure probability only depends on the value of $D_{req}^{(\tau)}$, because once the routing trees are identified, the end-to-end reliability between any node to its gateway is determined, too. The larger the value of $D_{req}^{(\tau)}$, the greater the throughput failure probability.

We then analyze the probability that the actual service cost is greater than the expected service cost of the solution delivered by algorithm `Min_Cost`, and we refer to this probability as *the cost exceeding probability*, denoted by $Pr[C^{(\tau)} \geqslant (1+\theta)C_{ex}]$ where $\theta > 0$ and $C^{(\tau)}$ is the actual service cost, $C^{(\tau)} = m \cdot c_f + \sum_{i=1}^{m} max\{0, (D^{(\tau)}(i) - Q) \cdot c_p\}$. The actual service cost $C^{(\tau)}$ depends on the value of $D^{(\tau)}(i)$, the actual volume of data received by each gateway $g_i \in GW$ within a period of τ, while the value of $D^{(\tau)}(i)$ depends only on the topological structure of $T_i \in \mathcal{F}$. Any two values of $D^{(\tau)}(i)$ and $D^{(\tau)}(j)$ are independent of each other when $i \neq j$. Therefore, the actual service cost of each gateway can be considered as an *i.i.d* random variable. Following the Chenorff bound [13], the probability that the actual service cost $C^{(\tau)}$ is greater than the expected service cost C_{ex} is no more than $\left(\frac{e^{\theta}}{(1+\theta)^{(1+\theta)}}\right)^{C_{ex}}$. Two special cases of this general setting can be derived: (i) when $0 < \theta \leqslant 1$, $Pr[C^{(\tau)} \geqslant (1+\theta) \cdot C_{ex}] \leqslant e^{-C_{ex} \cdot \theta^2/3}$; (ii) when $\theta \geqslant 5, Pr[C^{(\tau)} \geqslant (1+\theta) \cdot C_{ex}] \leqslant 2^{-(1+\theta) \cdot C_{ex}}$. We therefore conclude that a large value of θ will lead to a small cost exceeding probability.

In summary, both the throughput failure probability and cost exceeding probability in the solution delivered by the proposed algorithm `Min_Cost` are bounded.

6. Performance evaluation

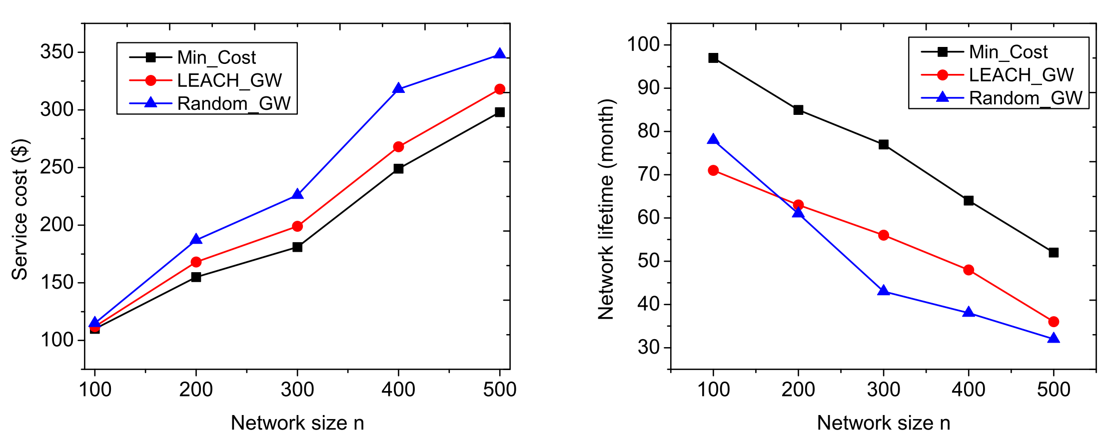
In this section we evaluate the performance of the proposed algorithm and investigate the impact of several constraint parameters on the network performance in terms of service cost and network lifetime (in month).

We consider a sensor network consisting of 100 to 500 sensors randomly deployed in a $1,000m \times 1,000m$ square region. We adopt the `CC2420` radio [1] and a typical 3G radio based on WCDMA 2100@24dBm standard [2] at each sensor. The initial energy capacity of each sensor IE is $1,000 Joules$. We adopt three different data plans provided by Vodafone [3]: (I) $Q = 2GB, c_f = \$19; (II)\ Q = 4$ GB, $c_f = \$29; (III)\ Q = 10GB, c_f = \$39.$ *Theseplansthesamepenaltyrate* $c_p = \$0.02/MB$. In the default simulation setting, Plan (II) will be adopted, the data generation rate $r_g = 50 Bytes/s$, the search space percentage $\beta = 0.1$, the weight adjustment parameter $\lambda = 2$, the link reliability is a random value within the interval $[0.5, 0.9]$, and the network throughput threshold $\alpha = 0.7$. Each value in figures is the mean of the results by applying the mentioned algorithm to 50 different network topologies of the same size.

6.1. Performance comparison of different algorithms

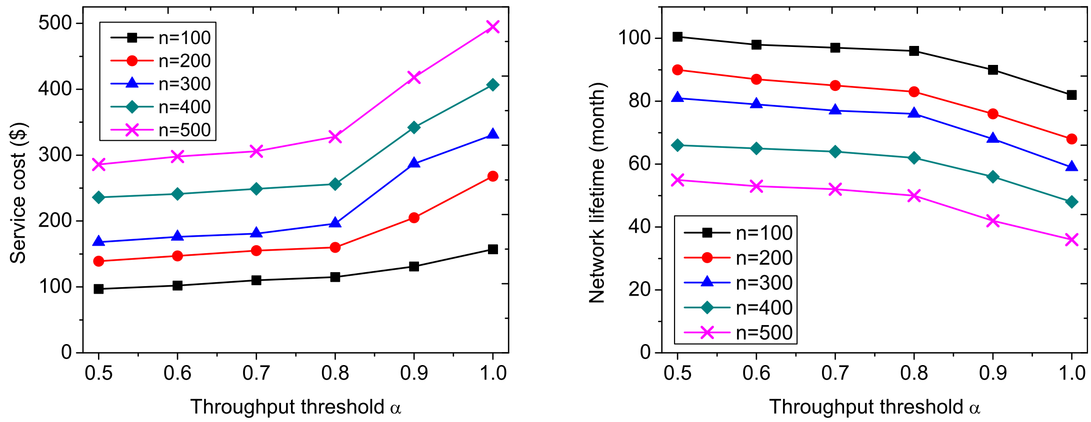
We first study the performance of algorithm `Min_Cost` against that of the other two algorithms. The only difference between these algorithms is in the identification of gateways. The number of gateways m is delivered by algorithm `Min_Cost`. One algorithm randomly selects m sensor nodes from all sensors as the gateways. We refer to this algorithm as `Random_GW`. The other is a variant of algorithm `LEACH` [8] which selects P percentage of nodes as the gateways, where $P = m/n$ is the ratio of the number of gateways to the total number of sensors. Nodes serving as gateways in the current charging period cannot be selected as gateways for the next $1/P$ periods. This algorithm is referred to as `LEACH_GW`. The rest of these three algorithms is identical, that is, the routing forest is built by adopting the forest establishment Algorithm 1. We compare the performance of these three algorithms in terms of the service cost and network lifetime by varying n from 100 to 500 while fixing $\alpha = 0.7$ and $r_g = 50\ Bytes/s, \lambda = 2$, and Plan (II) is adopted.

Fig. 2 shows that algorithm `Min_Cost` outperforms the other two in both the service cost and the network lifetime. On average, the service cost delivered by algorithm `Min_Cost` is 20% and 8% less than, and the network lifetime is 32% and 27% longer than those of algorithms `Random_GW` and `LEACH_GW`, respectively. From Fig. 2(a), it is observed that with the increase in n, the service cost of the solution delivered by each algorithm goes up, because larger volume of data is required to be sent to the remote monitoring center and a higher cost is incurred. With the growth of n, the gap between the three service cost curves is further enlarged. Fig. 2(b) illustrates that the curves of network lifetime drop with the growth of n. The superiority of algorithm `Min_Cost` lies in a more efficient gateway identification strategy to better balance the energy consumption among the sensor nodes. We also note that in terms of network lifetime, algorithm `LEACH_GW` outperforms algorithm `Random_GW` in most cases. The reason behind



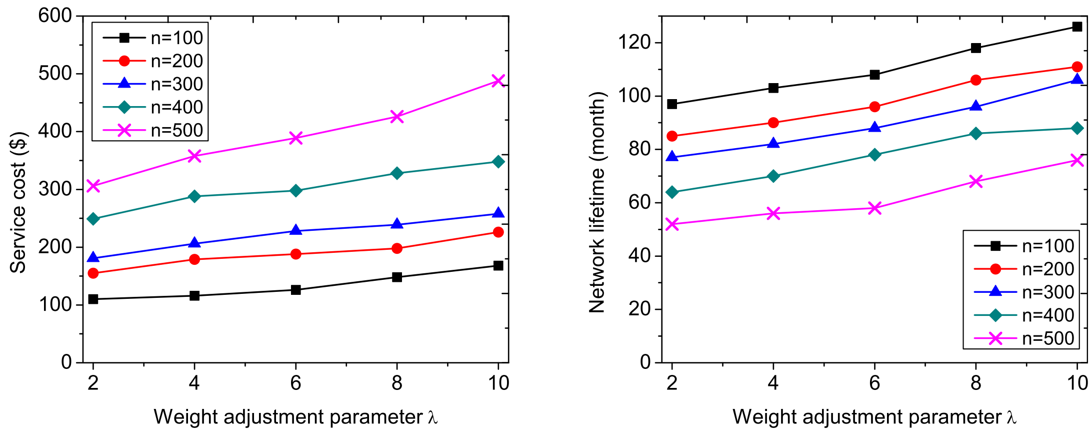
(a) Service cost (b) Network lifetime

Fig. 2 – Performance comparison of three algorithms when $\alpha = 0.7$, $r_g = 50$ Bytes/s, $\lambda = 2$, and Plan (II) is adopted.



(a) Service cost (b) Network lifetime

Fig. 3 – The performance of algorithm `Min_Cost` with different throughput thresholds α when r_g =50 Bytes/s, $\lambda = 2$, and Plan (II) is adopted.



(a) Service cost (b) Network lifetime

Fig. 4 – The performance of algorithm `Min_Cost` with different weight adjustment parameter λ when $r_g = 50$ Bytes/s, $\alpha = 0.7$, and Plan (II) is adopted.

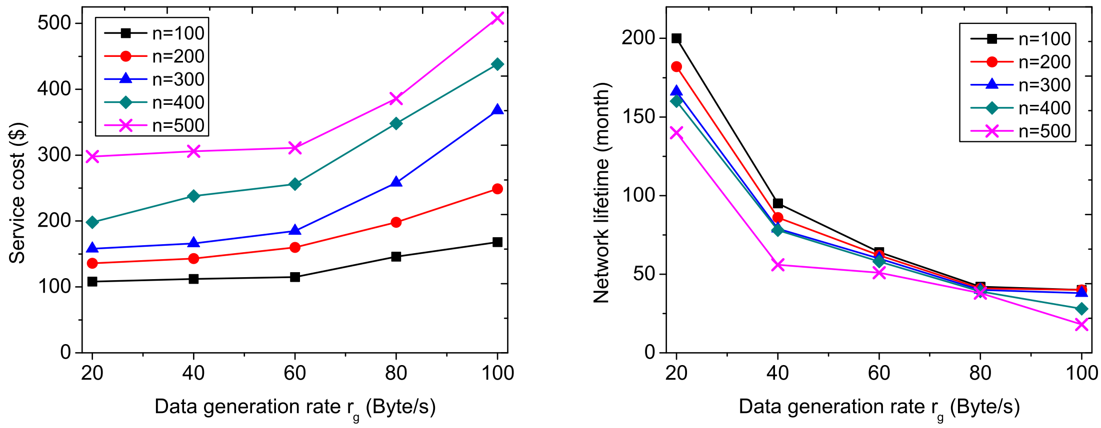
is that the nodes in algorithm `LEACH_GW` cannot be repeatedly selected as the gateways in a number of consecutive rounds, while algorithm `Random_GW` does not pose such a restriction, which potentially increases the chances for more balanced energy distribution and a longer network lifetime.

6.2. Impact of constraint parameters on network performance

We now investigate the impact of different constraint parameters on the performance of algorithm `Min_Cost` in terms of the service cost and network lifetime.

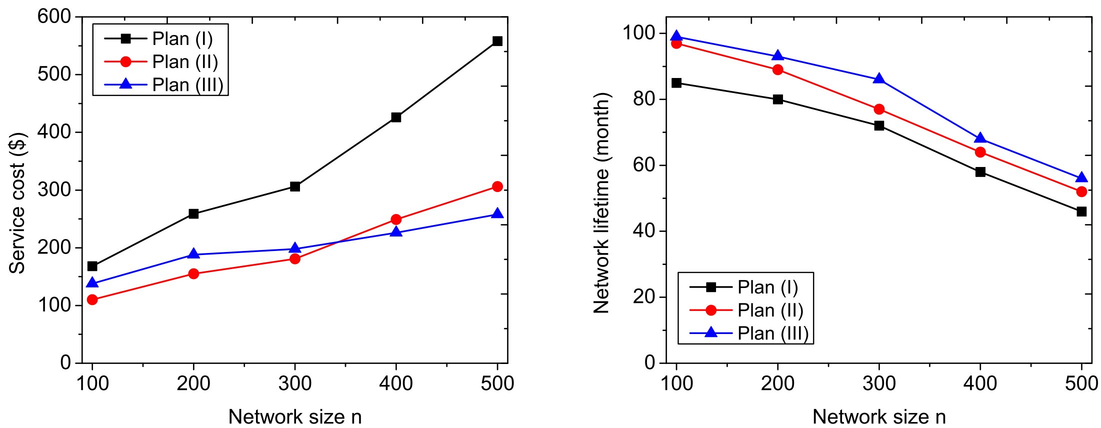
We start with the impact of the network throughput threshold α on the network performance by varying α from 0.5 to 1.0. As shown in Fig. 3, the service cost increases while the network lifetime decreases as the value of α goes up. This is because the higher the throughput requirement, the larger the volume of sensing data collected from the sensor network, thereby resulting in a higher service cost and more energy consumption among sensors, thus a shorter network lifetime. However, note that when prior to $\alpha = 0.8$, the service cost and network lifetime vary slowly and are flat, while α reaches 0.8, both of them change dramatically. The rationale behind is that a larger α does not necessarily mean that the amount of data relayed by each gateway increases accordingly. With the growth of α from 0.5 to 0.8, the forest of routing trees may not experience many changes, resulting in slight changes in the service cost and network lifetime. However, with further increase in α, a large number of gateways is expected to be used in order to meet the network throughput requirement, resulting in a greater service cost. The similar explanation applies to the trend of network lifetime, too.

We then investigate the impact of weight adjustment parameter λ on the network performance by varying λ from 2 to 10. Fig. 4 indicates that the increase of λ results in a higher service cost yet a longer network lifetime. Recall that the weight of a directed edge $\langle v, u\rangle$ is $\omega(v,u) = IE \cdot \lambda^{1-er(v)/IE}/p(v,u)$. The value of λ affects weights of edges thus the routing forest construction. The larger the value of λ, the greater the impact of the residual energy on the edge weight, and the more balanced energy consumption. However, a larger λ leads to a higher service cost, as shown in Fig. 4(a). In each iteration of



(a) Service cost (b) Network lifetime

Fig. 5 – The performance of algorithm *Min_Cost* with different data generation rate r_g when $\alpha = 0.7$, $\lambda = 2$, and Plan (II) is adopted.



(a) Service cost (b) Network lifetime

Fig. 6 – The performance of algorithm *Min_Cost* with different data plans when $\alpha = 0.7$, $\lambda = 2$, and $r_g = 50$ Bytes/s.

searching the optimal number of gateways m, the routing forest algorithm with a larger λ delivers a forest with a lower throughput, compared to that delivered by the same algorithm with a smaller λ, which very likely does not meet the specified throughput requirement. As a result, a larger number of gateways is required and a higher service cost is incurred.

We also address the impact of the data generation rate r_g on the network performance by varying r_g from 20 Bytes/s to 100 Bytes/s with the increment of 20 Bytes/s. Fig. 5 shows that with a fixed r_g, the larger the network size n, the higher the service cost, and the shorter the network lifetime, because more data is required to be transmitted and a larger number of gateways is required. This can also explain that with the increase in the data generation rate r_g, the service cost will go up while the network lifetime will drop.

What follows is to investigate the impact of different data plans on the network performance. Fig. 6(a) indicates that adopting Plan (I) incurs the highest service cost among the three plans. Adopting Plan (II) is the cheapest when $n \leqslant 300$, and when $n > 300$ Plan (III) leads to the smallest service cost. Recall that the number of gateways m is searched around $m_0 = \lfloor \frac{\alpha \cdot n \cdot \tau \cdot r_g}{Q} \rfloor$. A smaller data quota Q indicates a larger number of gateways needed and a higher service cost incurred, it is because transmitting the same amount of data by adopting a data plan with a smaller data quota usually incurs a higher cost than that of adopting a plan with a larger data quota, as illustrated by the following example. Assume that there is 20 GB collected data to be sent, and the amount of data sent through each gateway is equal to the data quota. Corresponding to plans (I), (II), and (III), the numbers of gateways needed are 10 (2 GB at each gateway), 5 (4 GB at each gateway), and 2 (10 GB at each gateway), and the corresponding service costs are \$190, \$145, \$79 respectively. Though the penalty is not considered, the fixed cost is dominant in the service cost. This explains the higher cost caused by Plan (I) in comparison with the other two plans. It is also interesting to see that when $n > 300$, adopting Plan (III) results in a lower cost compared with Plan (II). It is because when Plan (II) is adopted, a higher penalty is incurred, which depends on the quota usage on individual gateways. In other words, adopting a plan with a larger quota (e.g., Plan (III)) means a smaller fixed cost yet might be accompanied with an expensive penalty, and results in a high service cost in the end.

We finally evaluate the impact of charging period τ on the network lifetime by the proposed algorithm. Assume that

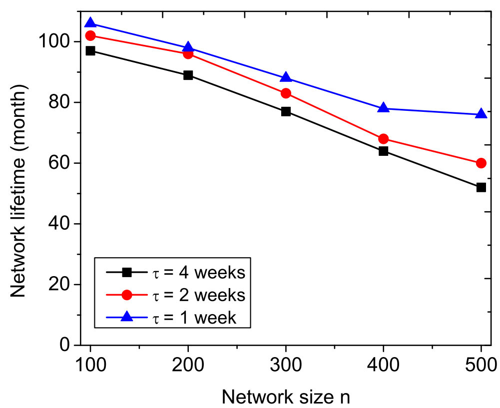


Fig. 7 – The performance of algorithm *Min_Cost* **with different charging periods** τ **when** $r_g = 50$ **Bytes/s,** $\alpha = 0.7$**, and** $\lambda = 2$**.**

two varieties of Plan (II) are adopted: one is with Q = 2 GB, $c_f = \$14.5$, and $\tau = 2$ weeks; another is with $Q = 1GB, c_f = \$7.25$, and $\tau =$ 1 week. Both plans have the same penalty rate $c_p = \$0.02/MB$. We evaluate the monthly service cost by adopting Plan (II) ($\tau = 4$ weeks) and its two varieties. Fig. 7 shows that the shorter the charging period, the longer the network lifetime delivered by the proposed algorithm, as this results in more frequent changes of gateways, thus more balanced energy consumption among the sensors.

7. Conclusions

In this paper we studied a novel remote monitoring cost minimization problem for a link-unreliable dual-radio sensor network, subject to the user-specified network throughput requirement. We formulated the problem as a constrained optimization problem and showed its NP-completeness. We then proposed a heuristic for the problem by jointly identifying gateways and finding energy-efficient routing trees rooted at the gateways spanning the other nodes. We also conducted theoretical analysis on the quality of the solution obtained. We finally evaluated the performance of the proposed algorithm by simulations. Experimental results demonstrate that the proposed algorithm outperforms other heuristics in terms of both the service cost and the network lifetime significantly.

R E F E R E N C E S

[1] CC2420 2.4 GHz IEEE 802.15.4/ZigBee-ready RF transceiver. <www.ti.com/lit/ds/symlink/cc2420.pdf.2012>.

[2] PCI express minicard and LGA modules high-speed multimode 3G. <www.embeddedworks.net/ewdatasheets/option/EW-Gobi3000.pdf>; 2012.

[3] Vodafone. <http://www.vodafone.com.au>; 2012.

[4] Al-Karaki JN, Kamal AE. Routing techniques in wireless sensor networks: a survey. IEEE Wireless Commun 2004;11(6):6–28.

[5] Braginsky D, Estrin D. Rumor routing algorithm for sensor networks. In: Proceedings of the first ACM international workshop on wireless sensor networks and applications, Atlanta, USA; 2002. p. 22–31.

[6] Chang J-H, Tassiulas L. Energy conserving routing in wireless ad-hoc networks. In: Proceedings of the 19th annual joint conference of the IEEE computer and communications societies, Tel Aviv, Israel; 2000. p. 22–31.

[7] Gummeson J, Ganesan D, Corner MD, Shenoy P. An adaptive link layer for heterogeneous multi-radio mobile sensor networks. IEEE J Selected Areas Commun 2010;28(7):1094–104.

[8] Heinzelman WR, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd annual Hawaii international conference on system sciences. Hawaii, USA; 2000. p. 1–10.

[9] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In: Proceedings of the 6th annual international conference on mobile computing and networking. Boston, USA; 2000. p. 56–67.

[10] Liang W, Liu L. On-line data gathering for maximizing network lifetime in sensor networks. IEEE Trans. Mobile Comput 2007;6(1):2–11.

[11] Liang W, Luo J, Xu X. Prolonging network lifetime via a controlled mobile sink in wireless sensor networks. In: Proceedings of the IEEE 2010 global telecommunications conference. Miami, USA; 2010. p. 1–6.

[12] Lymberopoulos D, Priyantha NB, Goraczko M, Zhao F. Towards efficient design of multi-radio platforms for wireless sensor networks. In: Proceedings of the 2008 international conference on information processing in sensor networks, St. Louis, USA; 2008. p. 257–268.

[13] Michael M, Upfal E. In: Probability and computing: randomized algorithms and probabilistic analysis. UK: Cambridge University Press; 2005. p. 61–70.

[14] Ren F, Zhang J, Lin C, He T, Ren S. Ebrp: energy-balanced routing protocol for data gathering in wireless sensor networks. IEEE Trans. Parallel Distributed Syst 2011;22(12):2108–25.

[15] Sengul C, Bakht M, Harris AF, Abdelzaher T, Kravets R. Improving energy conservation using bulk transmission over high-power radios in sensor networks. In: Proceedings of the 28th international conference on distributed computing systems. Beijing, China; 2008. p. 801–808.

[16] Servetto SD, Barrenechea G. Constrained random walks on random graphs: routing algorithms for large scale wireless sensor networks. In: Proceedings of the first ACM international workshop on wireless sensor networks and applications. Atlanta, USA; 2002. p. 12–21.

[17] Shepard TJA. Channel access scheme for large dense packet radio networks. In: Proceedings of the ACM 1996 conference on Applications, technologies, architectures, and protocols for computer communications. California, USA; 1996. p. 219–230.

[18] Stathopoulos T, Lukac M, Mclntire D, Heidemann J, Estrin D, Kaiser WJ. End-to-end routing for dual-radio sensor networks. In: Proceedings of the 26th IEEE international conference on computer communications. Alaska, USA; 2007. p. 2252–2260.

[19] Hussain S, Islam O. An energy efficient spanning tree based multi-hop routing in wireless sensor networks. In: Proceedings of the IEEE wireless communications and networking conference. Hong Kong, China; 2007. p. 4383–4388.

[20] Rivest RL, Cormen TH, Leiserson CE, Stein C. In: Introduction to algorithms. USA: MIT Press; 2009. p. 1128–40.

[21] Xu X, Liang W. Monitoring quality optimization in wireless sensor networks with a mobile sink. In: Proceedings of the 14th ACM international conference on modeling, analysis and simulation of wireless and mobile systems. Miami, USA; 2011. p. 77–84.

[22] Xu X, Liang W, Wark T. Data quality maximization in sensor networks with a mobile sink. In: Proceedings of the 2011

international conference on distributed computing in sensor systems and workshops. Barcelona, Spain; 2011. p. 1–8.
[23] Xu X, Liang W, Wark T, Jeong J. Maximizing network lifetime via 3G gateway assignment in dual-radio sensor networks. In: Proceedings of the 37th conference on local computer networks. Florida, USA; 2012. p. 479–486.
[24] Xu X, Liang W, Xu Z. Minimizing remote monitoring cost of wireless sensor networks. In: Proceedings of the 2013 IEEE wireless communications and networking conference. Shanghai, China; 2013. p. 1476–1481.
[25] Xu Z, Liang W, Xu Y. Network lifetime maximization in delay-tolerant sensor networks with a mobile sink. In: Proceedings of the 8th international conference on distributed computing in sensor systems. Hangzhou, China; 2012. p. 9–16.
[26] Xu Y, Heidemann J, Estrin D. Geography-informed energy conservation for Ad Hoc routing. In: Proceedings of the 7th annual international conference on Mobile computing and networking. Rome, Italy; 2001. p. 70–84.