Data Collection Maximization in Renewable Sensor Networks via Time-Slot Scheduling

Xiaojiang Ren, *Student Member, IEEE*, Weifa Liang, *Senior Member, IEEE*, and Wenzheng Xu

Abstract—In this paper we study data collection in an energy renewable sensor network for scenarios such as traffic monitoring on busy highways, where sensors are deployed along a predefined path (the highway) and a mobile sink travels along the path to collect data from one-hop sensors periodically. As sensors are powered by renewable energy sources, time-varying characteristics of ambient energy sources poses great challenges in the design of efficient routing protocols for data collection in such networks. In this paper we first formulate a novel data collection maximization problem by adopting multi-rate data transmissions and performing transmission time slot scheduling, and show that the problem is NP-hard. We then devise an offline algorithm with a provable approximation ratio for the problem by exploiting the combinatorial property of the problem, assuming that the harvested energy at each node is given and link communications in the network are reliable. We also extend the proposed algorithm by minor modifications to a general case of the problem where the harvested energy at each sensor is not known in advance and link communications are not reliable. We thirdly develop a fast, scalable online distributed algorithm for the problem in realistic sensor networks in which neither the global knowledge of the network topology nor sensor profiles such as sensor locations and their harvested energy profiles is given. Furthermore, we also consider a special case of the problem where each node has only a fixed transmission power, for which we propose an exact solution to the problem. We finally conduct extensive experiments by simulations to evaluate the performance of the proposed algorithms. Experimental results demonstrate that the proposed algorithms are efficient and the solutions obtained are fractional of the optimum.

Index Terms—Time-slot scheduling, approximation algorithms, online distributed algorithms, energy renewable sensor networks, mobile sinks, data collection, generalized assignment problems

1 Introduction

Wireless sensor network has emerged as a key technology for various applications such as environmental sensing, structural health monitoring, and area surveillance [10]. However, the limited energy in sensor batteries has hampered the large-scale deployment of such a network. A promising solution against the limited energy supplies is to enable sensor nodes to harvest ambient energy from their surroundings [18], [26] such as solar energy and wind energy. In addition to being environmentally friendly, renewable energy could also enable sensor nodes to operate perpetually, eliminating the cost for batteries [19].

In this paper, we consider an energy renewable sensor network deployed along a pre-defined path for surveillance or monitoring. Such an application can be a highway traffic surveillance, where sensors are deployed along both sides of a highway for traffic monitoring to get traffic related information such as the number of vehicles, types of vehicles, and individual vehicle speeds, which can later be used for road usage and maintenance, and driver behavior analysis. Another potential application scenario is the eco-system monitoring in a forest, e.g., such a network can be deployed for monitoring exotic plant growths and/or endangered animals (e.g., giant panda) existence and behavior observations, where humans or vehicles can only access the limited roads rather than everywhere in the forest. Also, a vehicle can receive sensing data from a sensor if the vehicle is within the transmission range of the sensor. The sensors that can communicate with the vehicle usually serve as gateways where the other sensors will forward their sensing data to them through multi-hop relays. There are many other applications that are also fitted in this application scenarios such as oil/gas/water pipeline monitoring [17], structural health monitoring for bridges [30], etc. We assume that each sensor in the network is powered by renewable energy (e.g., solar energy) to avoid its energy expiration. We will employ a mobile sink (e.g. a vehicle) to periodically travel along the pre-defined path at a constant speed to collect data from its one-hop sensors as it has been demonstrated that sink mobility can significantly improve various network performance, including reducing the energy consumption of sensors, balancing the workload among sensors, reducing data delivery delays, and improving network coverage [3], [4], [8], [11], [16], [23]. However, the time-varying characteristics of energy renewable sources poses a great challenge in the design of routing protocols for energy renewable sensor networks, that is, how to design a routing protocol for renewable sensor networks such that the volume of collected data is maximized, under the dynamic energy replenishment constraint. Specifically, the following issues must be addressed when a mobile sink

- *X. Ren and W. Liang are with the Research School of Computer Science, The Australian National University, Canberra, ACT0200, Australia. E-mail: richard.rxj@anu.edu.au, wliang@cs.anu.edu.au.*
- *W. Xu is with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 51006, P. R. China, and the Research School of Computer Science, The Australian National University, Canberra, ACT 0200, Australia. E-mail: wenzheng.xu3@gmail.com.*

Manuscript received 7 Oct. 2013; revised 11 June 2014; accepted 4 Aug. 2014. Date of publication 18 Aug. 2014; date of current version 10 June 2015. Recommended for acceptance by P. Bellavista.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TC.2014.2349521

0018-9340 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is used for data collection in a renewable sensor network. (1) Due to the time-varying characteristics of energy renewable sources, the energy replenishment rate of each sensor is unknown in advance, the sensor thus must have its time-varying energy budget (amount of available energy) for transmitting data to avoid its energy expiration. (2) For a given sensor, it requires using different data transmission rates to transmit its data when the mobile sink is at different locations, while different transmission rates will consume different amounts of its transmission energy. (3) During each tour of the mobile sink, it is very likely that multiple sensors can communicate to the mobile sink at the same time. Simultaneous transmissions of these sensors will result in a collision at the mobile sink and none of the transmissions will succeed. In this paper we will address these issues by scheduling sensors at which time slots to transmit their data to the mobile sink so that the accumulative volume of the data collected by the mobile sink per tour is maximized. We achieve this through incorporating the time-varying sensor energy budget and employing multi-rate wireless communications.

Our main contributions in this paper are as follows. We consider data collection in an energy renewable sensor network, using a path-constrained mobile sink. We first formulate a novel data collection maximization problem by incorporating multi-rate transmissions and transmission time slot scheduling, and show the NP-hardness of the problem. We then devise an offline algorithm with a provable approximation ratio for the problem, assuming that the global knowledge of the network and sensor profiles (their locations and available energy) are given. We also extend the proposed algorithm by minor modifications to solve a generalized case of the problem where the harvested energy at each sensor is not given and link communications are unreliable. We thirdly develop a fast, scalable online distributed algorithm for the problem without the global knowledge of the network and sensor profiles, which is more suitable for real distributed sensor networks. For a special case of the problem where each sensor has a fixed transmission power, we propose an exact solution for it. We finally conduct extensive experiments by simulations to evaluate the performance of the proposed algorithms. Experimental results demonstrate that the proposed algorithms are very promising and the solutions obtained are fractional of the optimum.

To the best of our knowledge, unlike most existing studies of data collection in renewable sensor networks that either formulated it as an integer linear programming (ILP) or provided heuristic solutions, the proposed algorithm is the first approximation algorithm for the problem, by exploiting the combinatorial property of the problem. To respond to the time-varying nature of energy harvesting, traditional ILP methods take too much time and suffer poor scalability. Worst of all, the solutions obtained may not be applicable due to their insensitivities to the dynamic changes of harvesting energy sources. On the other hand, although the heuristic solutions can be found quickly, there are no guarantees of the solutions from the optimal one. In comparison with its conference version [36], this extended paper has the following improvements: (1) Since wireless communication usually is unreliable, we show how to extend the proposed algorithm for solving this general case (see Sections 5.3 and 6.3). (2) We make use of a novel prediction technique to predict the amount of harvested energy at each sensor (see Section 5.3), and later experimental results show that the prediction is quite accurate (see Section 8.2). (3) We evaluate the performance of proposed algorithms against an existing algorithm through extensive experimental simulations, and the experimental outcomes show that the proposed algorithms outperform the existing one (see Section 8).

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 introduces the system model, notions, problem definition. Section 4 shows the NP-completeness of the problem. Section 5 devises an offline approximation algorithm with a provable approximation ratio for the problem. Sections 6 develops a fast, scalable online distributed algorithm, and Section 7 devises an exact solution to the problem when each sensor has only one fixed transmission power. Section 8 evaluates the performance of the proposed algorithms through experimental simulations, and Section 9 concludes the paper.

2 Related Work

Most existing solutions in such networks assumed that the collected data is routed to a fixed sink through multi-hop relays [13], [24], [28], [29], [40], [45]. For example, Liu et al. [28], [29] formulated the problem as a lexicographic maxmin rate allocation problem, and provided a centralized algorithm for the problem by solving an integer linear program. Liang et al. [24] developed a fair rate allocation algorithms by incorporating temporal-spatial sensing data correlations. Zhang et al. [45] studied the problem as a utility maximization problem by representing the utility gain at each sensor node as a concave utility function. They proposed an efficient algorithm for finding the accumulative sum of utility gains in a tree network. Although the data collection paradigm based on fixed sinks may be applicable to small to mediate size networks, it is definitely not suitable for large-scale networks due to long delays on data delivery by multi-hop relay, limited communication bandwidth, etc. To mitigate the deficiencies brought by fixed sinks, a feasible solution is to introduce mobile sinks.

Sink mobility in conventional sensor networks has been extensively studied in the past few year and demonstrated that it can significantly improve various network performance including reducing the energy consumption of sensors, balancing the workload among the sensors, reducing the data delivery delays, and prolonging the network lifetimes [5], [8], [11], [16], [23], [25], [39], [42], [43], [44]. Most existing studies focused on minimizing the energy consumption so as to prolong the network lifetime since sensors are powered by energy-limited batteries. The use of a path-constrained mobile sink for data collection in conventional sensor networks has been well studied. For example, Kansal et al. [20], [38] addressed a network infrastructure based on the use of a path-constrained mobile sink for data collection, where a sensor sends its data to the sink along a minimum number of hops routing path. They proposed a speed control algorithm to maximize the volume of data collected.

Assuming that the mobile sink moves at a constant speed, Gao et al. [16] addressed the energy minimization problem by proposing a novel data collection scheme, where sensors close to the trajectory of the mobile sink are chosen as 'subsinks' and other sensors make use of different subsinks for their data relays. They formulated the subsink choice problem as a problem of minimizing the number of hops from each sensor to its subsink by providing a heuristic solution. They also studied time slot allocations for subsinks when the mobile sink collect data from the subsinks. Chakrabarti et al. [8] considered the dependence of transmission setting and packet loss rate of the mobile data collection problem by modeling the process of data collection as an M/D/1 queue. They then proposed an algorithm that ensures adequate data collection and minimizes the energy consumption. Liang et al. [25] considered another data collection problem by assuming the subsinks (the gateways) are given in advance, they devised several approximation algorithms for the problem, by formulating the problem as a minimum cost capacitated forest problem that finds a minimum cost capacitated forest consisting of routing trees rooted at gateways and spanning all sensors. Unlike the mentioned work in conventional sensor networks that focused on energy conservation to prolong the network lifetime, maximizing network lifetime is no longer a main issue for energy renewable sensor networks as the sensors can be continuously recharged by renewable energies. Thus, in principle, such networks can be operational perpetually. Unfortunately, very little attention has been paid to data collection in energy renewable sensor networks, by using mobile sinks [34], [35]. Ren and Liang [34] considered this problem by assuming that the mobile sink sojourns at some strategic locations and the mobile sink only collects the sensing data from one-hop sensors. Recently they [35] further extended their work for quality-data collection by developing a heuristic and a distributed algorithms, assuming that a mobile sink travels along a pre-defined track for data collection. Furthermore, Ren and Liang [37] studied the quality data collection maximization problem in energy harvesting sensor networks.

3 Preliminaries

3.1 System Model

We consider an energy renewable sensor network $G = (V \cup \{s\}, E)$ where V is a set of n stationary sensors that are densely deployed along a pre-defined path, and a mobile sink s periodically travels along the path at a constant speed r_s without stops to collect data from one-hop sensors. Each sensor is powered by renewable energy (e.g., solar energy) and has stored enough sensing data for collection. There is a link in E between a sensor $v \in V$ and the mobile sink s when they are within the transmission range of each other. Assume that the maximum transmission range of each sensor is R, and the length of the pre-defined path is L. The duration per tour by the mobile sink is determined by its traveling speed r_s, which is referred to as the *data latency*. That is, the faster the mobile sink travels, the shorter the duration per tour is, resulting in a shorter delay on data delivery from its generation to its collection by the mobile sink.

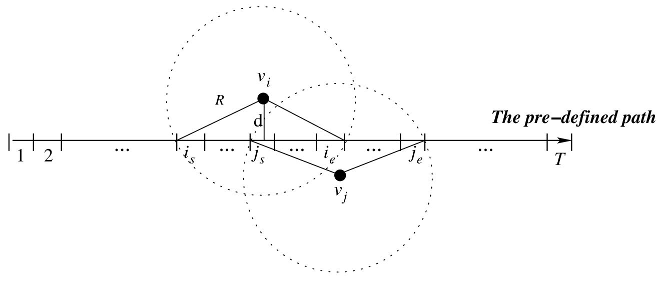


Fig. 1. An illustration of time slots covered by sensors v_i and v_j.

We here adopt a discrete-time system where the duration per tour is slotted into equal time slots with each lasting τ time units [27]. Given the mobile sink speed r_s, the number of time slots per tour can be determined, which is $T = \lceil \frac{L}{r_s \cdot \tau} \rceil$, where L is the length of the pre-defined path. We index the T time slots by $1, 2, \ldots, T$. Let $A(v)$ represent the set of consecutive time slots in which the data transmitted by sensor $v \in V$ can be collected by the mobile sink. Then, $A(v)$ will be determined by the maximum transmission range R of v and its distance from the pre-defined path. Fig. 1 uses an example to illustrate this concept. Given two sensors v_i and v_j, then $A(v_i) = \{i_s, i_s + 1, \ldots, i_e - 1, i_e\}$ and $A(v_j) = \{j_s, j_s + 1, \ldots, j_e - 1, j_e\}$ are the sets of time slots in which they can transmit their data to the mobile sink, $1 \le i_s \le i_e \le T$ and $1 \le j_s \le j_e \le T$. Notice that if $A(v_i) \cap A(v_j) \neq \emptyset$, they share some time slots at which they both can transmit their data to the mobile sink. However, following wireless communication interference model [41], the mobile sink at any given time slot can receive the data from one sensor only; otherwise, none of the transmitted data can be received by the mobile sink due to the channel interference. Thus, we need to allocate these time slots to the sensors such that each time slot is allocated to one sensor only with an objective to maximize the amount of data collected by the mobile sink.

3.2 Energy Model

As sensors are powered by renewable energy, the amount of energy harvested by a sensor at each different time slot is different. This implies that a sensor cannot transmit its data to the mobile sink without any restriction. In principle, a given sensor v can transmit its data to the mobile sink in all time slots in $A(v)$ if it has sufficient energy to support it doing so. However, it may not have enough energy at this moment to achieve that. Following a widely adopted assumption of renewable energy replenishment, we assume that the energy replenishment rate of each sensor is much slower than its energy consumption rate, and the amount of energy harvested in a future time period is uncontrollable but predictable [28]. Denote by $B(v)$ the battery capacity of each sensor v, and denote by $P_j(v)$ and $RE_j(v)$ the amounts of available energy of node v prior to and after tour j, respectively. Thus, sensor v consumes the amount of energy $P_j(v) - RE_j(v)$ for transmitting its data to the mobile sink in tour j. Let $Q_j(v)$ be the amount of harvested energy of sensor v between the $(j-1)$th tour and the jth tour, $P_j(v)$ thus can be expressed as $\min\{RE_{j-1}(v) + Q_j(v), B(v)\}$, where $0 \le P_j(v) \le B(v)$. Furthermore, to support long-term, continuous monitoring service, we assume that sensors should not consume more energy than they can collect in order to achieve ‘perpetual’ operations [19]. Hence, without loss of generality, we refer to $P_j(v)$ as the *energy budget of sensor* v at tour j. We also refer to $P(v)$ as the energy budget of sensor v per tour.

3.3 Communication Model

It is known that wireless signal suffers from path loss, fading, shadowing, interference and other impairments. The communication reliability of a receiver usually is determined by its received Signal-to-Noise Ratio (SNR). The communication reliability of the mobile sink can be maximized if a sensor uses its maximum transmission power level to transmit its data to the sink, this however incurs unnecessarily high energy consumption of the sensor. Motivated by the fact that radio hardware such as CC2500 RF Transceiver [6] allows not only adjusting its transmission power levels but also setting multiple data transmission rates, a multi-rate communication model between each sensor v_i and the mobile sink s is adopted. That is, let $P = \{P_{i,1}, P_{i,2}, \ldots, P_{i,l_i}\}$ be the set of transmission power levels and $R_i = \{r_{i,1}, r_{i,2}, \ldots, r_{i,k_i}\}$ the set of data transmission rates of sensor v_i. Given a time slot (i.e., the mobile sink is located at this moment), sensor v_i could adopt a pair of a transmission power level and a data transmission rate for its data transfer at the time slot if the Signal-to-Noise Ratio at the mobile sink is no less than a given threshold. Thus, there may have many such pairs that sensor v_i can adopt at a given time slot. However, in practice sensor v_i only adopts one specific pair by its PHY/MAC layer protocol that ensures that data can be received by the receiver reliably within the distance between the sender and the receiver. For the sake of simplicity, we assume that the pair of a transmission power level and a data transmission rate of each sensor at each time slot is given.

3.4 Approximation Algorithm

We say an algorithm for a maximization optimization problem is an α-approximation algorithm if the ratio of the approximate solution to the optimal solution is no less than α, where α is a constant with $0 < \alpha < 1$.

3.5 Problem Definition

Given an energy renewable sensor network G and T time slots per tour in which the mobile sink travels along with a pre-defined path to collect data from one-hop sensors, the *data collection maximization problem* is to maximize the volume of the data collected by the mobile sink through allocating the T time slots to individual sensors, under the constraints on both the energy replenishment rate and multi-rate data transmission rate at each time slot.

Intuitively, each sensor should transmit its data to the mobile sink at all available time slots to it in order to maximize its share on the collected data, thereby maximizing the volume of the data collected from the entire network. However, since the energy replenishment rate of each sensor is much slower than its energy consumption rate, each sensor may only make use of some of all available time slots to transmit its data due to its energy budget. What followed is which time slots it should choose for its data transmission. Since the sensor at different time slots will have different data transmission rates, this results in different amounts of its transmission energy consumption. Furthermore, it is very likely that multiple sensors sharing the same time slot will compete with each other for the time slot to transmit their own data, as sensors in the network are densely deployed. Thus, allocating each shared time slot to one of the competing sensors so as to maximize the accumulative data volume is a challenging task.

In other words, the data collection maximization problem in G can be described as follows. Given T time slots and a pre-defined path, the mobile sink travels along the path at a given constant speed to collect data from one-hop sensors. Associated with each sensor $v_i \in V$, there are $|A(v_i)|$ potentially available time slots for sensor v_i to transfer its data to the mobile sink, where $r_{i,j}$ is the average data transmission rate of v_i if it transmits data at time slot $j \in A(v_i)$ with the amount of energy consumption $P_{i,j} \cdot \tau$. We further assume that the number of different transmission rates of each sensor v_i, $r'_{i,1}, r'_{i,2}, \ldots, r'_{i,k_i}$, is given and $r'_{i,x} < r'_{i,y}$ if $1 \le x < y \le k_i$. Usually, k_i is a fixed integer. To ensure that the transmitted data can be received by the receiver successfully, the use of a different transmission rate for data transmission will consume a different amount of power of sensor v_i. For the sake of convenience, in the rest of the paper we assume that all sensors have the same number of transmission power levels k, i.e., $k_i = k$ for all i with $1 \le i \le n$. Also, it is well known that wireless communication is unreliable. In this paper we thus assume that the *link reliability* of the link between sensor v_i and the mobile sink at time slot j is $\rho_{i,j}$ with $0 \le \rho_{i,j} \le 1$ and $1 \le j \le T$. The data collection maximization problem in G thus is to allocate a subset of time slots $A'(v_i)$ ($\subseteq A(v_i)$) to the sensors such that the volume of data transmitted from all sensors, $\sum_{v_i \in V} \sum_{j \in A'(v_i)} (\rho_{i,j} \cdot r_{i,j} \cdot \tau)$ is maximized, subject to (i) each time slot is allocated to only one sensor if there are multiple sensors sharing the time slot; (ii) the total energy consumption of each sensor v_i per tour is no more than its energy budget $P(v_i)$, i.e., $\sum_{j \in A'(v_i)} P_{i,j} \cdot \tau \le P(v_i)$, where τ is the duration of each time slot and $A'(v_i) \subseteq A(v_i)$ for all i with $1 \le i \le n$.

4 NP-Hardness

We show that the data collection maximization problem is NP-hard by the following theorem.

Theorem 1. *The data collection maximization problem in an energy renewable sensor network is NP-hard.*

Proof. We show the claim by a reduction from a well known NP-complete problem—the generalized assignment problem (GAP), which is defined as follows. Given a set of bins and a set of items that have a different size and profit for each bin, pack a maximum profit subset of items into the bins. In other words, let $A = \{a_1, a_2, \ldots a_m\}$ be a set of m items and $B = \{B_1, B_2, \ldots B_n\}$ a set of bins, where each B_i has a capacity b_i for all i with $1 \le i \le n$. Assigning item a_j to bin B_i will consume the amount of resource $b_{i,j}$ of B_i, and the benefit brought by this assignment is $c_{i,j}$. The objective is to allocate the items in A to the bins in B such that the total profit is maximized, subject to the total amount of resources consumed of each bin B_i being no more than its capacity b_i for all i with $1 \leq i \leq n$.

We now show that a special case of the data collection maximization problem is equivalent to the defined GAP problem, where the equivalence means that a solution to one of them is the solution to another as well. The data collection maximization problem is given as follows: we assume that the maximum transmission range of each sensor R is large enough to cover the entire tour path, wireless communication is most reliable, i.e., the link reliability of each link is 1. We proceed the following reduction.

Each item in A corresponds a time slot, thus the set of items corresponds to the set of time slots. Each bin B_i in B corresponds to a sensor $v_i \in V$, the capacity b_i of B_i corresponds to the energy budget of sensor v_i, $P(v_i)$, to perform its data transmission for a certain number of time slots in $A(v_i)$, and $P_{i,j} \cdot \tau$ is the amount of transmission energy consumed by v_i if it sends its data to the mobile sink at time slot a_j, i.e., the amount of its resource consumed. The profit brought by allocating time slot a_j to sensor v_i is $c_{i,j}$ $(= r_{i,j} \cdot \tau)$, which is the amount of data transmitted, where $r_{i,j}$ is the average data transmission rate of v_i at time slot a_j, which usually is determined by the Euclidean distance $d_{i,j}$ between v_i and the mobile sink at time slot a_j and the transmission power adopted by v_i. This implies that at different time slots, different data transmission rates will be adopted, thereby leading to different amounts of data collected by the mobile sink. Allocating the T time slots to the n sensors such that the amount of data collected by the mobile sink is maximized is equivalent to maximizing the profit in the GAP. Hence, the data collection maximization problem is NP-hard. □

5 An Offline Approximation Algorithm

Since the data collection maximization problem is NP-hard, in this section we devise an approximation algorithm with a provable approximation ratio for the problem, by exploiting the combinatorial property of the problem, provided that the mobile sink has the global knowledge of the network topology and the profile of each sensor (e.g., the energy budget of each sensor at the current tour, the location of the sensor, the starting and ending time slots of the sensor, etc).

For the sake of convenience, in the following we first deal with the data collection maximization problem under the assumptions that the amount of available energy at each sensor for the current mobile sink tour is given and all links are reliable. We then show how to extend the proposed solution with minor modifications to the problem without the specified assumptions.

5.1 Approximation Algorithm

Cohen et al. [9] proposed a local search algorithm for the generalized assignment problem. We show how to adopt their algorithm to the data collection maximization problem by necessary modifications, as we have already shown that the data collection maximization problem is equivalent to GAP.

The technique they adopted is based on a novel combinatorial translation of any (exact or approximation) algorithm for the knapsack problem into an approximation algorithm for GAP. Thus, any β-approximation algorithm for the knapsack problem can be transformed into a $\frac{\beta}{1+\beta}$-approximation algorithm for GAP, where β is a constant with $0 < \beta < 1$. The theoretical foundation of their technique is based a local-ratio theorem [2]. Specifically, the Cohen et al. [9] algorithm proceeds iteratively. It essentially decomposes the profit function into two profit functions: one is used for the current bin packing; and another is used for the rest of bin packing. The initial profit matrix is defined as follows:

$$D^{(0)}_{i,j} = \begin{cases} r_{i,j} \cdot \tau, & \text{if time slot } j \in A(v_i), \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Within iteration l, it packs items in $A(v_l)$ into bin B_l, using the profit function $D^{(l)}_{i,j}$, i.e., it packs time slots $j \in A(v_l)$ to sensor v_l, based on the profit entries of row l in $D^{(l)}_{i,j}$, subject to the capacity constraint $P(v_l)$ of sensor v_l.

Let $\bar{S}_l$ be the set of time slots allocated to sensor v_l by a β-approximation algorithm for the knapsack problem, clearly $\bar{S}_l \subseteq A(v_l)$. Then, the profit function $D^{(l)}_{i,j}$ is decomposed into two profit functions $D^{(l+1)}_{i,j}$ and $T^{(l+1)}_{i,j}$ as follows:

$$D^{(l+1)}_{i,j} = \begin{cases} D^{(l)}_{l,j}, & \text{if time slot } j \in \overline{S_l} \text{ or } i = l, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

and

$$T^{(l+1)}_{i,j} = D^{(l)}_{i,j} - D^{(l+1)}_{i,j}. \tag{3}$$

The decomposition of the profit function implies that $D^{(l+1)}_{i,j}$ is identical to $D^{(l)}_{i,j}$ with regard to bin B_l. In addition, if time slot $j \in \overline{S_l}$, then it is allocated in $D^{(l+1)}_{i,j}$ the same profit as that in $D^{(l)}_{i,j}$ for all bins l' if $j \in A(v_{l'})$. All other entries are zeros. The new profit function for bin B_{l+1}, $D^{(l+1)}_{i,j}$ then is $T^{(l+1)}_{i,j}$, i.e.,

$$D^{(l+1)}_{i,j} = T^{(l+1)}_{i,j}. \tag{4}$$

The procedure continues until the last bin B_n is packed. Then, an approximate solution to the data collection maximization problem finally is derived. That is, let S_l be the set of time slots allocated to sensor v_l. If $l = n$, then $S_n = \overline{S_n}$; otherwise, the set of time slots allocated to sensor v_l is $S_l = \overline{S_l} \setminus \cup^n_{j=l+1} S_j$ for all l with $1 \leq l \leq n-1$.

The offline approximation algorithm for the data collection maximization problem is thus as follows.

Initially, we sort the sensors in increasing order of indices of their starting time slots, (i.e., the index of the first time slot in $A(v_i)$ for sensor v_i). If there are multiple sensors with the same starting time slot, then sort them in increasing order of indices of their ending time slots. In case the indices of these ending time slots are also identical, the tie between the sensors will be broken arbitrarily. Without loss of generalization, assume that $v_1, v_2, \ldots v_n$ is the sorted sensor sequence starting from time slot indexed by 1, and the mobile sink starts its data collection tour from the first time slot. The detailed offline approximation algorithm `Offline_Appro` is presented in `Algorithm` 1.

Algorithm 1. `Offline_Appro`

Input: The number of time slots T, the set of sensors V, the energy budget $P(v_i)$ and the set of available time slots $A(v_i)$, the transmission rate $r_{i,j}$ and the corresponding energy consumption $P_{i,j}$ of each sensor $v_i \in V$, and the profit matrix $D_{i,j}^{(0)}$ for all i and j with $1 \leq i \leq n$ and $1 \leq j \leq T$.

Output: Allocate T time slots to the n sensors.

```
Sort all sensors in increasing order of the indices of their starting time slots, followed by their ending time slots. Let v_1, v_2, ..., v_n be the sorted sensor sequence;
Profit matrix's initialization: D_{i,j}^{(1)} <- D_{i,j}^{(0)} for all i and j with 1 <= i <= n and 1 <= j <= T;
for l <- 1 to n do
    /* Assume that A(v_l) = {l_s, ... l_e} */
    Apply a β-approximation algorithm for a single bin packing (knapsack problem) to allocate time slots in A(v_l) to sensor v_l, subject to the capacity of v_l, P(v_l), using the profit function D_{i,j}^{(l)}, i.e., the entries in row l of the matrix. Let \overline{S_l} be the result delivered by the approximation algorithm to sensor v_l, where \overline{S_l} ⊆ A(v_l);
  /* Decompose the profit function into two profit functions D_{i,j}^{(l+1)} and T_{i,j}^{(l+1)} */
D_{i,j}^{(l+1)} <- T_{i,j}^{(l+1)};
end for;
   /* construct a solution to the time slot allocation */
S_n <- \overline{S_n};
for l <- n - 1 downto 1 do
    S_l <- \overline{S_l} \ ∪_{j=l+1}^{n} S_j;
end for;
return S_l for all l with 1 <= l <= n.
```

5.2 Complexity Analysis

Theorem 2. *Given an energy renewable sensor network $G = (V \cup \{s\}, E)$ with $n = |V|$, there is an approximation algorithm for the data collection maximization problem with an approximation ratio of $\frac{1}{2+\epsilon}$. The time complexity of the proposed approximation algorithm is $O(n^2)$.*

Proof. Cohen et al. [9] have showed that algorithm `Offline_Appro` is a $\frac{\beta}{1+\beta}$-approximation algorithm, where β is the approximation ratio of an approximation algorithm for the single knapsack problem with $0 < \beta < 1$. Obviously, the approximation ratio of the approximation algorithm is $\beta = \frac{1}{1+\epsilon}$ [22], where ϵ is a constant with $0 < \epsilon < 1$, and it takes $O(|A(v_l)| \log \frac{1}{\epsilon} + \frac{1}{\epsilon^4}) = O(t_{max})$ time to find the subset $\overline{S_l}$ $(\subseteq A(v_l))$, where $t_{max} = \max \{|A(v)| \mid v \in V\}$. The updating of profit matrices $D_{i,j}^{(l)}$ and $T_{i,j}^{(l)}$ also takes time. However, it is noticed that there is no need to update all entries, we only need to update the entries in row l and the related columns $j \in \overline{S_l}$, thus, it takes $O(|A(v_l)| + \sum_{j \in \overline{S_l}} O(n)) = O(|A(v_l)| + O(n \cdot |\overline{S_l}|)) = O(nt_{max})$ time. The running time of allocating all time slots into the n sensors therefore is $\sum_{v_l \in V} O\ (t_{max} + nt_{max}) = O(nt_{max} + n^2 t_{max}) = O(n\Gamma + n^2\Gamma) = O(n^2)$ since $t_{max} \leq 2\Gamma$ and $\Gamma = \lfloor \frac{R}{r_s \cdot \tau} \rfloor$ usually is a constant in practice, where R is the maximum transmission range of sensors and r_s is the travelling speed of the mobile sink. The approximation ratio of the proposed algorithm for the data collection maximization problem thus is $\frac{\beta}{1+\beta} = \frac{1}{2+\epsilon}$. □

5.3 Harvesting Energy Estimation and Unreliable Link Reliability

The proposed approximation algorithm, `Algorithm 1`, is proposed, under the assumptions that the energy budget $P(v_i)$ of each sensor $v_i \in V$ is given and the link reliability of each link $\rho_{i,j}$ between sensor v_i and the mobile sink at each time slot j is reliable (i.e., $\rho_{i,j} = 1$) for all $v_i \in V$ and all $j \in A(v_i)$. In reality, the battery energy information $P(v_i)$ at each sensor v_i is not known, and the wireless communication between a sensor and the mobile sink is error-prone and not always noise free, interferences are not avoidable. Therefore, both harvested energy predictions and unreliable link reliability must be taken into account when dealing with the design of real protocols for energy renewable sensor networks. In this section we show how to extend the proposed algorithm for this general case.

To predict the harvested energy of each sensor node, Kansal et al. proposed the very first algorithm, referred to as the Exponentially Weighted Moving-Average (EWMA), which applies weighting factors to previously harvested sampling energy values that are constantly decreasing. At the same time, the prediction takes into account every single harvesting energy sample with different relevance [19]. A similar prediction strategy has also been adopted by Noh et al. [33]. Specifically, let $\overline{Q}(d,t)$ be the base prediction of the amount of harvested energy of sensor v_i between the $(t-1)$th tour and the tth tour on day d as follows:

$$\overline{Q}(d,t) = w \cdot \overline{Q}(d-1,t) + (1-w) \cdot Q(d-1,t), \qquad (5)$$

where $0 < w < 1$ is a weight factor and $Q(d-1,t)$ is its real amount of harvested energy between the $(t-1)$th tour and the tth tour on day $d-1$. The final prediction on the amount of harvested energy can be calculated by adjusting the base prediction with the current environmental conditions (e.g., a sunny day or a cloudy day), as follows:

$$\hat{Q}(d,t) = \overline{Q}(d,t) \cdot \frac{Q(d,t-1)}{\overline{Q}(d,t-1)}. \qquad (6)$$

We refer to this energy prediction as the Variance Exponentially Weighted Moving-Average algorithm, or algorithm `VEWMA` for short.

Assume that $RE(d,t-1)$ is the amount of residual energy of sensor v_i after the $(t-1)$th tour on day d, then the predicted amount of available energy $\hat{P}_j(v_i)$ for its tth tour is

$$\hat{P}_t(v_i) = \min\{RE(d,t-1) + \hat{Q}(d,t), B(v_i)\}, \qquad (7)$$

where $B(v_i)$ is the energy capacity of sensor v_i.

Assume that the mobile sink starts its tour t. We take the predicted energy budget $\hat{P}_t(v_i)$ and link reliability $\rho_{i,j}$ of each sensor into consideration when the mobile sink performs its next tour t, i.e., when the mobile sink performs packing time slots in $A(v_i)$ to bin v_i with the estimated energy budget constraint $\hat{P}_t(v_i)$ and the link reliability $\rho_{i,j}$ for all $j \in A(v_i)$, the β-approximation algorithm for the knapsack problem with reliable link reliability can still be applied to this general setting through a minor modification. That is, the profit brought by allocating time slot j to sensor v_i is $D^{(0)}_{i,j} = \rho_{i,j} \cdot r_{i,j} \cdot \tau$, not the original $r_{i,j} \cdot \tau$, when sensor v_i consumes the amount of energy $P_{i,j}$ to transmit data at time slot j with link reliability $\rho_{i,j}$. The rest is almost identical to the proposed algorithm, `Algorithm` 1, omitted.

6 Online Distributed Algorithm

In the previous section we provided an offline approximation algorithm with a provable approximation ratio for the data collection maximization problem. However, the solution obtained is based the assumptions that the global knowledge of the network topology and the profiles of sensors including their physical locations, energy budgets, starting and ending time slots are available. In reality, there is no way for the mobile sink to know the profile of each sensor unless it is within the transmission range of the sensor. Also, even if the mobile sink is able to collect the topological information of the entire network and the profiles of sensors at its previous tours, using the piggybacking strategy or linear regression prediction, it then performs time slot scheduling based on the collected information, the solution obtained however may not be applicable due to the fact that both the energy harvesting and the link reliability profiles of some sensors may have experienced drastic changes over the period of the mobile sink tour. In this section we will develop a fast, scalable online distributed algorithm for the problem without the mentioned assumptions. For the sake of discussion convenience, we first assume that all links are reliable, i.e., the link reliability of each link is one. We then extend the distributed solution to the unreliable link case through minor modifications.

6.1 Overview of the Distributed Algorithm

The overview of the proposed online distributed algorithm proceeds as follows. The mobile sink periodically broadcasts a 'Probe' message with a 'Registration' timer, announcing its presence once per time interval when it travels along the pre-defined path, where each *time interval* consists of $\Gamma = \lfloor \frac{R}{\tau \cdot r_s} \rfloor$ time slots. The 'Probe' message is broadcast in the beginning of each interval, which will be used to detect whether the mobile sink and the sensors are within the transmission range of each other. Each sensor receiving the 'Probe' message will send the mobile sink back an 'Ack' message which contains its current power level, the indices of its starting and ending time slots, its location coordinate, its link reliability, etc. The sensor then enters the waiting status to get the reply from the mobile sink when performing its next action. Once the 'Registration' timer expires, the mobile sink starts scheduling the Γ time slots to the registered sensors, using a time-slot scheduling algorithm $\mathcal{A}$ which will be detailed later. It finally broadcasts the time-slot allocation results to the registered sensors. Each registered sensor (in the waiting status) then sets its own scheduling, i.e., in which time slots it will transmits its data to the mobile sink.

Within the rest of the current time interval, each registered sensor transmits its data to the mobile sink at its allocated time slots. For the sake of simplicity, we here assume that the time spent by the mobile sink in probing and time slot scheduling is negligible in comparison with the time at each time slot for data transmission.

Algorithm 2. `Distributed_Algorithm (the mobile sink)`

$continue \leftarrow$ '$true$';/*the current tour finishes or not*/
$j \leftarrow 0$; /* the number of time intervals per tour */
while $continue$ **do**
  $j \leftarrow j + 1$; /* The current time interval j*/
  Mobile sink broadcasts a 'Probe' message with a 'Registration' timer to one-hop sensors;
  if the timer expires **then**
    if the mobile sink received 'Ack' messages from sensors **then**
      Call a time-slot scheduling algorithm, `Algorithm` $\mathcal{A}$, in the mobile sink to allocate the time slots in time interval t to the registered sensors, subject to the power constraint on each registered sensor;
      The mobile sink broadcasts the scheduled results to sensors in the network;
      /*Each registered sensor performs data transmissions in its allocated time-slots; */
      The mobile sink broadcasts a 'Finish' message to sensors when it finished the data collection from the last time slot in time interval j;
      /* The registered sensors update their energy profiles when they received the 'Finish' messages. That is, each registered sensor v_i updates its power: $P_j(v_i) \leftarrow P_j(v_i) - \sum_{j' \in S_i} \cdot P_{i,j'} \cdot \tau$, where S_i is the set of time slots assigned to v_i by algorithm $\mathcal{A}$ in the current time interval j and $S_i \subseteq A(v_i)$; */
    else
      $continue \leftarrow$ '$false$'; /* finish the tour */
    end if
  else
    Waiting for the replies from one-hop sensors;
  end if
end while

When the mobile sink received the data from the sensor at the last time slot in the current time interval, it sends a 'Finish' message to all the registered sensors. The registered sensors then update their own energy profiles after having received the 'Finish' message and wait for their scheduling in the next time interval. This procedure continues until there is no response from any sensor to the 'Probe' message sent by the mobile sink in some time interval, which means that the mobile sink finishes the tour already, as we assumed that the sensors are densely deployed along the pre-defined path and there is at least one sensor at each time interval. The detailed online distributed algorithm is given in `Algorithm` 2 and `Algorithm` 3.

Algorithm 3. `Distributed_Algorithm (sensor node` v_i`)`

At each time slot, sensor node $v_i \in V$ performs its data collection based on its duty-cycling

When it receives a 'Probe' message from the mobile sink, it responds by sending back of an 'Ack' message that includes its current energy $P(v_i)$ and link reliability in the last tour $\rho_{i,j}$, and waiting for the reply from the mobile sink;

When it receives the time-slot allocation information from the mobile sink, set its time-slot scheduling, and perform data transmission in its allocated time slots.

When it receives a 'Finish' message from the mobile sink, it updates its energy budget for next time interval.

6.2 GAP-Based Time Slot Scheduling

In the rest we devise a *GAP-based time-slot scheduling algorithm* as `Algorithm` $\mathcal{A}$ in algorithm 2. Recall that the starting and ending time slots of sensor $v_i \in V$ are the i_sth and the i_eth time slots, denote by $[i_s, i_e]$ the time slot interval in which sensor v_i can transmit its data to the mobile sink. Given the current time interval j, $[a_j, b_j]$ where a_j and b_j are the starting and ending time slots in the current time interval, then $|b_j - a_j| = \lfloor \frac{R}{r_s \cdot \tau} \rfloor$. If $[i_s, i_e] \cap [a_j, b_j] \neq \emptyset$, then sensor v_i can transmit its data to the mobile sink in time interval j within time slot interval $[i'_s, i'_e] = [i_s, i_e] \cap [a_j, b_j]$ with $i_s \leq i'_s$ and $i'_e \leq i_e$. Let $P_j(v_i)$ be the amount of power of sensor v_i in the beginning of time interval j, then it consumes the amount of energy $P_{i,j'} \cdot \tau$ when sensor v_i transmits its data in a time slot $j' \in [i'_s, i'_e]$. It may transmit its data within multiple time slots as long as its residual energy enables itself to do so. The mobile sink schedules the current Γ time slots to these registered sensors in the current time interval, using the offline approximation algorithm. This GAP-based algorithm is described in `Algorithm` 4.

Algorithm 4. `GAP-based_Time-slot_Scheduling` at time interval j

Let a_j and b_j be the starting and ending indices of time slots in time interval j, let TS_j be the set of sensor nodes responded to the 'Probe' request message issued by the mobile sink.

Let $v_i \in TS_j$ and i_s and i_e be the starting and ending indices of time slots of v_i in $[a_j, b_j]$. That is, let $A'(v_i) = \{i_s, i_s + 1, \ldots, i_e\} \subseteq A(v_i)$ be the subset of time slots of v_i in time interval j.

for each $v_i \in TS_j$ **do**

  Apply an β-approximation algorithm for bin packing to pack time slots in $A'(v_i)$ to sensor v_i with the bin capacity $P_j(v_i)$

  Let S_i be the set of allocated time slots to sensor v_i, i.e., $S_i \subseteq A'(v_i) \subseteq A(v_i)$.

end for

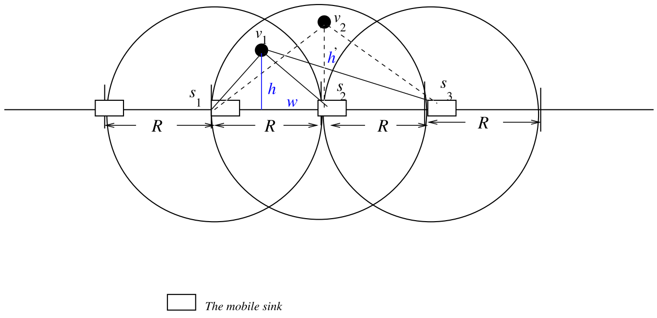


Fig. 2. A sensor v_1 (or v_2) cannot be in three consecutive time intervals.

We then have the following lemma and theorem.

Lemma 1. *Given the sensor network $G = (V \cup \{s\}, E)$, following the proposed online distributed algorithm,* `Algorithm` *2 and* `Algorithm` *3, we claim that each sensor is within at most two consecutive broadcasting regions (or two consecutive time intervals).*

Proof. We show the claim by contradiction. Considering Fig. 2, assume that a sensor v_1 is within three consecutive 'Probe' message broadcasting regions, i.e., when the mobile sink broadcasts its probing messages at s_1, s_2, and s_3 locations, sensor v_1 is able to receive the message three times. Following this assumption, we have $d(v_1, s_1) \leq R$, $d(v_1, s_2) \leq R$, and $d(v_1, s_3) \leq R$. We now show that this is impossible by the following three cases:

Case one. Sensor v_1 is in the left side of s_2, then $d(v_1, s_3) = \sqrt{h^2 + (w + R)^2} > \sqrt{R^2} = R$, which contradicts the fact that $d(v_1, s_3) \leq R$.

Case two. Sensor v_1 is in the right side of s_2, the proof is similar to Case one, omitted.

Case three. Sensor v_1 (i.e., sensor v_2) is just above s_2, then $d(v_2, s_1) = \sqrt{h'^2 + R^2} > \sqrt{R^2} = R$ and $d(v_2, s_3) = \sqrt{h'^2 + R^2} > \sqrt{R^2} = R$. This contradicts that v_2 is in the transmission ranges of s_1 and s_3. □

Theorem 3. *Given an energy renewable sensor network $G = (V \cup \{s\}, E)$ with $|V| = n$, there is an online distributed algorithm for the data collection maximization problem in G, which takes $O(n)$ time and $O(n)$ messages.*

Proof. Following Lemma 1, we notice that each sensor can receive the probing message and the finish message from the mobile sink at most twice per tour, and these messages are issued in two consecutive time intervals. Thus, the total number of probing and finish messages and the time slot allocation messages received by each sensor are four, respectively per tour of the mobile sink, while the number of acknowledgement messages by each sensor is two as well. Thus, the total number of messages transmitted per tour is $O(\sum_{v \in V} d_v) = O(n)$ as each sensor v has $O(d_v) = O(1)$ messages to be received and/or sent out. Clearly, the time for time-slot scheduling by the mobile sink in each interval j is $\sum_{l=1}^{N_j} O(t_{max} \log t_{max}) = O(N_j \cdot t_{max} \log t_{max})$ as sorting by the mobile sink for bin packing at each sensor in this interval takes $O(t_{max} \log t_{max})$ time, and the rest operations takes constant time, where N_j is the number of registered sensors in interval j and $t_{max} = \max_{v \in V} \{|A(v)|\}$. Thus, the time complexity of the online distributed algorithm is proportional to the number of time intervals per tour. As we assume that sensors are densely deployed, this implies that there is at least one sensor responded to each probing request in the beginning of each time interval, while each sensor is included in at most two consecutive time intervals by Lemma 1. Assume that there are K intervals of each tour, then $\sum_{j=1}^{K} N_j \leq 2n$. Thus, the time complexity of the online distributed algorithm is $\sum_{j=1}^{K} O(N_j \cdot t_{max} \log t_{max}) = O(n t_{max} \log t_{max}) = O(n \Gamma \log \Gamma) = O(n)$ as $t_{max} \leq 2\Gamma$ and $\Gamma = \lfloor \frac{R}{r_s \cdot \tau} \rfloor$ usually is a constant in practice, where R is the maximum transmission range of sensors and r_s is the travelling speed of the mobile sink. □

6.3 Unreliable Wireless Communication

The proposed online distributed algorithm is built upon the assumption that communications between sensors and the mobile sink are reliable. We now remove this assumption by dealing with a general case where wireless communications are not reliable, for which we will adopt the similar strategy as we did for the offline approximation algorithm. That is, within each time interval, when the mobile sink broadcasts a 'Probe' message, a responding sensor v_i receiving the 'Probe' message will respond by sending an 'Ack' message back to the mobile sink, the Ack message contains not only the current harvested energy $P(v_i)$ of v_i but also its link reliability $\rho_{i,j'}$ in the previous tour for each time slot $j' \in A(v_i)$. The mobile sink then proceeds a time-slot scheduling in this time interval, based on sensor energy budget and the estimation of link reliability. In terms of time slot allocation to a responded sensor, the energy consumption of the sensor by transmitting its data at any given time slot should incorporate its re-transmission energy consumptions (the link reliability). The rest operations are identical to the case for the perfect channel condition, omitted.

7 A Special Data Collection Maximization Problem

In this section we deal with a special case of the data collection maximization problem where each sensor $v_i \in V$ has only one fixed transmission power level with power P'_i. For this special case, we devise a fast, scalable online distributed algorithm for the problem as follows.

We reduce this special data collection maximization problem to the maximum weight matching problem in another auxiliary, bipartite graph $G = (X \cup Y, E_{XY})$, where X is the set of sensors which acknowledged the probing message by the mobile sink in the beginning of time interval j, Y is the set of Γ time slots to be allocated to the registered sensors in X. There is an edge between a sensor node v_i that corresponds to a node $x_i \in X$ and a time slot node $y_j \in Y$ if $y_j \in [i'_s, i'_e]$, i.e., y_j is a time slot in interval $[i'_s, i'_e]$. There are $m_i = |i'_s - i'_e| + 1$ edges incident to node x_i in G. The weight associated with edge $(x_i, y_j) \in E_{XY}$ is the average amount of data received by the mobile sink from sensor v_i at time slot y_j, $D^{(0)}_{i,j} = r_{i,j} \cdot \rho_{i,j} \cdot \tau$, where the average data transmission rate $r_{i,j}$ of sensor v_i at time slot y_j is determined by the distance between sensor v_i and the mobile sink at time slot y_j. Our objective thus is to maximize the data collected by the mobile sink in the current time interval through the time slot allocation. In terms of time slot allocation, we notice that each registered sensor v_i in the current time interval can make use of upto $n_i = |A(v_i)|$ time slots to transmit its data. Meanwhile, it is very likely that there are multiple sensors to compete with each other for each shared time slot to transmit their own data. The challenge thus is how to allocate these time slots to the registered sensors such that the sum of amounts of data transmitted is maximized. In the following we propose a solution to this special data collection maximization problem by reducing it to a maximum weight matching problem in another bipartite graph $G' = (\{x_i^{(k)} | x_i \in X, 1 \leq k \leq n'_i\} \cup Y, E')$, where G' is derived from the bipartite graph G as follows.

For each node $x_i \in X$ in G, there are n'_i corresponding node copies, $x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(n'_i)}$ in G', where $n'_i = \min\{\lfloor \frac{R}{r_s \cdot \tau} \rfloor, |i'_s - i'_e| + 1, \lfloor P(v_i)/(P'_i \cdot \tau) \rfloor\}$, where P'_i is the fixed transmission power of sensor v_i, and $P'_i \cdot \tau$ is the amount of energy needed by sensor v_i to transmit a message in a time slot. For each an edge $(x_i, y_j) \in E_{XY}$ in G, there are n'_i corresponding edge copies $(x_i^{(1)}, y_j), (x_i^{(2)}, y_j), \ldots, (x_i^{(n'_i)}, y_j)$ in E', and each of them has a weight $D^{(0)}_{i,j}$. Then, finding a solution of allocating the Γ time slots to the registered sensors such that the amount of data collected by the mobile sink in this time interval is maximized is equivalent to finding a maximum weight matching in G' such that the weighted sum of matched edges is maximized.

Let M be such a maximum weight matching in G'. Then, M corresponds to a time-slot allocation. That is, each edge $(x_i^{(k)}, y_j)$ in M implies that time slot y_j is allocated to sensor v_i, and sensor v_i will successfully transmit its data with the data transmission rate $r_{i,j}$ to the mobile sink. We refer to this online distributed algorithm as `Online_MaxMatch`, and have the following theorem.

Theorem 4. *Given an energy renewable unreliable sensor network $G = (V \cup \{s\}, E)$ with $|V| = n$, there is an online maximum weight matching-based distributed algorithm for a special data collection maximization problem in G where there is only one fixed transmission power for each sensor. The proposed distributed algorithm takes $O(n^{1.5})$ time and $O(n)$ messages.*

Proof. The analysis of time complexity and message complexity of the proposed online distributed algorithm are almost identical to the ones in Theorem 3. The rest will focus on the analysis of time complexity of the operations in each time interval. Let N_j be the number of registered sensors in time interval j. Then, the bipartite graph G' contains $O(N_j \cdot t_{max} + \Gamma)$ nodes and $O((N_j \cdot t_{max}) \cdot \Gamma)$ edges, while it takes $O(\sqrt{|V|} \cdot |E|)$ time to find a maximum weight matching in a bipartite graph $G = (V, E)$ [31]. Thus, it takes $O(N_j^{1.5} \cdot \Gamma^{2.5}) = O(N_j^{1.5})$ time in G' to find the maximum weight matching M, since $t_{max} \leq 2\Gamma$ and $\Gamma = \lfloor \frac{R}{r_s \cdot \tau} \rfloor$ usually is a constant in practice, where R is the maximum transmission range of sensors and r_s is the

TABLE 1
List of Experimental Setting Parameters

Parameter	Values
Number of sensors	100 – 400
Sensor transmission rates $r_{i,j}$ and energy consumption $P_{i,j}$	250 Kbps between 0 m–20 m at 170 mW 19.2 Kbps between 20 m–50 m at 220 mW 9.6 Kbps between 50 m–120 m at 300 mW 4.8 Kbps between 120 m–200 m at 330 mW
Link reliability $\rho_{i,j}$	[0, 1]
Sensor energy capacity $B(v)$	10,000 Joules
Sink travelling speed r_s	5 m/s – 30 m/s
Duration of time slot τ	1 sec.–10 sec.

travelling speed of the mobile sink. Notice that this maximum weight matching-based time-slot scheduling algorithm is performed by the mobile sink. Assuming that there are K intervals, following Lemma 1, each sensor appears at most twice in two consecutive time intervals, thus, $\sum_{j=1}^{K} N_j \leq 2n$. The total amount of time spent for finding maximum weight matchings in all intervals therefore is $\sum_{j=1}^{K} O(N_j^{1.5}) = O(n^{1.5})$. Considering the fact that N_j usually is bounded by a constant in practice, then the proposed online distributed algorithm takes only $O(n)$ time, and the message complexity is still $O(n)$. □

Notice that if the global knowledge of the entire network and the residual energy and location profiles of all sensors are given, an offline algorithm for the special data collection maximization problem based on the maximum weight matching can also be obtained, and delivers an exact solution in polynomial time. We refer to this offline algorithm as algorithm `Offline_MaxMatch`.

8 Performance Evaluation

In this section we first evaluate the accuracy of the energy prediction model. We then study the performance of the proposed algorithms through experimental simulation. We finally investigate the impact of parameters: the network size n, the mobile sink speed r_s, and the duration τ of each time slot on the performance of proposed algorithms.

8.1 Experimental Environment Setting

We consider an energy renewable sensor network consisting of 100 to 400 sensor nodes randomly deployed along two sides of a pre-defined path, and a mobile sink s travels along the path at constant speed r_s. We further assume that the length of the pre-defined path is $10{,}000\,\mathrm{m}$ and the path is a straight line, and the maximum distance between the location of any sensor and the path is $180\,\mathrm{m}$. Each sensor has an identical maximum transmission range of 200 meters and is powered by a $10\,\mathrm{mm} \times 10\,\mathrm{mm}$ square solar panel with the battery capacity of $10,000\,Joules$. The solar power harvesting profile is built upon real solar radiation measurements [29], in which the total amount of energy collected from a $37\,\mathrm{mm} \times 37\,\mathrm{mm}$ solar panel over a 48-hour period is $655.15\,\mathrm{mWh}$ in a sunny day and $313.70\,\mathrm{mWh}$ in a partly cloudy day, respectively. We here adopt the communication parameters of a real radio $CC2591$ by TI [7], where its transmission and corresponding distance parameters are listed in Table 1. In the default setting the duration of each time slot τ is 1 second. Each value in figures is the mean of the results obtained by applying each mentioned algorithm to 50 different network topologies of the same network size. Since the deviations of the 50 replication results are minor, for the sake of clarity, we do not provide error bars to indicate their standard deviations. We will adopt an existing offline algorithm `C_Schedule` [35] for a similar data gathering problem as the benchmark, which proceeds to allocate time slots iteratively, starting from time slot 1 and ending at time slot T. Within iteration j, time slot j will be allocated to the sensor with the maximum amount of its data to be transmitted.

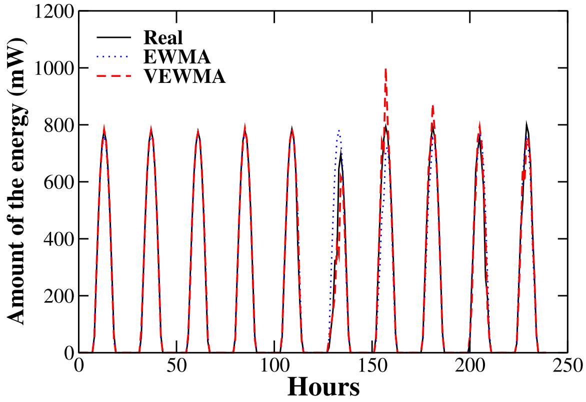


Fig. 3. The accuracy performance of prediction algorithms `VEWMA` and `EWMA` with weight $w = 0.5$.

8.2 Harvesting Energy Prediction

We first investigate the accuracy of the harvested energy prediction approach `VEWMA` in comparison with the one of a basic prediction approach `EWMA`, using the real solar data profiles obtained from The National Solar Radiation Data Base (NSRDB) in the States [32] which contains the most comprehensive collection of solar data and is freely available.

Fig. 3 shows the actual solar data measurements within 10 consecutive days under different weather conditions and the predicted values by algorithms `EWMA` and `VEWMA`, respectively, from which it can be seen that the accumulative error between the estimated ones and the real ones is given by the following equation:

$$Error = \frac{1}{M} \sum_{i=1}^{M} \left| 1 - \frac{Real}{Estimated} \right|, \tag{8}$$

where M is the number of predictions made in the past. By setting the weight w to be 0.5, both algorithms `EWMA` and `VEWMA` will deliver small accumulative errors. Specifically, the error by algorithm `VEWMA` is 9.1 percent, compared with 12.6 percent by algorithm `EWMA`. Given three different independent datasets, Fig. 4 implies that with the increase of w from 0.1 to 0.9, the errors by both algorithms `VEWMA` and `EWMA` decrease slightly. However, when the value of w is greater than 0.9, the errors by both algorithms `VEWMA` and `EWMA` increase significantly and can reach upto from 66 to 300 percent. In order to

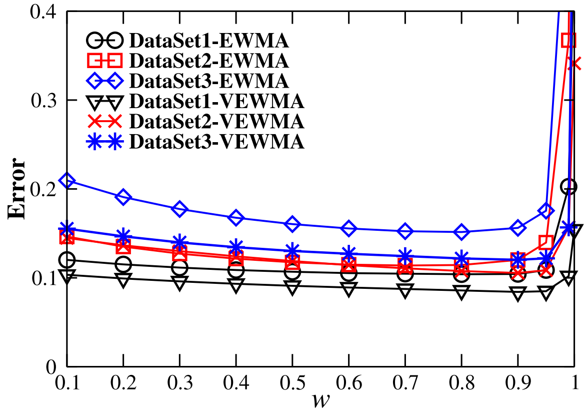


Fig. 4. Accumulative errors of prediction accuracy by algorithms VEWMA and EWMA with different weights w.

obtain better harvesting energy prediction performance, the value of w should be adjusted, following the historical harvesting energy profiles.

8.3 Performance Evaluation of Different Algorithms

We then evaluate the performance of algorithms Offline_Appro and Online_Appro by varying the network size n from 100 to 400 and setting the mobile sink speed r_s at $5\,\mathrm{m/s}$, and $10\,\mathrm{m/s}$, while the duration of time slot τ is fixed at $1s$, $4s$, and $8s$, respectively.

Fig. 5 demonstrates that algorithm Offline_Appro always outperforms algorithm Online_Appro slightly. However, they both outperform the benchmark algorithm C_Schedule significantly. For example, when $r_s = 5\,\mathrm{m/s}$ and $\tau = 1\mathrm{s}$, the network throughput of algorithm Online_Appro is no less than 93 percent of that of algorithm Offline_Appro, while their throughputs are no less than from 115 to 400 percent that of algorithm C_Schedule. The reason behind is that algorithm Online_Appro only has the local, rather global knowledge of the entire network. It can be also noticed that when the network size is fixed, the longer duration of time slot and the higher mobile sink speed will lead to a lower network throughput of each mentioned algorithm. In other words, to maximize the network throughput, a shorter duration of time slot should be chosen when the mobile sink travels at a faster speed.

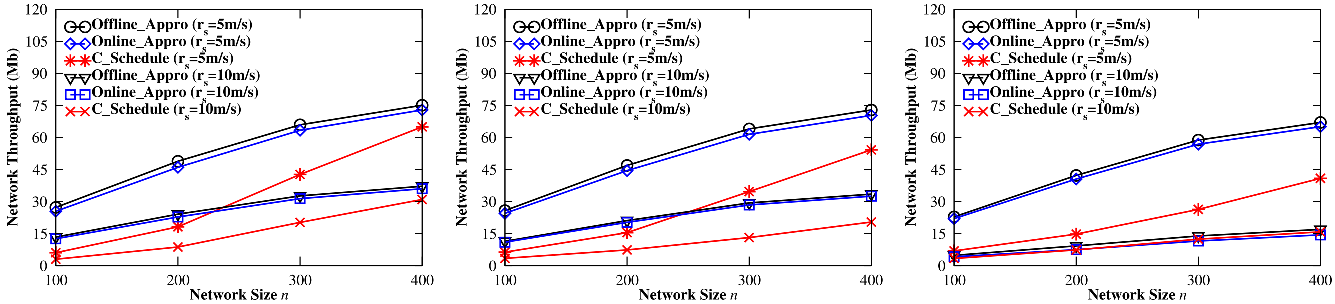
Fig. 6 shows that the network throughputs of the three mentioned algorithms Offline_Appro, Online_Appro and C_Schedule drop down significantly when varying the link reliability between 0 and 1 randomly in comparison with their counterparts in the link reliability case in Fig. 5.

8.4 Performance of Different Algorithms for the Special Data Collection Maximization Problem

When the transmission power of each sensor is fixed at $300\,\mathrm{mW}$, we now investigate both the performance of algorithms Offline_MaxMatch, Online_MaxMatch, Offline_Appro, and Online_Appro against algorithm C_Schedule and the impacts of the network size n and the mobile sink speed r_s on the performance, by varying n from 100 to 400 and setting r_s at $5\,\mathrm{m/s}$, $10\,\mathrm{m/s}$, and $30\,\mathrm{m/s}$ while the duration of time slot τ is fixed at $1s$.

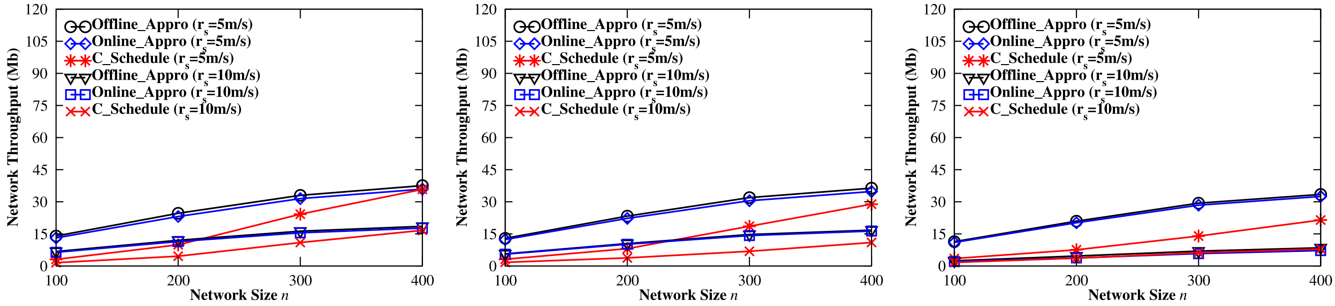
When the mobile sink speed is fixed at $5m/s$, Fig. 7a clearly indicates that algorithm Offline_MaxMatch outperforms the other four algorithms, and algorithm C_Schedule is the worst one among them. Moreover, it is observed that algorithm Online_MaxMatch is inferior to algorithm Offline_MaxMatch, as algorithm Online_MaxMatch only has the local knowledge of the network. However, the performance gap between them is only marginal. It is also noticed that algorithm Online_MaxMatch outperforms the other three algorithms, and the performance gaps among them increase with the growth of network size. Specifically, when $n = 100$, the performance of algorithms Online_MaxMatch, Offline_Appro, and Online_Appro are almost the same. When $n = 400$, the performance of algorithm Online_MaxMatch is 18 and 23 percent better than that of algorithms Offline_Appro and Online_Appro.

When the mobile sink speed is fixed at 10 or $30\,\mathrm{m/s}$ respectively, Figs. 7b and 7c exhibit the similar performance behaviors as Fig. 7a, omitted. In summary, Fig. 7 implies that when the network size is fixed, the network throughput delivered by each mentioned algorithm decreases, with the increase of the mobile sink speed. Specifically, the network throughput delivered by algorithm Offline_MaxMatch when $r_s = 5\,\mathrm{m/s}$ is at least 105 and 617 percent higher than that by itself when $r_s = 10$ and $30\,\mathrm{m/s}$, respectively. This is because when the mobile sink travels at a faster speed, the duration of the mobile sink travels the entire path will be shortened, while the data transmission rate is still keeping unchanged, thus, the amount of uploading data from sensors will be reduced. Although a faster speed leads to a shorter delay on data delivery, it will result in a less amount of data collected per tour.



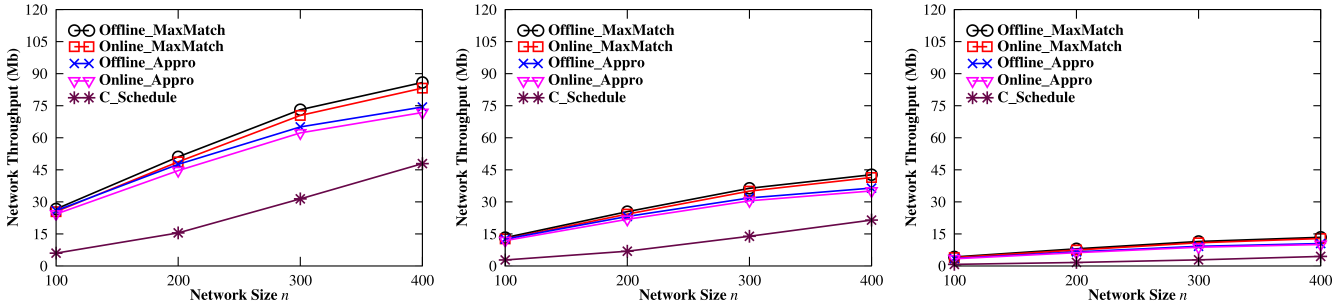
(a) $\tau = 1s$ (b) $\tau = 4s$ (c) $\tau = 8s$

Fig. 5. Network throughput by algorithms Offline_Appro, Online_Appro, and C_Schedule by varying the sink speed r_s and the network size n when all links are reliable.



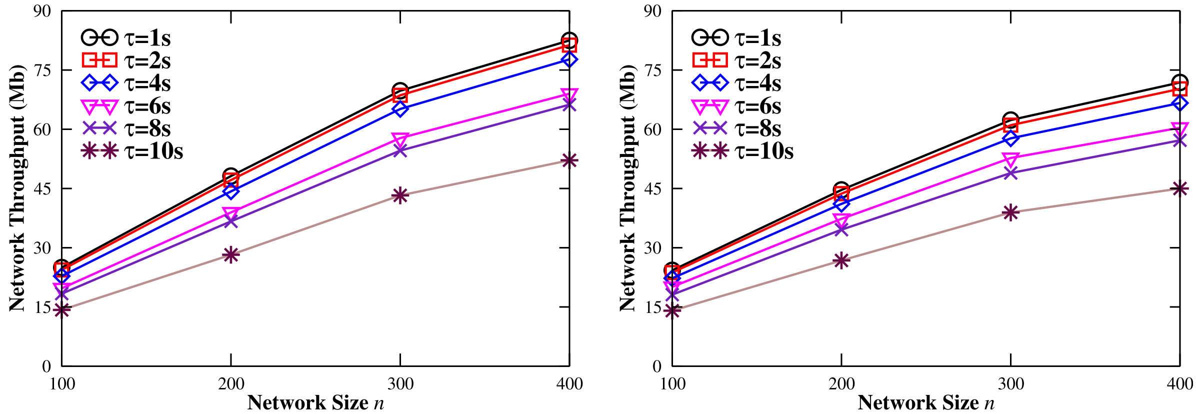
(a) $\tau = 1s$ (b) $\tau = 4s$ (c) $\tau = 8s$

Fig. 6. Network throughput delivered by algorithms `Offline_Appro`, `Online_Appro`, and `C_Schedule` through varying the sink speed r_s and the network size n when the link reliability is between 0 and 1.



(a) $r_s = 5m/s$ (b) $r_s = 10m/s$ (c) $r_s = 30m/s$

Fig. 7. Network throughput delivered by different algorithms for a special case through varying the mobile sink speed r_s and the network size n when all links are reliable.



(a) Algorithm `Online_MaxMatch` (b) Algorithm `Online_Appro`

Fig. 8. Impact of network size n and the time slot duration τ on the network throughput delivered by algorithms `Online_MaxMatch` and `Online_Appro` when all links are reliable.

We finally study the impact of the duration of time slot τ and the network size n on the performance of algorithms `Online_MaxMatch` and `Online_Appro`, by varying n from 100 to 400 and setting τ as 1s, 2s, 4s, 6s, 8s, and 10s respectively, while keeping the mobile sink speed r_s at 5m/s.

Figs. 8a and 8b illustrate that for both algorithms `Online_MaxMatch` and `Online_Appro`, the network throughput decreases with the increase of the duration of each time slot. Their performance gap becomes larger and larger, with the growth of the network size. Specifically, in Fig. 8a, the network throughput of algorithm `Online_MaxMatch` with $\tau = 1s$ is at least 3, 9, 21, 28, and 61 percent higher than that by itself when $\tau = 2s$, 4s, 6s, 8s, and 10s, respectively. In Fig. 8b, the network throughput of algorithm `Online_Appro` with $\tau = 1s$ is at least 2, 7, 18, 24 and 56 percent higher than that by itself when $\tau = 2s$, 4s, 6s, 8s, and 10s, respectively. The reason behind is that with a shorter time slot, the registered sensors can utilize their energy more efficiently.

9 Conclusions

In this paper we studied data collection in an energy renewable sensor network using a mobile sink that travels along a pre-defined path, by adopting multi-rate data transmission mechanisms and time-slot scheduling. We first formulated a novel data collection maximization problem and showed its NP-hardness. We then provided an offline approximation algorithm with a provable approximation ratio, by exploiting the combinatorial property of the problem, assuming that the global knowledge of the network is available. We also proposed a fast, scalable online distributed algorithm for realistic sensor networks without the global knowledge assumption. In addition, for a special case of the data collection maximization problem where each sensor has only one fixed transmission power, we proposed an exact solution to the problem. Finally, we conducted experiments by simulations to evaluate the performance of the proposed algorithms. Experimental results demonstrate that the proposed algorithms are efficient and scalable, and the solutions delivered are fractional of the optimum.

References

[1] Y. Alayev, F. Chen, Y. Hou, M. P. Johnson, A. Bar-Noy, T. La Porta, and K. K. Leung, "Throughput maximization in mobile WSN scheduling with power control and rate selection," in *Proc. IEEE 8th Int. Conf. Distrib. Comput. Sensor Syst.*, 2012, pp. 33–40.

[2] R. Bar-Yehuda and S. Even, "A local-ratio theorem for approximating the weighted vertex cover problem," *Annu. Discrete Math.*, vol. 25, pp. 27–45, 1985.

[3] S. Basagni, L. Bölöni, P. Gjanci, C. Petrioli, C. A. Phillips, and D. Turgut, "Maximizing the value of sensed information in underwater wireless sensor networks via an autonomous underwater vehicle," in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 988–996.

[4] S. Basagni, A. Carosi, C. Petrioli, and C. A. Phillips, "Coordinated and controlled mobility of multiple sinks for maximizing the lifetime of wireless sensor networks," *Wireless Netw.*, vol. 13, pp. 759–778, 2011.

[5] L. Bölöni, D. Turgut, S. Basagni, and C. Petrioli, "Scheduling data transmissions of underwater sensor nodes for maximizing value of information," in *Proc. IEEE Global Telecommun. Conf.*, 2013, pp. 460–465.

[6] CC2500 RF Transceiver [Online]. Available: http://www.ti.com/products/cc2500, 2014.

[7] CC2591 RF front end [Online]. Available: http://www.ti.com/products/cc2591, 2014.

[8] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Communication power optimization in a sensor network with a path-constrained mobile observer," *ACM Trans. Sensor Netw.*, vol. 2, pp. 297–324, 2006.

[9] R. Cohen, L. Katzir, and D. Raz, "An efficient approximation for the generalized assignment problem," *Inf. Process. Lett.*, vol. 100, pp. 162–166, 2006.

[10] P. Dutta, J. Hui, J. Jeong, S. Kim, C. Sharp, J. Taneja, G. Tolle, K. Witehouse, and D. Culler, "Trio: Enabling sustainable and scalable outdoor wireless sensor network deployments," in *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, 2006, pp. 407–415.

[11] M. DiFrancesco, S. K. Das, and G. Anastasi,"Data collection in wireless sensor networks with mobile elements: A survey," *ACM Trans. Sensor Netw.*, vol. 8, no. 1, article 7, 2011.

[12] M. L. Fisher, R. Jaikumar, and L.-N. Wassenhove, "A multiplier adjustment method for the generalized assignment problem," *Manage. Sci.*, vol. 32, pp. 1095–1103, 1986.

[13] K.-W. Fan, Z. Zheng, and P. Sinha, "Steady and fair rate allocation for rechargeable sensors in perpetual sensor networks," in *Proc. 6th ACM Conf. Embedded Netw. Sensor Syst.*, 2008, pp. 239–252.

[14] L. Fleischer, M. X. Goemans, V. S. Mirrokni, and M. Sviridenko, "Tight approximation algorithms for maximum general assignment problems," in *Proc. 17th Annu. ACM-SIAM Symp. Discrete Algorithm*, 2006, pp. 611–620.

[15] H. N. Gabow, "Data structures for weighted matching and nearest common ancestors with linking," in *Proc. 17th Annu. ACM-SIAM Symp. Discrete Algorithm*, 1990, pp. 434–443.

[16] S. Gao, H. Zhang, and S. K. Das, "Efficient data collection in wireless sensor networks with path-constrained mobile sinks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 4, pp. 592–608, Apr. 2011.

[17] I. Jawhar and N. Mohamed, "A hierarchical and topological classification of linear sensor networks," in *Proc. Wireless Telecommun. Symp.*, 2009, pp. 1–8.

[18] X. Jiang, J. Polastre, and D. Culler, "Perpetual environmentally powered sensor networks," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.*, 2005, pp. 463–468.

[19] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 6, no. 4, article 32, 2007.

[20] A. Kansal, A. A. Somasundara, D. D. Jea, M. B. Srivastava, and D. Estrin, "Intelligent fluid infrastructure for embedded networks," in *Proc. 2nd Int. Conf. Mobile Syst., Appl. Serv.*, 2004, pp. 111–124.

[21] P. S. Khairnar and N. B. Mehta, "Power and discrete rate adaptation for energy harvesting wireless nodes," in *Proc. Int. Conf. Commun.*, 2011, pp. 1–5.

[22] E. Lawler, "Fast approximation algorithms for knapsack problems," *Math. Oper. Res.*, vol. 4, pp. 339–356, 1979.

[23] W. Liang and J. Luo, "Network lifetime maximization in sensor networks with multiple mobile sinks," in *Proc. IEEE 36th Conf. Local Comput. Netw.*, 2011, pp. 350–357.

[24] W. Liang, X. Ren, X. Jia, and X. Xu, "Monitoring quality maximization through fair rate allocation in harvesting sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 9, pp. 1827–1840, Sep. 2013.

[25] W. Liang, P. Schweitzer, and Z. Xu, "Approximation algorithms for capacitated minimum forest problem in wireless sensor networks with a mobile sink," *IEEE Trans. Comput.*, vol. 62, no. 10, pp. 1932–1944, Oct. 2013.

[26] K. Lin, J. Yu, J, Hsu, S. Zahedi, D. Lee, J. Friedman, A. Kasal, V. Raghunathan, and M. Srivastava,"Heliomote: Enabling long-lived sensor networks through solar energy harvesting," in *Proc. 3rd Int. Conf. Embedded Netw. Sensor Syst.*, 2005, p. 309.

[27] L. Lin, N. B. Shroff, and R. Srikant,"Asymptotically optimal energy-aware routing for multihop wireless networks with renewable energy sources," *IEEE/ACM Trans. Netw.*, vol. 15, no. 5, pp. 1021–1034, Oct.2007.

[28] R.-S. Liu, K.-W. Fan, Z. Zheng, and P. Sinha, "Perpetual and fair data collection for environmental energy harvesting sensor networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 4, pp. 947–960, Aug. 2011.

[29] R.-S. Liu, P. Sinha, and C. E. Koksal, "Joint energy management and resource allocation in rechargeable sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, 2010, pp. 1–9.

[30] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon, "Health monitoring of civil infrastructures using wireless sensor networks," in *Proc. 6th Int. Symp. Inf. Process. Sensor Netw.*, 2007, pp. 254–263.

[31] S. Micali and V. V. Vazirani, "An $O(\sqrt{|V|}|E|)$ algorithm for finding maximum matching in general graphs," in *Proc. 21st Annu. Symp. Found. Comput. Sci.*, 1980, pp. 17–27.

[32] The National Solar Radiation Data Base [Online]. Available: http://www.rredcnrel.gov/solar/old-data/nsrdb

[33] D. K. Noh, L. Wang, Y. Yang, H. K. Le, and T. Abdelzaher, "Minimum variance energy allocation for a solar-powered sensor system," in *Proc. IEEE 8th Int. Conf. Distrib. Comput. Sensor Syst.*, 2009, pp. 44–57.

[34] X. Ren and W. Liang, "Delay-tolerant data gathering in energy harvesting sensor networks with a mobile sink," in *Proc. IEEE Global Telecommun. Conf.*, 2012, pp. 93–99.

[35] X. Ren and W. Liang, "The use of a mobile sink for quality data collection in energy harvesting sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2013, pp. 1145–1150.

[36] X. Ren, W. Liang, and W. Xu, "Use of a mobile sink for maximizing data collection in energy harvesting sensor networks," in *Proc. 42nd Int. Conf. Parallel Process.*, 2013, pp. 439–448.

[37] X. Ren and W. Liang, "Exploiting mobility for quality-maximized data collection in energy harvesting sensor networks," to appear in *Proc. 25th Annu. Int. Symp. Personal, Indoor, Mobile Radio Commun.*, 2014.

[38] A. A. Somasundara, A. Kansal, D. D. Jea, D. Estrin, and M. B. Srivastava,"Controllably mobile infrastructure for low energy embedded networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 8, pp. 958–973, Aug. 2006.

[39] D. Turgut and L. Bölöni, "Heuristic approaches for transmission scheduling in sensor networks with multiple mobile sinks," *The Comput. J.*, vol. 54, pp. 332–344, 2011.

[40] D. Turgut and L. Bölöni, "IVE: Improving the value of information in energy-constrained intruder tracking sensor networks," in *Proc. IEEE Int. Conf. Commun.*, 2013, pp. 6360–6364.

[41] Q. Wang, D. O. Wu, and P. Fan,"Delay-constrained optimal link scheduling in wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4564–4577, Nov. 2010.

[42] X. Xu and W. Liang, "Monitoring quality optimization in wireless sensor networks with a mobile sink," in *Proc. 14th ACM Int. Conf. Model., Anal. Simul. Wireless Mobile Syst.*, 2011, pp. 77–84.

[43] W. Xu, W. Liang, and X. Lin, "Approximation algorithms for min-max cycle cover problems,"*IEEE Trans. Comput.*, to appear, DOI.10.1109/TC.2013.2295609.

[44] Z. Xu, W. Liang, and Y. Xu, "Network lifetime maximization in delay-tolerant sensor networks with a mobile sink," in *Proc. IEEE 8th Int. Conf. Distrib. Comput. Sensor Syst.*, 2012, pp. 9–16.

[45] B. Zhang, R. Simon, and H. Aydin, "Maximum utility rate allocation for energy harvesting wireless sensor networks," in *Proc. 14th ACM Int. Conf. Model., Anal. Simul. Wireless Mobile Syst.*, 2011, pp. 7–16.

Xiaojiang Ren received the BE and ME degrees from the Huazhong University of Science and Technology, China, in 2004 and 2007, respectively, and is currently working toward the PhD degree in the Research School of Computer Science at the Australian National University. His research interests include wireless sensor networks, routing protocol design for wireless networks, and optimization problems. He is a student member of the IEEE.

Weifa Liang (M'99-SM'01) received the BSc degree from Wuhan University, China, in 1984, the ME degree from the University of Science and Technology of China in 1989, and the PhD degree from the Australian National University in 1998, all in computer science. He is currently an associate professor in the Research School of Computer Science at the Australian National University. His research interests include design and analysis of energy-efficient routing protocols for wireless ad hoc and sensor networks, cloud computing, graph databases, design and analysis of parallel and distributed algorithms, approximation algorithms, combinatorial optimization, and graph theory. He is a senior member of the IEEE.

Wenzheng Xu received the BSc and ME degrees in computer science from Sun Yat-Sen University, Guangzhou, China, in 2008 and 2010, respectively. He is currently working toward the PhD degree at Sun Yat-Sen University and a visiting student at the Australian National University. His research interests include routing algorithms and protocols design for wireless ad hoc and sensor networks, approximation algorithms, combinatorial optimization, and graph theory.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.