Minimizing Remote Monitoring Cost of Wireless Sensor Networks

Xu Xu, Weifa Liang, Zichuan Xu

The Australian National University, Canberra, Australia

grace.xu@anu.edu.au, wliang@cs.anu.edu.au, edward.xu@anu.edu.au

Abstract—In this paper we consider a remote monitoring scenario where the monitoring center is geographically located far away from the region of the deployed sensor network, and the sensing data by the sensors is transmitted to the monitoring center through a third party telecommunication service, thus a cost associated with this service will be incurred, which is related to the amount of data successfully received by the monitoring center within a specified period. For this scenario, we first formulate a novel optimization problem, namely, the throughput guaranteed service cost minimization problem with an objective to minimize the service cost while the specified network throughput requirement is guaranteed. We show that the problem 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. Finally, we conduct experiments by simulation to evaluate the performance of the proposed heuristic. The experimental results demonstrate the proposed algorithm outperforms other two mentioned algorithms in terms of both service cost and the network lifetime.

I. INTRODUCTION

Traditionally, the base station is deployed at a pre-defined strategic location in a wireless sensor network (WSN) 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 is deployed in a different region from its monitoring center and the sensing data by the sensors is to be sent to the center in real-time. Clearly, the traditional multi-hop data gathering paradigm in which both the base station and the sensors are deployed in the same region is not applicable to this scenario due to the long distance (at least several hundreds of kilometers) between the monitoring center and the deployed network. To enable the monitoring center to receive sensing data from the remote sensor network in real-time, long-distance data transmission service has to be leased from a third party telecommunication company. If not all, some sensors in the sensor network must register themselves to this service so that they are able to access to the third party network to transmit data to the monitoring center. Such sensors are referred to as the *gateways*. In our discussion, we assume that each sensor is equipped with dual-radios: one with high-bandwidth radio such as 3G or 4G, and the other with low-power radio such as IEEE 802.15.4. The high-bandwidth radios are used by the gateways to communicate with the facilities in the third party network. Such high-bandwidth links are considered to

be reliable yet incur costs, and the cost incurred by high-bandwidth links is defined as the *service cost*. This service cost is usually comprised of a *fixed cost* for a data *quota* as well as a *penalty cost* for any exceeding data usage beyond the quota on a fixed period basis, referred to as the *charging period*. The other sensors in the sensor network will forward their sensing data to the gateways for further relay. The low-power radios are adopted for local data transmissions in the sensor network which are unreliable due to time-varying fading of wireless channels [6] and thus cause data loss. The actual volume of data received by the monitoring center within a charging period is defined as the *network throughput*. In this paper we refer to a given percentage of all generated data within a charging period as the *network throughput requirement*. If high throughput is required, a large volume of data has to be sent via the third party network and a more expensive service cost is incurred, and vice versa. Our objective in this paper thus is to minimize the service cost while maintaining the network lifetime as long as possible, subject to a specified network throughput requirement. To solve this constrained optimization problem, sensors need to be allocated to gateways, and a set of routing trees rooted at gateways are built to span all sensors, such that the sum of the cost for relaying data by each individual gateway is minimized while the expected volume of data relayed by all gateways meets the throughput requirement.

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 and finding a forest of routing trees rooted at the gateways. We finally conduct experiments by simulation to evaluate the performance of the proposed algorithm and study the impact of different constraint parameters on its performance. The experimental results demonstrate that the proposed algorithm outperforms other two 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 is considered, and a novel solution to the problem is provided.

II. RELATED WORK

Data gathering is one primary function of wireless sensor networks, which has been extensively explored in the past decade [8], [15]. Previous studies on this problem can be clas-

sified into three categories based on different routing protocols adopted [4]: flat-based routing, hierarchical-based routing, and location-based routing. In the flat-based routing, all sensors are assigned with equal capabilities and play the same roles. These kinds of routing protocols include directed diffusion, rumor routing, random-walks-based routing, and so on. In the hierarchical-based routing, sensors are clustering into different clusters, the network has intra-cluster and inter-cluster layers, and sensors will serve as different roles (cluster heads and cluster members) [8], [14], [15]. 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 [16]. This paper falls into the hierarchical-based routing category. The idea of periodic gateway identification in this paper is similar to the cluster head selection and rotation scheme in LEACH [8]. However, we assume that wireless links are not reliable in the sensor network and focus on building the routing trees rooted at gateways to guarantee that the expected volume of data collected from trees meets the network throughput requirement. Furthermore, unlike most existing works that focused on the network lifetime, we focus on minimizing the service cost caused by transmitting data to the remote monitoring center.

The dual-radio wireless sensor network has been studied recently. Most studies focused on energy conservation [7], [10]. The main challenge of these works is to find a fine tradeoff between minimizing the amount of time spent by the high-bandwidth radio in idle status and using the lower-power radio as a paging and control channel for resource discovery and mobility support [11]. For example, Stathopoulos *et al.* [12] 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.* [10] concluded that high bandwidth radios are energy efficient only when the amount of data to be sent is large. Different from the usage of dual-radio in these studies, the dual-radio model adopted by this paper is not for the purpose of saving energy, but for the remote data transmission. Instead of using both radios for data transmission within the network, only the low-power radio is used. We investigate the cost incurred by data transmission over the high-bandwidth radios, which is a significant departure from existing studies.

III. 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 sensing 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 $p(e)$, with $0 \le p(e) \le 1$, however is determined by the path loss of wireless channels. We assume that the successful probability of data transmission over the same link e only depends on $p(e)$. Whereas the data transmission over the high-bandwidth radios is considered reliable. A sensor that engages both of the two radios is a *gateway*. The high-bandwidth radio at a gateway is only turned on when its buffered data needs to be sent immediately. 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 for communication in the network.

To reach the remote monitoring center, the sensing data 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 $V(T_i)$ the set of nodes in T_i, $1 \le i \le m$. 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 $p(v, g_i) = \prod_{i=1}^{h} p(e_i)$. The expected volume of data collected by gateway g_i within a period of t is $E[D^{(t)}(i)] = r_g \cdot t \cdot \sum_{v \in V(T_i)} p(v, g_i)$.

A. Service Cost

The service provided by the third party telecommunication company is charged according to the volume of data transmitted via individual gateways. Different data plans are provided, each of which has a *fixed cost* c_f for a data *quota* Q within a fixed *charging period* τ, and a *penalty* rate c_p for exceeding every MB data usage during this period. And the exceeding data charging rate usually is much higher than the rate for quota data, i.e., $c_p > \frac{c_f}{Q}$. For each gateway, the fixed cost is for any amount of data usage no more than the quota, while a penalty will be applied once the quota is exceeded and the amount of penalty is proportional to the exceeding data volume. Thus, if a gateway cannot consume its data quota within the charging period, it may choose another cheaper plan with a smaller data quota in the future. Otherwise, it under-utilizes the data quota by the current plan and wastes money in some sense. On the contrary, if a gateway always exceeds its quota a lot, it is wise to choose a more expensive plan with a larger data quota to reduce the prohibitive penalty. For simplicity, in this paper we assume that all gateways adopt the same data plan and the set of gateways is fixed during a charging period. We investigate the sum of the cost for sending the expected volume of data collected by each individual gateway within one charging period, and refer to it as the *service cost*. It is denoted by C_{ex}, and

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

where $E[D^{(\tau)}(i)]$ is the expected volume of data received by gateway g_i within a charging period τ. The first term in the right hand side of Eq. (1) is the sum of the fixed cost of the m gateways, and 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 $(E[D^{(\tau)}(i)] - Q) \cdot c_p$.

B. Network Throughput

The volume of data received by the monitoring center within a charging period is defined as 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 data generated during the charging period τ. And α is a pre-defined constant referred to as the *network throughput threshold* with $0 < \alpha \leq 1$. As we assume that data transmission within the sensor network is not reliable and results in data loss, the network throughput which is the sum volume of data received by all gateways during the period of τ, $\sum_{i=1}^{m} D^{(\tau)}(i)$, is not deterministic too. Recall that $p(v, g_i)$ represents the end-to-end reliability between v and g_i. The expectation of the network throughput is

$$
\begin{aligned}
E[\sum_{i=1}^{m} D^{(\tau)}(i)] &= \sum_{i=1}^{m} E(D^{(\tau)}(i)) \\
&= \sum_{i=1}^{m} r_g \cdot \tau \cdot \sum_{v \in V(T_i)} p(v, g_i) = r_g \cdot \tau \cdot \sum_{v \in V} p(v, g_i).
\end{aligned} \tag{2}
$$

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

C. Problem Definition

Given a dual-radio 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. The cost of transferring sensing data is determined by how much data is to be transmitted through a third party network and which plan is chosen. Since wireless communication is unreliable, to ensure the monitoring quality of the monitored region, α percentage of the 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 gateways for the period 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.

D. 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 [13]. 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 $T = max\{ a_i \mid 1 \leq i \leq n\} + 1$ 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. ∎

IV. HEURISTIC

A. Overview

Due to the NP-completeness of the problem, we here propose a heuristic for it. 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 sum of cost for relaying data collected from each individual tree is minimized while the expected volume of data collected meets the specified throughput requirement. As different number of gateways result in different service costs, we aim to find the one with the minimum service cost. On one hand, a small number of gateways means the total fixed cost is relatively low yet the quotas might be severely overused at some gateways and expensive penalties will be applied. 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 may be severely under-utilized, and a large fraction of the fixed costs of these gateways will be wasted. Therefore, an appropriate number of gateways is to be found to fully utilize the data quota at each gateway and minimize penalties if unavoidable. To this end, we first find the routing forest with a given number of gateways, and then find the number of gateways.

B. Finding Routing Trees with a Given Number of Gateways

Assuming that the number of gateways m has been given, this section deals with identifying m gateways from all sensor nodes and finding the m routing trees rooted at the identified gateways.

To select m gateways from n nodes in V, 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 $er(v'_i) \geq er(v'_j)$, $1 \leq i < j \leq n$. Select the first $m' = \lceil n \cdot \beta \rceil > m$ nodes, where $0 < \beta \leq 1$ 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 $V \setminus GW$. To this end, we 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 $\omega(s, g) = 0$ 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 $\langle v, u \rangle$, its weight is $\omega(v, u) = IE \cdot \lambda^{(1-er(v)/IE)}/p(v, u)$ [9], where IE is the initial energy of each sensor, $p(u, v)$ is the link reliability 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 over time, the residual energy of each node becomes smaller and the weights of outgoing edges from the node will increase. The less the energy left at node v, the higher the weights its outgoing edges. Having the auxiliary graph G_d, we now describe the construction of a forest consisting 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 [13]. 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 are respectively s and a gateway $g \in GW$. That is because the virtual sink can only access to gateways thus the path from s to any other node must be via a gateway. After the removal of s and its incident edges from tree F, a forest $\mathcal{F} = \{T_i \mid 1 \leq i \leq m\}$ consisting of routing trees rooted at the m gateways then follows. 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. (2) and the service cost is calculated by Eq. (1). The detailed description is given in the following routine `Iden_GW`.

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, and the service cost

/*stage 1: identify m gateways */
1. Let $v'_1, v'_2, \ldots v'_n$ be the node sequence in V sorted in non-increasing order of residual energy;
2. $m' \leftarrow \lceil n \cdot \beta \rceil > m$;
3. Randomly select m nodes from the first m' nodes and form the set of gateways GW;
/*stage 2: build routing trees rooted at the m gateways*/
4. Construct a weighted directed graph $G_d(V', E', \omega)$, with $V' = V \cup \{s\}$ and
$E' = \{\langle v, u \rangle, \langle u, v \rangle | (v, u) \in E\} \cup \{\langle s, g \rangle \mid g \in GW\}$;
5. Find a shortest path tree F in G_d rooted at s;
6. 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;
7. Calculate $\sum_{i=1}^{m} E[D^{(\tau)}(i)]$ according to Eq.(2);
8. Calculate C_{ex} according to Eq.(1);
return $\sum_{i=1}^{m} E[D^{(\tau)}(i)]$, $\mathcal{F} = \{T_i \mid 1 \leq i \leq m\}$, *and* C_{ex}.

C. Finding the Number of Gateways

So far we have assumed the number of gateways m is given, we now remove this assumption and propose an algorithm for the problem as follows. We focus on finding an appropriate value of m to minimize the service cost. Consider a scenario that the total volume of generated data is evenly distributed (and relayed) to m_0 gateways, where $m_0 = \lfloor \frac{\alpha \cdot n \cdot \tau \cdot r_g}{Q} \rfloor$, and the expected volume of data collected by the m_0 gateways meets the throughput requirement. This will lead to a lower bound of 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 not exist as it 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 its corresponding service cost is the minimum among the found solutions so far while the throughput requirement is met.

The proposed heuristic proceeds to search in the two intervals $[1, m_0]$ and $[m_0 + 1, n]$ separately to find m. We first search 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 the routine `Iden_GW` with the current value of m as the input, and obtains a solution with a corresponding service cost and network throughput. It then checks whether the throughput meets the requirement and the cost is the minimum one among the found solutions so far. The procedure terminates if either condition fails or the value of m is decreased to 1, and a solution is found. We then start from $m = m_0 + 1$, increase the value of m by one in each iteration, call the routine `Iden_GW` and repeat the above process. In the end, a feasible solution to the problem with a service cost will be delivered. The algorithm is referred to as `Min_Cost` and its computational complexity is analyzed as follows.

Theorem 2: Given a dual-radio sensor network $G(V, E)$ with unreliable link reliability and a data plan, there is an algorithm for the throughput guaranteed service cost minimization problem, 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 $O(|V|) = O(n)$ time, while finding routing trees rooted at the gateways takes $O(|E| + |V| \log |V|) = O(n^2)$ time [13]. There are at most n iterations to search the appropriate value of m from 1 to n. Thus the computational complexity of algorithm `Min_Cost` thus is $O(n^3)$. ∎

V. PERFORMANCE EVALUATION

In this section we evaluate the performance of the proposed algorithm and investigate the impact of constraint parameters on the network performance. We consider a sensor network consisting of 100 to 500 sensors randomly deployed in a $1,000m \times 1,000m$ square region. The initial energy capacity of each sensor IE is $1,000$ $Joules$. The energy consumption parameters of IEEE 802.15.4 and 3G radios are referred to [1] and [2]. We adopt four different data plans provided by Vodafone [3]: (I) $Q = 500MB$, $c_f = \$50$; (II) $Q = 2GB$, $c_f = \$65$; (III) $Q = 3GB$, $c_f = \$85$; (IV) $Q = 4GB$, $c_f = \$100$. Each plan has the same penalty rate $c_p = \$0.25/MB$. In the default simulation setting, we adopt Plan (II), the charging period $\tau = 30days \times 24hours \times 3,600seconds$, the data generation rate $r_g = 50$ $Bytes/s$, the network throughput threshold $\alpha = 0.7$, the search space percentage $\beta = 0.1$, the weight adjustment parameter $\lambda = 2$, and the link reliability is a random value within the interval $[0.5, 0.9]$. Each value in

figures is the mean of the results by applying the mentioned algorithm to 50 different network topologies of the same size.

A. Performance Comparison

We first compare the performance of algorithm `Min_Cost` against that of other two algorithms, where the only difference between these algorithms is in the identification of m gateways. 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`. Having identified the m gateways, these two algorithms build trees rooted at the gateways by adopting algorithm `Iden_GW`. We compare the performance of these algorithms in terms of service cost and network lifetime by varying n from 100 to 500.
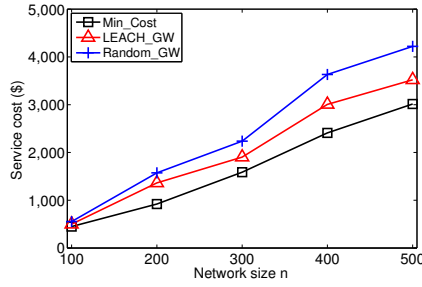
Service cost: From Fig. 1(a), it is observed that with the increase of n, the service cost of the solution delivered by each algorithm goes up, because a larger volume of data is required to be relayed to the remote monitoring center. Notice that algorithm `Min_Cost` always outperforms the other two. And with the further growth of n, the gap between them is further enlarged.

Network lifetime: The network lifetime is defined as the time of the first sensor's failure due to the depletion of energy [5]. Fig. 1(b) illustrates that the curves of network lifetime drop with the growth of n. The network lifetime delivered by Algorithm `Min_Cost` is about 1.4 and 1.8 times longer than those of algorithms `LEACH_GW` and `Random_GW`, respectively. The superiority of algorithm `Min_Cost` lies in a more efficient gateway identification strategy. It selects m gateways with the objective to better balance the energy consumption among nodes that results in a longer network lifetime, while the other two algorithms just randomly choose the m gateways from all nodes. We also note that most of the time, algorithm `LEACH_GW` outperforms algorithm `Random_GW` but sometimes `Random_GW` does deliver a better result. The reason behind is that nodes in algorithm `LEACH_GW` cannot be repeatedly selected as the gateways in several consecutive rounds while the nodes in algorithm `Random_GW` do not have such a restriction.
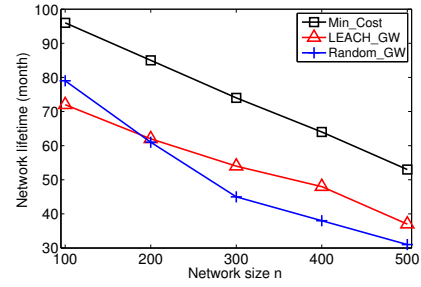
B. Impact of Constraint Parameters on Service Cost

We then study the impact of different constraint parameters on the service cost of algorithm `Min_Cost`.

Network throughput threshold α: We now study the impact of the network throughput threshold on the service cost by varying α from 0.5 to 1.0 with the increment of 0.1. As shown in Fig. 2, the curves are flat before $\alpha = 0.8$ and increase rapidly after that point. The rationale behind is that a large α does not necessarily mean that the amount of data relayed by each individual gateway increases too. With the growth of the

Fig. 1. Performance comparison of three different algorithms in terms of service cost and network lifetime
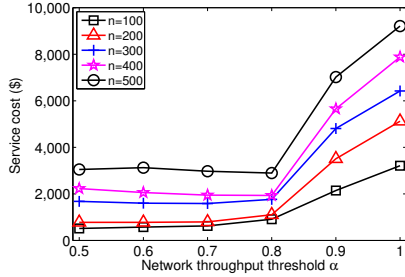

Fig. 2. The service cost of algorithm `Min_Cost` with different throughput thresholds α

value of α from 0.5 to 0.8, the forest of routing trees for data gathering does not experience many changes, resulting in the similar costs. However, with the further growth of the value of α, a larger number of gateways is expected to be employed in order to meet the higher network throughput requirement, resulting in a greater service cost, too.
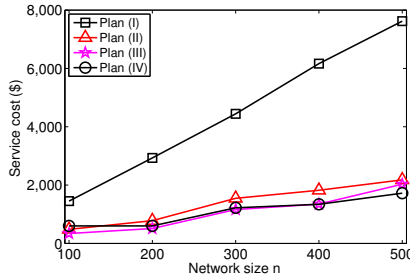

Fig. 3. The service cost of algorithm `Min_Cost` with different data plans

Data plan: What follows is to investigate the impact of different data plans on the service cost. Fig. 3 indicates that Plan (I) will incur the highest service cost among the mentioned data plans while the others result in the similar service costs. This is because the quota Q in Plan (I) is small thus more gateways are needed to relay data, and more expensive fixed cost is caused in every charging period. It is interesting to see that adopting a data plan with a large data quota does not necessarily mean that a small penalty will be incurred. The penalties depend on the quota usage on individual gateways, which could be either over-utilized or under-utilized. That is, a plan (e.g., Plan (IV)) resulting in a small fixed cost might be accompanied with an expensive penalty. That is why there is a marginal difference in service

costs with Plan (II), (III), and (IV) in Fig. 3.

VI. CONCLUSION

In this paper we studied a novel problem of minimizing the service cost of an 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 by jointly identifying gateways and finding routing trees rooted at the gateways spanning the other sensors. We finally evaluated the performance of the proposed algorithm by simulation. The experimental results demonstrate that the proposed algorithm outperforms other two heuristics in terms of both the service cost and the network lifetime.

REFERENCES

[1] CC2420 2.4 GHz IEEE 802.15.4/ZigBee-ready RF transceiver. www.ti.com/lit/ds/symlink/cc2420.pdf.
[2] PCI express minicard and LGA modules high-speed multi-mode 3G. www.embeddedworks.net/ewdatasheets/option/EW-Gobi3000.pdf.
[3] Vodafone. http://www.vodafone.com.au, Sept 2012.
[4] J. N. Al-Karaki and A. E. Kamal. Routing techniques in wireless sensor networks: a survey. *IEEE Wireless Communications*, 11:6–28, 2004.
[5] J-H Chang and L. Tassiulas. Energy conserving routing in wireless ad-hoc networks. In *Proc. of INFOCOM*. IEEE, 2000.
[6] G. D. Durgin, T. S. Rappaport, and D. A. de Wolf. New analytical models and probability density functions for fading in wireless communications. *IEEE Trans. on Communications*, 50(6):1005–1015, 2002.
[7] J. Gummeson, D. Ganesan, M. D. Corner, and P. Shenoy. An adaptive link layer for heterogeneous multi-radio mobile sensor networks. *IEEE Journal on Selected Areas in Communications*, 28:1094–1104, 2010.
[8] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proc. of HICSS*. IEEE, 2000.
[9] W. Liang and Y. Liu. On-line data gathering for maximinizing network lifetime in sensor networks. *IEEE TMC*, 6:2–11, 2007.
[10] D. Lymberopoulos, N. B. Priyantha, M. Goraczko, and F. Zhao. Towards efficient design of multi-radio platforms for wireless sensor networks. In *Proc. of IPSN*. IEEE, 2008.
[11] C. Sengul, M. Bakht, A. F. Harris, T. Abdelzaher, and R. Kravets. Improving energy conservation using bulk transmission over high-power radios in sensor networks. In *Proc. of ICDCS*. IEEE, 2008.
[12] T. Stathopoulos, M. Lukac, D. McIntire, J. Heidemann, D. Estrin, and W. J. Kaiser. End-to-end routing for dual-radio sensor networks. In *Proc. of INFOCOM*. IEEE, 2007.
[13] R. L. Rivest T. H. Cormen, C. E. Leiserson and C. Stein. *Introduction to Algorithms*. MIT Press, 3rd edition, 2009.
[14] X. Xu and W. Liang. Monitoring quality optimization in wireless sensor networks with a mobile sink. In *Proc. of MSWiM*. IEEE, 2011.
[15] X. Xu, W. Liang, and T. Wark. Data quality maximization in sensor networks with a mobile sink. In *Proc. of DCOSS*. IEEE, 2011.
[16] Y. Xu, J. Heidemann, and D. Estrin. Geography-informed energy conservation for Ad Hoc routing. In *Proc. of MobiCom*. ACM, 2001.