On the Construction of Data Aggregation Tree with Minimum Energy Cost in Wireless Sensor Networks: NP-Completeness and Approximation Algorithms

Tung-Wei Kuo and Ming-Jer Tsai
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan, ROC
E-mail: mjtsai@cs.nthu.edu.tw

Abstract—**In many applications, it is a basic operation for the sink to periodically collect reports from all sensors. Since the data gathering process usually proceeds for many rounds, it is important to collect these data efficiently, that is, to reduce the energy cost of data transmission. Under such applications, a tree is usually adopted as the routing structure to save the computation costs for maintaining the routing tables of sensors. In this paper, we work on the problem of constructing a data aggregation tree that minimizes the total energy cost of data transmission in a wireless sensor network. In addition, we also address such a problem in the wireless sensor network where relay nodes exist. We show these two problems are NP-complete, and propose $O(1)$-approximation algorithms for each of them. Simulations show that the proposed algorithms each have good performance in terms of the energy cost.**

I. Introduction

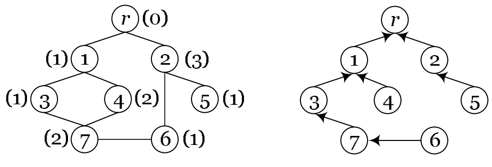
In many applications, sensors are required to send reports to a specific target (e.g. base station) periodically [1]. In habitat monitoring [2] and civil structure maintenance [3], it is a basic operation for the sink to periodically collect reports from sensors. Since the data gathering process usually proceeds for many rounds, it is necessary to reduce the number of the packets, which carries the reports, transmitted in each round for energy saving. In this paper, we undertake the development of data gathering in wireless sensor networks.

Data aggregation is a well-known method for data gathering, which can be performed in various ways. In [1], a fixed number of reports received or generated by a sensor are aggregated into one packet. In other applications, a sensor can aggregate the reports received or generated into one report using a divisible function (e.g. SUM, MAX, MIN, AVERAGE, top-k, etc.) [4]. Data compression, which deals with the correlation between data such that the number of reports is reduced, is another method for data gathering [5], [6]. In many applications, the spatial or temporal correlation does not exist between data (e.g. status reports [1]), and data aggregation is a more suitable method for data gathering.

The effectiveness of data aggregation is mainly determined by the routing structure. In many data aggregation algorithms, a tree is used as the routing structure [7], [8], [9], [10], [11], [12], especially for the applications that have to monitor events continuously. The reason is that sensors, which usually have limited resources, can save relatively high computational costs for maintaining routing tables if sensors route packets based on a tree. While several papers target at the maximization of the network lifetime [7], [8], it is sometimes desirable to minimize the energy cost. For example, in rechargeable sensor networks [13], [14], [15], as it is hard to predict the energy replenishment profile, minimizing the energy cost is a simple way to prolong the network lifetime. For some indoor applications, sensors may have AC power plugs. Under such circumstance, energy saving then becomes the major issue. In this paper, the problem of constructing a data aggregation tree with minimum energy cost will be studied. **Our contributions are described below:**

- We prove the problem of constructing a data aggregation tree with minimum energy cost, termed MECAT, is NP-complete and provide a 2-approximation algorithm.
- We study the variant of such a problem, in which the relay nodes exist, termed MECAT_RN. We show the MECAT_RN problem is NP-complete and demonstrate a 7-approximation algorithm.
- We show any λ-approximation algorithm of the Capacitated Network Design (CND) problem [16] can be used to obtain a 2λ-approximation algorithm of the MECAT_RN problem.
- We conduct several simulations to evaluate the performances of the proposed algorithms.

The remainder of this paper is organized as follows. Section II describes the network model and shows the MECAT problem is NP-complete. Section III provides a 2-approximation algorithm for the MECAT problem. In Section IV, we show the MECAT_RN problem is NP-complete and give a 7-approximation algorithm. We show a 2λ-approximation algorithm of the MECAT_RN problem can be obtained using a λ-approximation algorithm of the CND problem in Section



(a)

(b)

Fig. 1: The network model. (a) A wireless sensor network, where each node has a weight shown in parentheses. (b) A routing tree.

V. Using simulations, we evaluate the performances of the proposed algorithms in Section VI. Related works are studied in Section VII. Finally, we conclude the paper in Section VIII.

II. Network Model and Problem Definition

We first illustrate the network model in Section II-A. Subsequently, our problem is described and shown to be NP-complete in Section II-B.

A. The Network Model

We model a network as a connected graph $G = (V, E)$ with weights $s(v) \in \mathbb{Z}^+$ and 0 associated with each node $v \in V \setminus \{r\}$ and r, respectively, where V is the set of nodes, E is the set of edges, and $r \in V$ is the sink. Each node v has to send a report of size $s(v)$ to sink r periodically in a multi-hop fashion based on a routing tree. A routing tree constructed for a network $G = (V, E)$ with sink r is a directed tree $T = (V_T, E_T)$ with root r, where $V_T = V$ and a directed edge $(u, v) \in E_T$ only if an undirected edge $\{u, v\} \in E$. A node u can send a packet to a node v only if $(u, v) \in E_T$, in which case u is a child of v, and v is the parent of u. For the energy consumption, we only consider the energy cost of the radio [8]. Let Tx and Rx be the energy needed to send and receive a packet, respectively. While routing, a hop-by-hop aggregation is performed according to the aggregation ratio, q, which is the size of reports that can be aggregated into one packet. Because it would be meaningless if the aggregation ratio is set to a non-integer, the aggregation ratio is assumed to be an integer through this paper.

Example 1. Fig. 1(b) is a routing tree constructed for the wireless sensor network shown in Fig. 1(a). Assume the aggregation ratio is 3, and both Tx and Rx are equal to 1. Using the routing tree, node 6 first sends a packet containing its report to node 7. After node 7 receives the packet from node 6, node 7 aggregates the reports of nodes 6 and 7 into one packet and then sends the packet to node 3. The process proceeds until node r receives the reports of all nodes. Clearly, 3, 2, 2, 1, 1, 1, and 1 packets are sent by nodes 1, 2, 3, 4, 5, 6, and 7, respectively; therefore, a total of 11 packets are sent (and received) by the nodes. It is easy to verify that a total of 9 packets are required to be sent if the parent of node 6 is set to node 2.

B. The Problem and Its Hardness

We first describe our problem in the following.

Problem 1. Given a network $G = (V, E)$ with weights $s(v) \in \mathbb{Z}^+$ and 0 associated with each node $v \in V \setminus \{r\}$ and r, respectively, a sink $r \in V$, an aggregation ratio $q \in \mathbb{Z}^+$, energy costs $Tx \in \mathbb{R}^+$ and $Rx \in \mathbb{R}^+$ for transmitting and receiving a packet, respectively, and $C \in \mathbb{R}^+$, the **M**inimum **E**nergy-**C**ost **A**ggregation **T**ree (**MECAT**) problem asks for a routing tree $T = (V_T, E_T)$ with root r and $V_T = V$, such that the total transmission and reception energy consumed by all sensors is not greater than C. In addition, MECAT(G, r, q, Tx, Rx, C) denotes an instance of the MECAT problem, and $COST(T)$ denotes the energy cost of a routing tree T.

Next, we prove that the MECAT problem is NP-complete by showing a polynomial-time reduction from the Load-Balanced Semi-Matching problem, an NP-complete problem, as described below.

Definition 1. A **semi-matching** in a bipartite graph $G = (U \cup V, E)$ is an edge set $M \subseteq E$, such that every node in U incident to exactly one edge in M. Given a semi-matching M and $v \in V$, $Adj_M(v)$ denotes the set of nodes u with $\{v, u\} \in M$.

Problem 2. [17] Given a bipartite graph $G = (U \cup V, E)$ with a weight $w(u) \in \mathbb{Z}^+$ associated with each node $u \in U$ and $k \in \mathbb{Z}^+$, the **L**oad-**B**alanced **S**emi-**M**atching (**LBSM**) problem asks for a semi-matching M such that $k \geq \max_{v \in V} \sum_{u \in Adj_M(v)} w(u)$. Furthermore, LBSM$(G, k)$ denotes an instance of the LBSM problem.

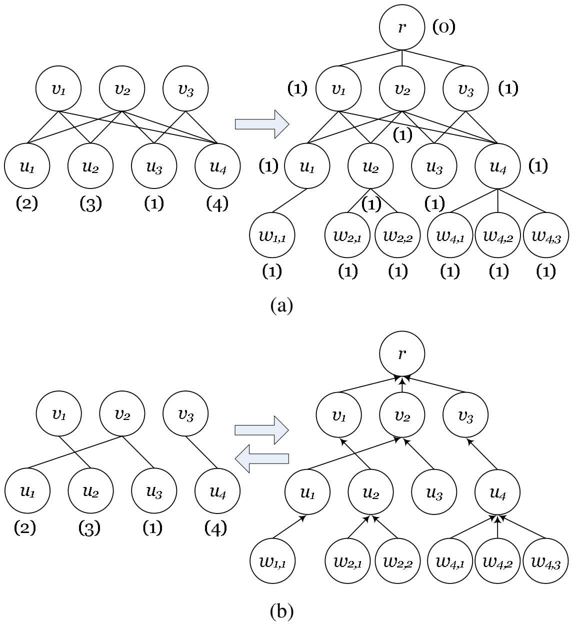
Theorem 1. *The MECAT problem is NP-complete.*

Proof: First, the MECAT problem is clearly in NP, since we can verify in polynomial time if a candidate solution is a tree and satisfies the energy cost constraint. Next, we prove that the MECAT problem is NP-hard by showing a polynomial-time reduction from the LBSM problem to the MECAT problem. For every instance LBSM$(G = (U \cup V, E), k)$, we construct an instance MECAT(G', r, q, Tx, Rx, C) as follows:

1) $G' = (\{r\} \cup U \cup V \cup W, E \cup E_R \cup E_W)$ with weights 1 and 0 associated with nodes in $U \cup V \cup W$ and $\{r\}$, respectively,
2) $q = k + 1$,
3) $Tx = Rx = 1$, and
4) $C = 2(|W| + \sum_{1 \leqslant i \leqslant |U|} \lceil \frac{|W_i| + 1}{q} \rceil + |V|)$,

where $W = \bigcup_{1 \leqslant i \leqslant |U|} W_i$, $W_i = \{w_{i,j} | 1 \leqslant j \leqslant w(u_i) - 1\}$, $E_R = \{\{r, v_i\} | 1 \leqslant i \leqslant |V|\}$, and $E_W = \{\{u_i, w_{i,j}\} | 1 \leqslant i \leqslant |U|, 1 \leqslant j \leqslant w(u_i) - 1\}$. Clearly, this instance is constructed in polynomial time. See Fig. 2, for example.

We need to show LBSM(G, k) has a feasible solution if, and only if, MECAT(G', r, q, Tx, Rx, C) has a feasible solution. For the "only if" part, let M be a semi-matching in G such that $k \geq \max_{v \in V} \sum_{u \in Adj_M(v)} w(u)$. We show that using M, a routing tree that spans $\{r\} \cup U \cup V \cup W$ and has a total energy cost not greater than C can be constructed. Let $T = (V_T, E_T)$ be a routing tree with $V_T = \{r\} \cup U \cup V \cup W$ and



(a)

(b)

Fig. 2: NP-hardness of the MECAT problem. (a) Reduction from the LBSM problem to the MECAT problem, where $w(u_i)$ is shown in parentheses and k is equal to 4. (b) Correspondence between the solutions for the LBSM and MECAT problems.

$E_T = \{(v_i, r)|1 \leqslant i \leqslant |V|\} \cup \{(u_i, v_j)|\{u_i, v_j\} \in M\} \cup \{(w_{i,j}, u_i)|1 \leqslant i \leqslant |U|, 1 \leqslant j \leqslant w(u_i) - 1\}$. We show the total energy cost of T is not greater than C. Since $Tx = Rx = 1$, the total transmission (or reception) energy of the nodes in W (or U) is $|W|$ and the total transmission (or reception) energy of the nodes in U (or V) is $\sum_{1 \leqslant i \leqslant |U|} \lceil \frac{|W_i|+1}{q} \rceil$. Since $C = 2(|W| + \sum_{1 \leqslant i \leqslant |U|} \lceil \frac{|W_i|+1}{q} \rceil + |V|)$ and $Tx = Rx = 1$, we only need to show each node in V sends exactly one packet. The size of reports sent by node v_j in V is

$$\begin{aligned} 1 + \sum_{u_i \in Adj_M(v_j)} (1 + |W_i|) &\leqslant 1 + \max_{v \in V} \sum_{u_i \in Adj_M(v)} (1 + |W_i|) \\ &= 1 + \max_{v \in V} \sum_{u_i \in Adj_M(v)} w(u_i) \\ &\leq k + 1. \end{aligned} \tag{1}$$

Thus, each node in V needs to send reports with a total size at most $k + 1$, which can be aggregated into 1 packet.

For the "if" part, let $T = (V_T, E_T)$ be a routing tree with minimum energy cost not greater than C. We show a semi-matching in G such that $k \geq \max_{v \in V} \sum_{u \in Adj_M(v)} w(u)$ can be constructed using T. Clearly, $(v_j, u_i) \notin E_T$ for all $v_j \in V$ and $u_i \in U$; otherwise, there exists a routing tree $T' = (V_T, E_T \setminus \{(v_j, u_i)\} \cup \{(v_j, r)\})$ with less energy cost than T. Let $M = \{\{u_i, v_j\}|(u_i, v_j) \in E_T\}$. We show

$$k \geq \max_{v \in V} \sum_{u \in Adj_M(v)} w(u). \tag{2}$$

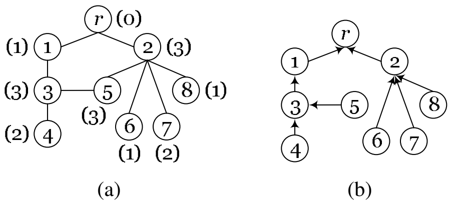
As in the proof of the "only if" part, the total transmission and reception energy of the nodes in U and W plus the reception energy of the nodes in V is

$$2(|W| + \sum_{1 \leqslant i \leqslant |U|} \lceil \frac{|W_i| + 1}{q} \rceil). \tag{3}$$

In addition, the total energy cost of T is not greater than

$$C = 2(|W| + \sum_{1 \leqslant i \leqslant |U|} \lceil \frac{|W_i| + 1}{q} \rceil + |V|). \tag{4}$$

(3) and (4) imply each node in V sends only one packet. Thus, each node in V receives at most k reports, implying (2). ■



(a)

(b)

Fig. 3: (a) A wireless sensor network, where $q = 9$ and $Tx = Rx = 1$. (b) A non-shortest path tree with minimum energy cost.

III. Approximation Algorithm

As the MECAT problem is NP-complete, we provide an approximation algorithm. Observe that while sending a packet to the sink, the longer the routing path is, the greater the energy cost is. Naturally, we would route each packet via a shortest path to the sink. The resulting routing structure is then a shortest path tree. There are at least two benefits to route packets using a shortest path tree. First, a shortest path tree is easy to construct in a distributed manner, as described in the following two steps. The sink node first broadcasts a message such that each node can evaluate the hop distance from the sink [18]. Then, each node sets its parent to the node with a smaller hop distance from the sink. Second, in many time-critical applications, it is necessary to route packets using a shortest path tree to achieve the minimum packet transmission delay. Although a shortest path tree may not have minimum energy cost (see Fig. 3, for example), Theorem 2 shows a shortest path tree algorithm has an approximation ratio of 2. Definition 2 and Lemma 1 are necessary for the proof of Theorem 2.

Definition 2. Given a graph $G = (V, E)$ and a root $r \in V$, a **Minimum Descendant Tree** is a tree T rooted at r and spanning V, such that $\sum_{v \neq r} des_T(v)$ is minimized, where $des_T(v)$ is the total size of reports to be sent by v's descendants in T.

Lemma 1. *Every shortest path tree is a minimum descendant tree.*

Proof: The lemma directly follows two claims below:

1) Every minimum descendant tree is a shortest path tree.
2) Every shortest path tree T has the same value of $\sum_{v \neq r} des_T(v)$.

We show claim 1 by contradiction. Suppose that there exists a minimum descendant tree $T' = (V_{T'}, E_{T'})$ that is not a shortest path tree. Let $D_G(v)$ and $D_{T'}(v)$ be the hop distances from v to r in G and T', respectively. Let $V' = \{v|D_G(v) < D_{T'}(v)\}$ and $v' = \arg\min_{v \in V'} D_G(v)$, i.e., v' is the node in V' with minimum hop distance to r in G. Then, $V' \neq \emptyset$ and v' must exist. Let u be the parent of v' in T', and u' be v' neighboring node with a smaller hop distance from r in G. Let $T” = (V_{T'}, E_{T'} \setminus \{(v', u)\} \cup \{(v', u')\})$. Clearly, $T”$ is a tree with $\sum_{v \neq r} des_{T”}(v) < \sum_{v \neq r} des_{T'}(v)$, a contradiction.

For claim 2, let T_1 and T_2 be any two shortest path trees. Clearly, T_1 and T_2 have the same height, say H. Let $L_k(T)$ be the set of nodes whose hop distances from the root in a tree T are k. We have

$$\sum_{v \in L_H(T_1)} des_{T_1}(v) = \sum_{v \in L_H(T_2)} des_{T_2}(v) = 0, \quad (5)$$

and

$$\sum_{v \in L_k(T)} des_T(v) = \sum_{v \in L_{k+1}(T)} des_T(v) + \sum_{v \in L_{k+1}(T)} s(v), \quad \forall 1 \leqslant k \leqslant H-1. \quad (6)$$

Since T_1 and T_2 are shortest path trees,

$$\sum_{v \in L_k(T_1)} s(v) = \sum_{v \in L_k(T_2)} s(v), \forall 1 \leqslant k \leqslant H. \quad (7)$$

By (5), (6), and (7), we have claim 2. ■

Theorem 2. *Every shortest path tree algorithm is a 2-approximation algorithm.*

Proof: Let T be a routing tree. Since the number of packets sent by nodes equals that received by nodes in T,

$$COST(T) = (Tx + Rx) \sum_{v \neq r} \lceil \frac{des_T(v) + s(v)}{q} \rceil. \quad (8)$$

Let T_{OPT} be a routing tree with minimum energy cost and T_{SPT} be an arbitrary shortest path tree. By Lemma 1, we obtain

$$\begin{aligned} COST(T_{OPT}) &= (Tx + Rx) \sum_{v \neq r} \lceil \frac{des_{T_{OPT}}(v) + s(v)}{q} \rceil \\ &\geqslant (Tx + Rx) \sum_{v \neq r} \frac{des_{T_{OPT}}(v) + s(v)}{q} \\ &\geqslant (Tx + Rx) \sum_{v \neq r} \frac{des_{T_{SPT}}(v) + s(v)}{q}. \quad (9) \end{aligned}$$

In addition, by (8), we have

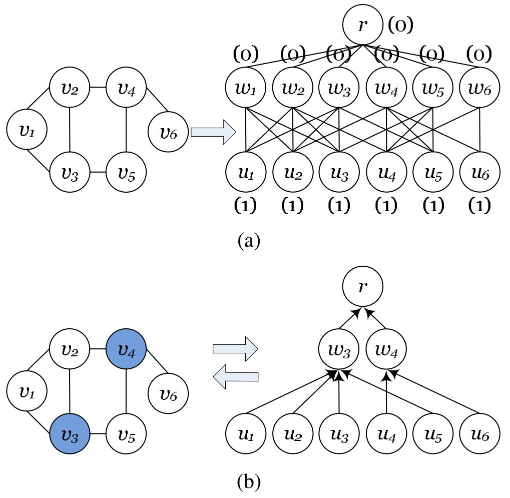
$$COST(T_{SPT}) = (Tx+Rx) \sum_{v \neq r} \lceil \frac{des_{T_{SPT}}(v) + s(v)}{q} \rceil \quad (10)$$

Therefore, by (9) and (10), we get

$$COST(T_{SPT}) - COST(T_{OPT}) < (Tx+Rx)(|V|-1). \quad (11)$$

In addition, each node has to send at least one packet, and these packets must be received by some nodes. Thus,

$$COST(T_{OPT}) \geqslant (Tx + Rx)(|V| - 1). \quad (12)$$



(a)

(b)

Fig. 4: NP-hardness of the MECAT_RN problem. (a) Reduction from the Dominating Set problem to the MECAT_RN problem. (b) Correspondence between the solutions for the Dominating Set and MECAT_RN problems.

Combining (11) and (12), we obtain

$$COST(T_{SPT}) < 2 \cdot COST(T_{OPT}). \quad (13)$$

■

IV. Data Aggregation with Relay Nodes

To improve the network connectivity or survivability, the relay node placement problem in a wireless sensor network has been extensively investigated in the literature [19], [20], [21]. These relay nodes, which do not produce reports, are used to forward the packets received from other nodes. In this section, we study the problem of constructing a data aggregation tree with minimum energy cost in the presence of relay nodes.

A. The Problem and Its Hardness

Here, a routing tree only needs to span all non-relay nodes. For the convenience of description, we assume every relay node has a zero-sized report. In the following, the problem is described and shown to be NP-complete.

Problem 3. Given a network $G = (V, E)$ with weights $s(u) \in \mathbb{Z}^+$ and 0 associated with each source $u \in U \subseteq V \setminus \{r\}$ and $v \in V \setminus U$, respectively, a set of sources U, a sink $r \in V$, an aggregation ratio $q \in \mathbb{Z}^+$, energy costs $Tx \in \mathbb{R}^+$ and $Rx \in \mathbb{R}^+$ for transmitting and receiving a packet, respectively, and $C \in \mathbb{R}^+$, the **M**inimum **E**nergy-**C**ost **A**ggregation **T**ree with **R**elay **N**odes (**MECAT_RN**) problem asks for a routing tree $T = (V_T, E_T)$ with root r and $V_T \supseteq U \cup \{r\}$, such that the total transmission and reception energy consumed by all sensors is not greater than C. Moreover, MECAT_RN(G, U, r, q, Tx, Rx, C) denotes an instance of the MECAT_RN problem, and $COST(T)$ denotes the energy cost of a routing tree T.

Theorem 3. *The MECAT_RN problem is NP-complete.*

Proof: First, it is easy to see that the problem is in NP since a non-deterministic algorithm just needs to guess a tree spanning all nodes in U and check in polynomial time if the energy cost of the tree is not greater than C. Next, to show the MECAT_RN problem is NP-hard, we demonstrate a polynomial-time reduction from the Dominating Set problem [22], which asks for a dominating set D in G with $|D| \leqslant k$ for a given instance DS(G, k), to the MECAT_RN problem. For any instance DS$(G = (V, E), k)$, we construct an instance MECAT_RN(G', U, r, q, Tx, Rx, C) as follows:

1) $G' = (\{r\} \cup W \cup U, E_S \cup E_R)$ with weights 0 and 1 associated with nodes in $\{r\} \cup W$ and U, respectively,
2) $q = \Delta(G) + 1$,
3) $Tx = Rx = 1$, and
4) $C = 2(|V| + k)$,

where $W = \{w_i | 1 \leqslant i \leqslant |V|\}$, $U = \{u_i | 1 \leqslant i \leqslant |V|\}$, $E_S = \{\{r, w_i\} | 1 \leqslant i \leqslant |V|\}$, and $E_R = \{\{w_i, u_i\} | 1 \leqslant i \leqslant |V|\} \cup \{\{w_i, u_j\} | \{v_i, v_j\} \in E\}$, and $\Delta(G)$ denotes the maximum degree of G. Clearly, this instance is constructed in polynomial time. See Fig. 4, for example.

We need to show DS(G, k) has a feasible solution if, and only if, MECAT_RN(G', U, r, q, Tx, Rx, C) has a feasible solution. For the ”only if” part, let D be a dominating set in G with $|D| \leqslant k$. We show that a routing tree that spans U and has a total energy cost not greater than C can be constructed using D. Let $W' = \{w_i | v_i \in D\}$. We construct $T = (V_T, E_T)$ as follows. Let $V_T = \{r\} \cup W' \cup U$. Set the parent of w_i to r for all $w_i \in W'$ and the parent of u_i to an arbitrary neighboring node in W' for all $u_i \in U$. Since D is a dominating set in G, each node in U has a parent in T. Thus, T is a routing tree spanning U. We show the total energy cost of T is not greater than C. Since $Tx = Rx = 1$, the total transmission (or reception) energy of the nodes in U (or W) is $|U| = |V|$. In addition, since each node in W' receives at most $\Delta(G) + 1$ reports, it sends exactly one packet to the parent in T. Since $|W'| = |D| \leq k$, the total transmission (or reception) energy of the nodes in W (or r) is at most k. Thus, the total energy cost of T is not greater than $2 \cdot (|V| + k) = C$.

For the ”if” part, let $T = (V_T, E_T)$ be a routing tree that spans U and has minimum energy cost not greater than $C = 2(|V| + k)$. Let $W' = V_T \setminus (U \cup \{r\})$. We claim that $D = \{v_i | w_i \in W'\}$ is a dominating set in G with $|D| \leqslant k$. In T, the parent of each node in U is in W'. This implies D is a dominating set in G. In addition, as in the proof of Theorem 1, the parent of each node in W' is r; otherwise, a routing tree that spans U and has less energy cost exists. Since the total transmission (or reception) energy of the nodes in U (or W') is $|V|$, the total transmission (or reception) energy of the nodes in W' (or r) is at most k. Thus, $|D| = |W'| \leqslant k$. ■

B. Approximation Algorithm

A Steiner tree algorithm and a shortest path tree algorithm provide solutions with minimum number of edges and minimum average hop distance from sources to the sink for the MECAT_RN problem, respectively. However, both of them have bad approximation ratios, as described in Theorems 4 and 5. Their proofs are given in the appendix.

Theorem 4. *The approximation ratio of a Steiner tree algorithm is at least* $\Theta(|U|)$.

Theorem 5. *The approximation ratio of a shortest path tree algorithm is at least* $\Theta(|U|)$.

Theorems 4 and 5 tell us that a routing tree with a constant approximation ratio cannot be found by minimizing either the number of edges or the average hop distance from sources to the sink. Our method (Algorithm 2) is to construct a routing tree that approximates both a Steiner tree and a shortest path tree based on Salman's algorithm [23] (Algorithm 1) for the Capacitated Network Design problem [16]. The Capacitated Network Design problem, Salman's algorithm, and the Light Approximate Shortest-path Tree (LAST) [24] used in Salman's algorithm are introduced below.

Problem 4. [16] Given a graph $G = (V, E)$ with weight $w(e) \in \mathbb{R}^+$ associated with each edge $e \in E$ indicating the length and weight $s(u) \in \mathbb{Z}^+$ associated with each source $u \in U \subseteq V$ indicating the demand size to route to sink $r \in V$, a set of sources U, a sink r, and a transmission facility capacity $q \in \mathbb{Z}^+$, the **C**apacitated **N**etwork **D**esign (**CND**) problem is to find a path from u to sink r for each source $u \in U$, such that the total cost of installing all facilities is minimized, where the cost of installing k facilities on an edge with length l is $k \cdot l$. Note that a node might have multiple outgoing edges in a feasible solution of the CND problem. That is, a feasible solution of the CND problem might not be a tree. Moreover, CND(G, U, r, q, C) denotes an instance of the CND problem, and $COST_{CND}(R)$ denotes the cost of installing facilities of a route R.

Definition 3. [24] Given a graph $G = (V, E)$ with weight $w(e) \in \mathbb{R}^+$ associated with each edge $e \in E$, a spanning tree T rooted at r is called an (α, β)**-LAST**, where $\alpha \geqslant 1$ and $\beta \geqslant 1$, if the following two conditions are satisfied:

1) For every node v, the distance from v to r in T is at most α times the minimum distance from v to r in G.
2) The weight of T is at most β times that of the minimum spanning tree of G.

Algorithm 1 : Salman's Algorithm for the CND Problem

```
Input: $G$, $U$, $r$, $q$, $C$
Construct a complete graph $G'$ with node set $U \cup \{r\}$.
Set the weight of each edge $(u, v)$ in $G'$ to the length of the shortest path from $u$ to $v$ in $G$.
Compute a (3,2)-LAST $T_L$ in $G'$.
Let $(u, u_1, \cdots, u_n, r)$ be the shortest path from $u$ to $r$ in $T_L$. Then, the concatenation of paths $P_{u,u_1}$, $P_{u_1,u_2}$, $\cdots$, and $P_{u_n,r}$ is the output path from $u$ to $r$, where $P_{x,y}$ denotes the shortest path from $x$ to $y$ in $G$.
Return the output path from $u$ to $r$ for each $u \in U$.
```

Algorithm 2 : Our Algorithm for the MECAT_RN Problem

```
Input: $G$, $U$, $r$, $q$, $Tx$, $Rx$, $C$
1: Construct a complete graph $G'$ with node set $U \cup \{r\}$.
2: Set the weight of each edge $(u, v)$ in $G'$ to the hop distance from $u$ to $v$ in $G$.
3: Compute a (3,2)-LAST $T_L$ in $G'$.
4: Compute $G” = (V”, E”)$, where $V” = \{w|w \in P_{u,v}$ for some $(u, v) \in T_L\}$, $E” = \{\{x, y\}|\{x, y\} \in P_{u,v}$ for some $(u, v) \in T_L\}$, and $P_{u,v}$ is the shortest path from $u$ to $v$ in $G$.
5: Construct a shortest path tree $T_{SPT}$ rooted at $r$ and spanning $U$ in $G”$.
6: Return $T_{SPT}$.
```

Theorem 6 shows Algorithm 2 is a 7-approximation algorithm of the MECAT_RN problem. Lemma 2, derived from the proof of Lemma 2.1 in [23], is used in the proof of Theorem 6. We omit the proof of Lemma 2 due to the page limit.

Lemma 2. *Let $R = \bigcup_{u \in U} P_{u,r}$ be a route of the CND problem, where $P_{u,r}$ is the routing path from source u to sink r, and let R_{OPT} be the route with minimum cost of the CND problem. Then, $COST_{CND}(R) \leqslant (\alpha' + \beta')COST_{CND}(R_{OPT})$, if the following two conditions are satisfied:*

1) *For every source u, the length of $P_{u,r}$ is at most α' times the minimum distance from u to r in G.*
2) *The total lengths of the edges of R is at most β' times that of the Steiner tree of G spanning U.*

Theorem 6. *Algorithm 2 is a 7-approximation algorithm of the MECAT_RN problem.*

Proof: Let Algorithm $\mathcal{A}$ be obtained from Algorithm 2 by replacing Line 2 with Line 2 of Algorithm 1 and modifying Line 6 to return the path from u to r in T_{SPT} for each $u \in U$ instead of T_{SPT}. We first claim that Algorithm $\mathcal{A}$ is a 7-approximation of the CND problem. Let R_1 and $R_{\mathcal{A}}$ be the solutions generated by Algorithms 1 and $\mathcal{A}$, respectively. Clearly, the following two facts hold:

1) For every source u, the length of $P_{u,r}$ in $R_{\mathcal{A}}$ is less than that in R_1.
2) The total lengths of the edges of $R_{\mathcal{A}}$ is is less than that of R_1.

[23] tells us that the length of $P_{u,r}$ in R_1 is at most 3 times the minimum distance from u to r in G for every source u and the total lengths of the edges of R_1 is at most 4 times that of the Steiner tree of G spanning U. Thus, Algorithm $\mathcal{A}$ is a 7-approximation of the CND problem by Lemma 2.

Next, given MECAT_RN$(G_1, U, r, q, Tx, Rx, C)$, we construct CND$(G_2, U, r, q, C)$, where G_2 is obtained from G_1 by setting the weight of each edge to $Tx + Rx$. Let T_2 and T_{OPT} be a routing tree generated by Algorithm 2 and the routing tree with minimum energy cost for MECAT_RN$(G_1, U, r, q, Tx, Rx, C)$, respectively. Let $R_{\mathcal{A}}$ and R_{OPT} be a route generated by Algorithm $\mathcal{A}$ and the route with minimum cost of installing facilities for CND(G_2, U, r, q, C), respectively. Note that for each $u \in U$, the sequence of the nodes in the path from u to r in T_2 is equal to that in the path from u to r in $R_{\mathcal{A}}$. Thus,

$$COST(T_2) = COST_{CND}(R_{\mathcal{A}}). \tag{14}$$

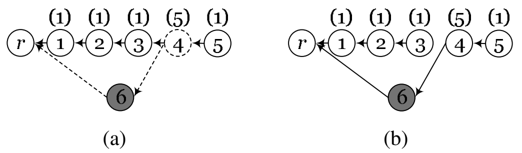
It is also noted that a collection of the path from u to r in T_{OPT} for each $u \in U$ can be a route R for CND(G_2, U, r, q, C), in which case $COST(T_{OPT}) = COST_{CND}(R)$. It implies

$$COST_{CND}(R_{OPT}) \leq COST(T_{OPT}). \tag{15}$$

Combining (14) and (15) together with the fact that $COST_{CND}(R_{\mathcal{A}}) \leq 7COST_{CND}(R_{OPT})$, we obtain

$$COST(T_2) \leq 7COST(T_{OPT}). \tag{16}$$

■



(a) (b)

Fig. 5: Example of Algorithm 3, where node 6 is a relay node. q is equal to 5. Weights are shown in parentheses. (a) The $P_{u,r}$ after the execution of Line 2, where node 4 sends its report along the path (4, 6, 1) and node 5 sends its report along the path (5, 4, 3, 2, 1). (b) The output of Algorithm 3.

V. Discussion

In Section IV-B, we obtain a 7-approximation algorithm of the MECAT_RN problem from Salman's 7-approximation algorithm of the CND problem. In this section, we show any λ-approximation algorithm of the CND Problem $\mathcal{A}$ can be used to obtain a 2λ-approximation algorithm of the MECAT_RN problem, as described in Algorithm 3 and Theorem 7. See Fig. 5 for an example.

Algorithm 3 : CND-Based Algorithm for the MECAT_RN Problem

```
Input: $G$, $U$, $r$, $q$, $Tx$, $Rx$, $C$, $\mathcal{A}$
1: Obtain a graph $G'$ from $G$ by setting the weight of each edge in $G$ to $Tx + Rx$.
2: Execute $\mathcal{A}$ with inputs $G'$, $U$, $r$, $q$, and $C$ to obtain $P_{u,r}$, the path from $u$ to $r$ in $G'$, for each $u \in U$.
3: Compute $G” = (V”, E”)$, where $V” = \{w|w \in P_{u,r}\}$ and $E” = \{\{x, y\}|\{x, y\} \in P_{u,r}\}$.
4: Construct a shortest path tree $T_{SPT}$ rooted at $r$ and spanning $U$ in $G”$.
5: Return $T_{SPT}$.
```

Theorem 7. *Algorithm 3 is a 2λ-approximation algorithm of the MECAT_RN problem, given that $\mathcal{A}$ is a λ-approximation algorithm of the CND problem.*

Proof: Let $P_{u,r}$ be the path from u to r output by algorithm $\mathcal{A}$, and let $R = \bigcup_{u\in U} P_{u,r}$. We have

$$\frac{COST_{CND}(R)}{Tx+Rx} = \sum_{u\in U}\frac{s(u)}{q}l(P_{u,r}) + \sum_{e\in R}(\lceil z(e)\rceil - z(e)), \tag{17}$$

where $z(e) = \sum_{u:e\in P_{u,r}} s(u)/q$ and $l(p)$ is the length of path p. Let T_{SPT} be the routing tree generated by Algorithm 3, and let $R' = \bigcup_{u\in U} P'_{u,r}$, where $P'_{u,r}$ denotes the path from u to r in T_{SPT}. Then,

$$\begin{aligned}\frac{COST(T_{TSP})}{Tx+Rx} &= \sum_{u\in U}\frac{s(u)}{q}l(P'_{u,r}) + \sum_{e\in R'}(\lceil z'(e)\rceil - z'(e)) \\ &< \sum_{u\in U}\frac{s(u)}{q}l(P'_{u,r}) + \sum_{e\in R'} 1,\end{aligned} \tag{18}$$

where $z'(e) = \sum_{u:e\in P'_{u,r}} s(u)/q$. Let R_{OPT} be the route with minimum cost of installing facilities. (17) together with the fact that $COST_{CND}(R) \leqslant \lambda COST_{CND}(R_{OPT})$ implies

$$\sum_{u\in U}\frac{s(u)}{q}l(P'_{u,r}) \leqslant \sum_{u\in U}\frac{s(u)}{q}l(P_{u,r}) \leqslant \frac{\lambda COST_{CND}(R_{OPT})}{Tx+Rx}. \tag{19}$$

In addition,

$$\sum_{e\in R'} 1 \leqslant \sum_{e\in R} 1 \leqslant \frac{COST_{CND}(R)}{Tx+Rx} \leqslant \frac{\lambda COST_{CND}(R_{OPT})}{Tx+Rx}. \tag{20}$$

By (18), (19), and (20), we have

$$COST(T_{TSP}) < 2\lambda COST_{CND}(R_{OPT}). \tag{21}$$

As in the proof of Theorem 6,

$$COST_{CND}(R_{OPT}) \leq COST(T_{OPT}). \tag{22}$$

Combining (21) and (22), we obtain

$$COST(T_{TSP}) < 2\lambda COST(T_{OPT}). \tag{23}$$

■

When all reports have the same size, Hassin *et. al.* propose a $(1+\rho_{st})$-approximation algorithm of the CND problem [16], where ρ_{st} denotes the approximation ratio of the algorithm of the Steiner tree problem. Since a 1.55-approximation algorithm of the Steiner tree problem is proposed [25], so far the best in the literature, we can obtain a 5.1-approximation algorithm of the MECAT_RN problem by Algorithm 3. As the reports have different sizes, the algorithm proposed by Hassin *et. al.* for the CND problem [16] has an approximation ratio $(2+\rho_{st})$, in which case a 7.1-approximation algorithm of the MECAT_RN problem can be obtained by Algorithm 3.

VI. Numerical Results

Two simulations were conducted here. In the first and second simulations, algorithms of the MECAT problem (data aggregation without relay nodes) and the MECAT_RN problem (data aggregation with relay nodes) were compared, respectively. We also compared our algorithms with the lower bound of the minimum energy cost LB evaluated by (24).

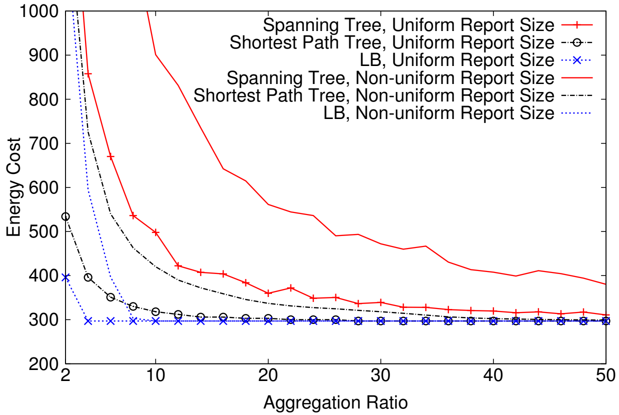


Fig. 6: Effect of aggregation ratio on energy cost under different algorithms for data aggregation without relay nodes.

$$LB = (Tx+Rx)\cdot\max\{\sum_{u\in U}\frac{s(u)}{q}l(u,r), |E(T_{Steiner})|\}, \tag{24}$$

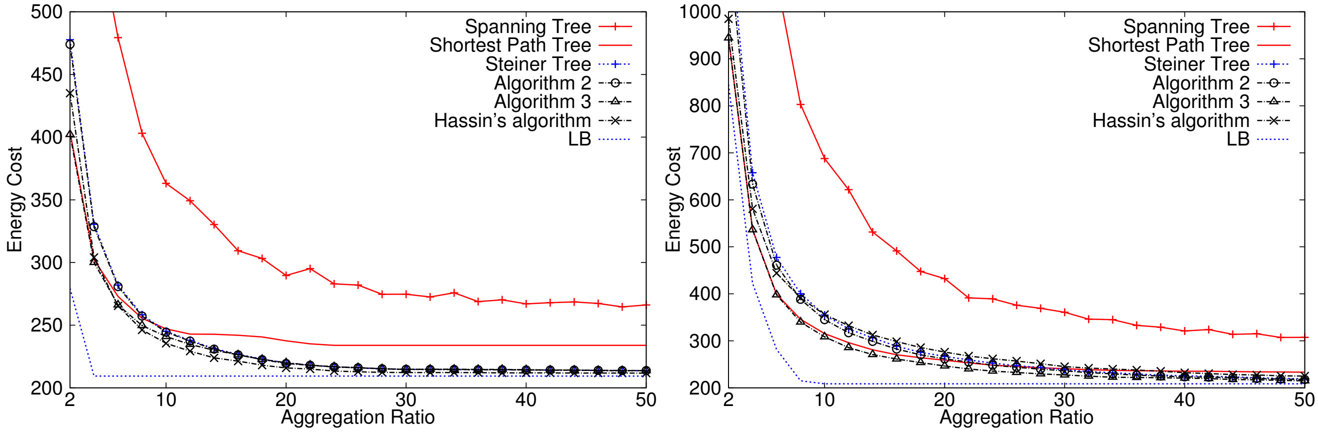
where U is the set of sources, q is the aggregation ratio, $l(u,r)$ is the hop distance from u to r in a shortest path tree and $|E(T_{Steiner})|$ is the number of edges in a Steiner Tree. LB is evaluated by (24) due to the fact that the corresponding minimum energy cost of the MECAT problem and the MECAT_RN problem is impossible to be smaller than each of $(Tx+Rx)\cdot\sum_{u\in U}\frac{s(u)}{q}l(u,r)$ and $(Tx+Rx)\cdot|E(T_{Steiner})|$. Since a Steiner tree cannot be obtained in polynomial time, we use a 2-approximation algorithm to construct a Steiner tree $T(V_{ST}, E_{ST})$ [26], and evaluate $|E(T_{Steiner})|$ by (25).

$$|E(T_{Steiner})| = \max\{\frac{|E_{ST}|}{2}, |U|\}. \tag{25}$$

In a wireless sensor network, 100 sensor nodes were uniformly deployed in a 100 × 100 field. A link exists between two sensor nodes with distance less than or equal to the transmission range $R = 20$. Since the transmission power is about two times the reception power [27], Tx and Rx are set to 2 and 1, respectively. If all reports have the same size (uniform report size), the size is set to 1; otherwise (non-uniform report size), the sizes are randomly set to range from 1 to 5. The energy cost of each algorithm was evaluated under different aggregation ratios 2, 4, 6, $\cdots$, 50. In the second simulation, each node has probabilities 0.7 and 0.3 to be a source and a relay node, respectively. Empirical data were obtained by averaging data of 30 different networks. Table I summarizes the simulation settings.

A. Results for Data Aggregation without Relay Nodes

Fig. 6 shows the energy cost of algorithms for data aggregation without relay nodes under different aggregation ratios. It can be seen that the shortest path tree algorithm (the proposed algorithm for the MECAT problem) significantly outperforms the spanning tree algorithm. When the aggregation ratio is greater than or equal to the sum of the sizes of the reports



(a)

(b)

Fig. 7: Effect of aggregation ratio on energy cost under different algorithms for data aggregation with relay nodes. (a) Uniform report size. (b) Non-uniform report size.

TABLE I: Simulation Settings

Number of nodes	100
Field	100 × 100
Sink location	(50, 50)
R (transmission range)	20
q (aggregation ratio)	2, 4, 6, ..., 50
Tx (transmission energy cost per packet)	2
Rx (reception energy cost per packet)	1
Uniform report size	1
Non-uniform report size	1, 2, 3, 4, 5
Probability of being relay nodes	0.3

sent by most of the nodes, the energy cost of each algorithm approaches $(Tx + Rx) \cdot |V| = 300$.

B. Results for Data Aggregation with Relay Nodes

Fig. 7 shows the energy cost of algorithms for data aggregation with relay nodes under different aggregation ratios. Algorithms 2 and 3 (the proposed algorithms for the MECAT_RN problem) construct data aggregation trees based on Salman's algorithm [23] and Hassin's algorithm [16], respectively. A non-tree routing structure established by Hassin's algorithm [16] is also compared here. Four observations are noteworthy. First, the energy cost of each algorithm approaches LB as the aggregation ratio is great. Second, although a shortest path tree algorithm and a Steiner tree algorithm have bad performances in the worst cases (see Theorems 4 and 5), they have good performances in average cases. Third, a shortest path tree performs better and worse than a Steiner tree algorithm when the aggregation ratio is small and great, respectively. This is because as the aggregation ratio is great, a packet can carry a large number of reports, and thus, the energy cost highly depends on the number of edges in the data aggregation tree. On the contrary, the energy cost highly depends on the lengths of the paths from sources to the sink as the aggregation ratio is small. Fourth, Algorithm 3 outperforms Hassin's algorithm as the reports have different sizes, in which case Hassin's algorithm usually utilizes only half of the size of a packet, and in contrast, Algorithm 3 utilizes the packet efficiently.

VII. Related Works

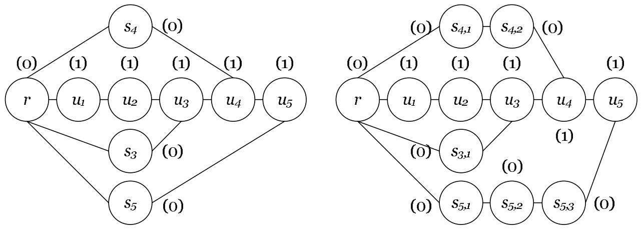
In [7], an algorithm is demonstrated to find the best shortest path tree that maximizes the network lifetime. In [8], the authors prove the problem of finding an optimal aggregation tree that maximizes the network lifetime is NP-complete and propose an approximation algorithm. In [10], a randomized $O(1)$-approximation algorithm is given to construct a simultaneous optimal aggregation tree based on the geographic correlation of reports. In [11], a data gathering tree is constructed to overcome the changes of the network topology.

In [1], the problem of finding a routing structure minimizing the number of transmitted packets is studied, in which neither the proof of NP-completeness nor an approximation algorithm is given. They show that routing packets on any two shortest path trees does not significantly affect the effectiveness of data aggregation. In addition, all reports are assumed to have the same size and the existence of relay nodes are not taken into consideration, which is different from this paper.

A problem similar to ours is studied in [9], but the aggregation model is different. It is assumed that any j reports can be aggregated into $f(j)$ packets, where f is concave and non-decreasing, and $f(0) = 0$. However, in our data aggregation model, $f(j) = \lceil \frac{j}{q} \rceil$ is not concave. Thus, the analysis in [9] cannot be applied here.

VIII. Conclusion

In this paper, we study the problem of constructing energy-efficient data aggregation trees. Two types of this problem are investigated: the one without relay nodes and the one with relay nodes. Both of them are shown to be NP-complete. For the problem without relay nodes, we find that a shortest path tree algorithm turns out to be a 2-approximation algorithm and can be easily implemented in a distributed manner. For the problem with relay nodes, we first show that a shortest path tree algorithm and a Steiner tree algorithm each have



(a) (b)

Fig. 8: Instances of the MECAT_RN problem. (a) An instance with the approximation ratio of a Steiner tree algorithm $\Theta(|U|)$. (b) An instance with the approximation ratio of a shortest path tree algorithm $\Theta(|U|)$.

bad performance in the worst cases. We then obtain an $O(1)$-approximation algorithm by constructing a shortest path tree on the routing structure of the Capacitated Network Design problem. Simulations show that the proposed algorithms each have good performance in terms of the energy cost. Simulations also show that for data aggregation with relay nodes, a tree might outperform a non-tree structure in terms of the energy cost. The reason is in a tree, the data is concentrated in a small number of nodes, resulting in efficient utilization of packets.

Appendix

A. Proof of Theorem 4

Proof: Consider one kind of instance of the MECAT_RN problem, MECAT_RN(G, U, r, q, Tx, Rx, C), where $G = (V, E)$ with weights 1 and 0 associated with each node $u \in U$ and $v \in V \setminus U$, respectively, $U = \{u_1, u_2, \cdots, u_{|U|}\}$, $V = \{r\} \cup U \cup \{s_i | 3 \leq i \leq |U|\}$, $E = \{\{r, u_1\}\} \cup \{\{u_i, u_{i+1}\} | 1 \leq i \leq |U| - 1\} \cup \{\{r, s_i\}, \{s_i, u_i\} | 3 \leq i \leq |U|\}$, $q = 2$, and $Tx = Rx = 1$. See Fig. 8(a), for the instance with $|U| = 5$. Clearly, $T_S = (V_S, E_S)$ constructed by a Steiner tree algorithm has energy cost $\Theta(|U|^2)$, where $V_S = \{r\} \cup U$ and $E_S = \{(u_1, r)\} \cup \{(u_{i+1}, u_i) | 1 \leq i \leq |U| - 1\}$. However, $T_O = (V_O, E_O)$ has energy cost $\Theta(|U|)$, where $V_O = V$ and $E_O = \{(u_2, u_1), (u_1, r)\} \cup \{(u_i, s_i), (s_i, r) | 3 \leq i \leq |U|\}$. ■

B. Proof of Theorem 5

Proof: Consider one kind of instance of the MECAT_RN problem, MECAT_RN(G, U, r, q, Tx, Rx, C), where $G = (V, E)$ with weights 1 and 0 associated with each node $u \in U$ and $v \in V \setminus U$, respectively, $U = \{u_1, u_2, \cdots, u_{|U|}\}$, $V = \{r\} \cup U \cup \{s_{i,j} | 3 \leq i \leq |U|, 1 \leq j \leq i - 2\}$, $E = \{\{r, u_1\}\} \cup \{\{u_i, u_{i+1}\} | 1 \leq i \leq |U| - 1\} \cup \{\{r, s_{i,1}\}, \{s_{i,i-2}, u_i\} | 3 \leq i \leq |U|\} \cup \{\{s_{i,j}, s_{i,j+1}\} | 3 \leq i \leq |U|, 1 \leq j \leq i - 3\}$, $q = |U|$, and $Tx = Rx = 1$. See Fig. 8(b), for the instance with $|U| = 5$. Clearly, $T_S = (V_S, E_S)$ constructed by some shortest path tree algorithm has energy cost $\Theta(|U|^2)$, where $V_S = V$ and $E_S = \{(u_2, u_1), (u_1, r)\} \cup \{(s_{i,1}, r), (u_i, s_{i,i-2}) | 3 \leq i \leq |U|\} \cup \{(s_{i,j+1}, s_{i,j}) | 3 \leq i \leq |U|, 1 \leq j \leq i - 3\}$. However, $T_O = (V_O, E_O)$ has energy cost $\Theta(|U|)$, where $V_O = \{r\} \cup U$ and $E_O = \{(u_1, r)\} \cup \{(u_{i+1}, u_i) | 1 \leq i \leq |U| - 1\}$. ■

References

[1] C. Liu and G. Cao, "Distributed monitoring and aggregation in wireless sensor networks," in *IEEE INFOCOM*, 2010.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *ACM WSNA*, 2002.

[3] N. Xu, S. Rangwala, K. K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin, "A wireless sensor network for structural monitoring," in *ACM SenSys*, 2004.

[4] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 755–764, 2005.

[5] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *IEEE INFOCOM*, 2004.

[6] J. Li, A. Deshpande, and S. Khuller, "On computing compression trees for data collection in wireless sensor networks," in *IEEE INFOCOM*, 2010.

[7] D. Luo, X. Zhu, X. Wu, and G. Chen, "Maximizing lifetime for the shortest path aggregation tree in wireless sensor networks," in *IEEE INFOCOM*, 2011.

[8] Y. Wu, S. Fahmy, and N. B. Shroff, "On the construction of a maximum-lifetime data gathering tree in sensor networks: NP-Completeness and approximation algorithm," in *IEEE INFOCOM*, 2008.

[9] A. Goel and D. Estrin, "Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk," in *SODA*, 2003.

[10] M. Enachescu, A. Goel, R. Govindan, and R. Motwani, "Scale free aggregation in sensor networks," in *ALGOSENSORS*, 2004.

[11] N. Thepvilojanapong, Y. Tobe, and K. Sezaki, "On the construction of efficient data gathering tree in wireless sensor networks," in *IEEE ISCAS*, 2005.

[12] B. Yu, J. Li, and Y. Li, "Distributed data aggregation scheduling in wireless sensor networks," in *IEEE INFOCOM*, 2009.

[13] C. Park and P. H. Chou, "Ambimax: Autonomous energy harvesting platform for multi-supply wireless sensor nodes," in *IEEE SECON*, 2006.

[14] X. Jiang, J. Polastre, and D. Culler, "Perpetual environmentally powered sensor networks," in *IPSN*, 2005.

[15] P. Dutta, J. Hui, J. Jeong, S. Kim, C. Sharp, J. Taneja, G. Tolle, K. Whitehouse, and D. Culler, "Trio: Enabling sustainable and scalable outdoor wireless sensor network deployments," in *IPSN*, 2006.

[16] R. Hassin, R. Ravi, and F. S. Salman, "Approximation algorithms for a capacitated network design problem," *Algorithmica*, vol. 38, pp. 417–431, 2004.

[17] C. P. Low, "An approximation algorithm for the load-balanced semi-matching problem in weighted bipartite graphs," *Information Processing Letters*, vol. 100, pp. 154 – 161, 2006.

[18] M.-J. Tsai, H.-Y. Yang, and W.-Q. Huang, "Axis-based virtual coordinate assignment protocol and delivery-guaranteed routing protocol in wireless sensor networks," in *IEEE INFOCOM*, 2007.

[19] X. Cheng, D.-Z. Du, L. Wang, and B. Xu, "Relay sensor placement in wireless sensor networks," *Wireless Networks*, vol. 14, pp. 347–355, 2008.

[20] X. Han, X. Cao, E. L. Lloyd, and C.-C. Shen, "Fault-tolerant relay node placement in heterogeneous wireless sensor networks," in *IEEE INFOCOM*, 2007.

[21] S. Misra, S. D. Hong, G. Xue, and J. Tang, "Constrained relay node placement in wireless sensor networks to meet connectivity and survivability requirements," in *IEEE INFOCOM*, 2008.

[22] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman, 1979.

[23] F. S. Salman, J. Cheriyan, R. Ravi, and S. Subramanian, "Approximating the single-sink link-installation problem in network design," *SIAM J. on Optimization*, vol. 11, pp. 595–610, 2000.

[24] S. Khuller, B. Raghavachari, and N. Young, "Balancing minimum spanning trees and shortest-path trees," *Algorithmica*, vol. 14, pp. 305–321, 1995.

[25] G. Robins and A. Zelikovsky, "Improved steiner tree approximation in graphs," in *SODA*, 2000.

[26] L. Kou, G. Markowsky, and L. Berman, "A fast algorithm for steiner trees," *Acta Informatica*, vol. 15, pp. 141–145, 1981.

[27] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication paradigm for sensor networks," in *ACM MobiCom*, 2000.