Use of a Mobile Sink for Maximizing Data Collection in Energy Harvesting Sensor Networks

Xiaojiang Ren[†] Weifa Liang[†] Wenzheng Xu[†¶]
[†]Research School of Computer Science, Australian National University, Canberra, ACT 0200, Australia
[¶]School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, 510006, China
Email: {richard.rxj, wenzheng.xu}@anu.edu.au, wliang@cs.anu.edu.au

***Abstract*—In this paper we study data collection in an energy harvesting sensor network for traffic monitoring and surveillance purpose on busy highways, where sensors are densely deployed along a pre-defined path and a mobile sink travels along the path to collect data from one-hop sensors periodically. As the sensors are powered by renewable energy sources, the time-varying characteristics of energy harvesting poses great challenges on the design of efficient routing protocols for data collection in such energy harvesting sensor networks. In this paper we first formulate a novel data collection maximization problem that deals with multi-rate transmission mechanism and transmission time slot scheduling among the sensors. We then show the NP-hardness of the problem and devise an offline algorithm with a provable approximation ratio for the problem by exploiting the combinatorial property of the problem, assuming that the global knowledge of the network topology and the profile of each sensor are given. We also develop a fast, scalable online distributed solution for the problem without the global knowledge assumption, which is more suitable for real distributive sensor networks. In addition, we consider a special case of the problem for which a optimal polynomial solution is given. We finally conduct extensive experiments by simulations to evaluate the performance of the proposed algorithms. Experimental results demonstrate that the proposed algorithms are very efficient, and the solutions are fractional of the optimum.**

I. Introduction

Wireless sensor network has emerged as a key technology for various applications such as environmental sensing, structural health monitoring, and area surveillance. However, the limited lifetime of battery-powered sensors has hampered the large-scale deployment of such networks. A viable solution against the limited energy supplies is to enable sensor nodes to harvest ambient energy from their surroundings. In addition to being environmentally friendly, harvesting energy could also enable sensor nodes to function indefinitely, allowing the network to operate perpetually and eliminating the cost for battery. However, the time-varying characteristics of energy harvesting sources poses a great challenge in the design of routing protocols for such networks due to dynamic energy replenishment. That is, algorithms that rely on pre-determined activations to sense, transmit, or receive cannot be used.

A. Related work

Sink mobility in conventional sensor networks has been extensively studied [2], [4], [8], [15], [18], [25], [26], and has been shown to improve various network performance including reducing the energy consumption of sensors, balancing the workload among sensors, thereby prolonging the network lifetime. In general, existing research can be classified into three categories in terms of sink mobility: sinks with random mobility which are often mounted on some humans or animals which move randomly in the monitored area to collect data from sensors [10]; sinks with controlled mobility which actively control their trajectories [7], [15], [9]; sinks with deterministic mobility (or *path-constrained mobile sinks*) which move along a pre-defined path [2], [8], [12], [23], [25]. Most existing studies focused on minimizing the energy consumption so as to prolong the network lifetime since sensors are powered by energy-limited batteries. However, network lifetime maximization is no longer a main issue in energy harvesting sensor networks as the sensors can be continuously recharged by renewable energies.

In terms of data collection with a path-constrained mobile sink, the closely related work in conventional sensor network is briefly described as follows. Chakrabarti *et al.* [2] 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. Kansal *et al.* [12], [23] 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 hop routing path. They proposed a speed control algorithm to improve the amount of data collected. Assuming that the mobile sink moves at a constant speed, Gao *et al.* [8] 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 relay. They formulated the subsink choice problem as a problem of minimizing the number of hops from each sensor to its subsink and provided a heuristic algorithm. They also studied the time allocation problem for subsinks by dividing the communication time between the mobile sink and all subsinks into several time intervals and proposed some practical time allocation methods. Liang *et al.* [18] recently considered another data collection problem by assuming the subsinks (they termed as gateways) are given, they devised an approximation algorithm for finding a forest consisting of routing trees rooted at gateways

0190-3918/13 $26.00 © 2013 IEEE
DOI 10.1109/ICPP.2013.53

and spanning all sensors. In contrast, very little attention has been paid to data collection in energy harvesting sensor networks with mobile sinks, and most existing solutions in such networks assumed that the collected data is routed to a fixed sink through multi-hop relays [14], [16], [27]. For example, Liu *et al.* [14], [16] formulated the problem as a lexicographic maximin rate allocation problem, and provided a centralized algorithm for the problem by solving an integer linear program. Zhang *et al.* [27] 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 tree networks. However, the fixed sink-based data collection paradigm may be applicable to small to mediate size networks, but it is not suitable for large-scale networks due to limited bandwidth, etc. Ren and Liang [21] recently considered the use of a mobile sink with controlled mobility for data collection by assuming that the mobile sink sojourns at some locations and only collects sensing data generated within that sojourn period from one-hop sensors. Orthogonal to these existing works, in this paper we consider data collection in an energy harvesting sensor network with a path-constrained mobile sink, where the sensor network is deployed along a highway for traffic-surveillance and a mobile vehicle at a constant speed is employed to patrol the highway for collecting data from its one-hop sensors. We formulate the problem as a data collection maximization problem by incorporating multi-rate wireless communication mechanism between the sensors and the mobile sink. The key technique to solve the problem of concern is a reduction which reduces the problem to the generalized assignment problem (GAP), and approximation techniques for the latter [3], [6] in turn will lead to an approximate solution to the former. The only similar work is [22] conducted by Ren and Liang. However, they considered a data quality problem and only developed greedy algorithms without performance guarantee, assuming that transmission power of sensors is fixed.

B. Contributions

Our major contributions in this paper are as follows. We consider data collection in an energy harvesting sensor network using a path-constrained mobile sink. We first formulate a novel data collection maximization problem by incorporating the multi-rate transmission mechanism and the transmission time slot scheduling among the sensors, and show the NP-hardness of the problem through a reduction from a NP-Complete problem - the generalized assignment problem. 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 global knowledge of the network and sensor profiles (e.g. their locations and energies) is available. We also develop a fast, scalable online distributed solution that is more suitable for real distributive sensor networks, by removing the global knowledge and sensor profile assumptions. We finally conduct extensive experiments by simulations to evaluate the performance of the proposed algorithms. Experimental results demonstrate that the proposed algorithms are very efficient.

To the best of our knowledge, unlike most existing solutions to this type of optimization problem by formulating and solving an integer linear programming (ILP), the proposed algorithm exploits the combinatorial property of the problem and provides the first approximate solution to the data collection maximization problem in energy harvesting sensor networks. To respond to time varying nature of renewable energy harvesting, traditional ILP methods take too much time and suffer poor scalability, and the worst of all, the solution delivered may be no longer applicable due to the quick changes of energy profiles at sensors. Thus, to develop a fast, scalable solution to the problem is desperately needed, to this end, an online distributed algorithm for the problem is devised, which can be applicable to the realistic energy harvesting sensor networks. For a special case where each sensor has only one fixed transmission power, we provide a polynomial solution to the data collection maximization problem.

C. Paper organization

The remainder of the paper is organized as follows. Section II introduces the system model, notions, problem definition. Section III shows the NP-completeness of the problem. Section IV is devoted to present an offline algorithm with a provable approximation ratio for the problem. Sections V and VI develop two online distributed algorithms for the problem. Section VII evaluates the performance of the proposed algorithms through experimental simulations, and Section VIII concludes the paper.

II. Preliminaries

A. System model

We consider an energy harvesting sensor network $G = (V \cup \{s\}, E)$ where V is a set of n homogeneous stationary sensors that are densely deployed along a pre-defined path, and s is a mobile sink traveling 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 moving 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 will be, resulting in a shorter delay on data delivery from its generation to its collection by the mobile sink. For the sake of discussion, we also assume the pre-defined path is a straight line in the rest of paper, which can be easily extended to real scenarios.

We here adopt a discrete-time system where the duration per tour is slotted into equal time slots with each lasting

τ time units. [17]. Given the mobile sink speed r_s, the number of time slots per tour can be determined, which is $T = \lceil \frac{L}{r_s \cdot \tau} \rceil$, and they are indexed by $1, 2, \cdots, T$, where L is the length of the pre-defined path. 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. Note that some sensors share some time slots at which they can transfer their data to the mobile sink. However, the mobile sink at any given time slot can receive the data from one sensor only otherwise a collision occurs [24]. We thus 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.

B. Energy model

As sensors in this network are powered by renewable energy, the amount of energy harvested by a sensor varies with time in a non-deterministic manner. 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 $A(v)$. However, it may not have enough power 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 based on the source type and harvesting history [14]. Denote by $B(v)$ the energy storage capacity and $P_j(v)$ the amount of energy stored at each node $v \in V$ in the beginning of tour j, $P_j(v)$ can be expressed as $\min\{P_{j-1}(v) + Q_{j-1}(v) - O_{j-1}(v), B(v)\}$, where $Q_{j-1}(v)$ and $O_{j-1}(v)$ are amounts of energy harvested and consumed at tour $j-1$, and $0 \leq P_j(v) \leq B(v)$. Furthermore, to support long-period, continuous monitoring service, we assume that sensors should not consume more energy than they can collect in order to achieve 'perpetual' operations [11]. Hence, we use $P_j(v)$ as the energy budget of sensor v for tour j. Without loss of generality, we define $P(v)$ as the energy budget of sensor v per tour.

C. Multi-rate communication

It is known that wireless signals suffer from path loss, fading, shadowing, interference and other impairments, and the communication performance is determined by the received Signal to Noise Ratio (SNR). Hence reliability and efficiency are often at odds with each other. Reliability can be improved by transmitting packets at the maximum transmission power. However, this introduces unnecessarily high energy consumption. Motivated by the fact that popularly used radio hardware such as CC2420 has multiple output power settings and offers a register to dynamically control the transmission power level during runtime, a multi-rate communication between v_i and the mobile sink s is adopted [19]. That is, the average transmission rates of v_i at two different time slots j and k, $r_{i,j} \propto \frac{P_{v_i}}{d^{\alpha}_{i,j}}$ and $r_{i,k} \propto \frac{P_{v_i}}{d^{\alpha}_{i,k}}$ are determined by their distances $d_{i,j}$ and $d_{i,k}$, where P_{v_i} is the transmission power of sensor v_i at that moment and $\alpha \geq 2$ is the path loss rate. We thus assume that $r_{i,j}$ and $r_{i,k}$ are given in the rest of discussions.

D. Problem definition

Given an energy harvesting 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 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 different data transmission rates of each sensor.

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 data collected from the entire network. However, since the energy replenishment rate of each sensor is much slower than its energy consumption rate, due to its energy budget, the sensor may only make use of some of the available time slots to transmit its data. What followed is which time slots should be chosen, since the sensor at different time slots will have different data transmission rates. 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, to allocate each shared time slot to which one among 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 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 does transmit its data at time slot $j \in A(v_i)$. We assume that there are a given number of different transmission rates for each sensor v_i, $r_{i,1}, r_{i,2}, \ldots r_{i,k_i}$. To ensure that the transmitted data can be received by the receiver successfully, we further assume that a different transmission rate $r_{i,j}$ consumes a different amount of power $P_{i,j}$ of sensor v_i, $1 \leq j \leq k_i$. Usually k_i is a fixed integer. For the sake of discussion convenience, in the rest of the paper we assume that $k_i = |A(v_i)|$. The data collection maximization problem in G thus is to allocate the time slots to the sensors such that $\sum_{v_i \in V} \sum_{j \in A(v_i)} (x_{i,j} \cdot r_{i,j} \cdot \tau)$ is maximized, subject to

$$x_{i,j} \in \{0, 1\}, \ \forall v_i \in V, j \in A(v_i) \quad (1)$$

$$x_{i,j} = 0, \ \forall v_i \in V, j \notin A(v_i) \quad (2)$$

$$\sum_{i=1}^{n} x_{i,j} \leq 1, \ \forall v_i \in V, \quad 1 \leq j \leq T \tag{3}$$

$$\sum_{j \in A(v_i)} P_{i,j} \cdot \tau \cdot x_{i,j} \leq P(v_i), \ \forall v_i \in V \tag{4}$$

where $P(v_i)$ is the energy budget of sensor v_i. $x_{i,j}$ is a boolean value: $x_{i,j} = 1$ if time slot $j \in A(v_i)$ is allocated to sensor v_i, otherwise $x_{i,j} = 0$. Constraints (1) and (2) ensure that at any given time slot, a sensor can transmit its data to the mobile sink only when the sink is within its transmission range. Constraint (3) enforces that at most one sensor can transfer its data to the mobile sink if multiple sensors share the time slot. Constraint (4) ensures that the energy consumption of each sensor per time-slot cannot exceed its energy budget at that moment.

III. NP-HARDNESS

Theorem 1: The data collection maximization problem in an energy harvesting 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 $A = \{a_1, a_2, \cdots a_m\}$ be a set of m items and $B = \{B_1, B_2, \cdots B_n\}$ a set of bins, where each B_i has a capacity b_i for all i with $1 \leq i \leq n$. 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 , $1 \leq i \leq n$.

We now show that a special case of the data collection maximization problem is equivalent to the defined GAP problem. The special case of 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. That is, a sensor can utilize each time slot in the pre-defined path to transmit its data. We then proceed the following reduction.

Each item in A corresponds a time slot, thus the set of time slots corresponds to the set of items A. 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 $P(v_i)$ of sensor v_i to perform its data transmission for a certain number of time slots, 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. This implies that at different time slots, the data transmission rates of sensor v_i are different, 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 GAP. Hence, the data collection maximization problem is NP-hard. ■

IV. AN OFFLINE APPROXIMATION ALGORITHM

Since the data collection maximization problem is NP-hard, in this section we instead devise an approximation algorithm with a provable approximation ratio for it, 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).

A. Approximation algorithm

Cohen et al. [3] proposed a local search algorithm for the generalized assignment problem (GAP). We adopt their algorithm for the data collection maximization problem, 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 $\frac{1}{1+\beta}$-approximation algorithm for GAP. The theoretical foundation of their technique is based a local-ratio theorem [1]. Specifically, the Cohen et al. [3] 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.

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

Within iteration l with $1 \leq l \leq n$, 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 $\overline{S_l}$ be the set of time slots allocated to sensor v_l by a β-approximation algorithm for the knapsack problem, clearly $\overline{S_l} \subseteq A(v_l)$. Then, the profit function $D^{(l)}_{i,j}$ is decomposed into two profit functions $D^{(l+1)}_{i,j}$ and $T^{(l+1)}_{i,j}$, where

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

$$T^{(l+1)}_{i,j} = D^{(l)}_{i,j} - D^{(l+1)}_{i,j}. \tag{5}$$

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.,

$$D^{(l+1)}_{i,j} = T^{(l+1)}_{i,j}. \tag{6}$$

The procedure continues until the last bin B_n is packed.

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 $S_l = \overline{S_l} \setminus \cup_{j=l+1}^{n} S_j$.

Initially we sort the sensors in increasing order of indices of their starting time slots. If there are multiple sensors with the same starting time slot, then sort them in increasing order of the 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. For the sake of completeness, we present the offline centralized algorithm for the data collection maximization problem `Offline_Appro` as **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 by the indices of their starting time slots, followed by their ending time slots. Let $v_1, v_2, \ldots, v_n$ be the sorted sensor sequence;
Profit matrix's Initialization: $D_{i,j}^{(1)} \leftarrow D_{i,j}^{(0)}$ for all i and j with $1 \leq i \leq n$ and $1 \leq j \leq T$;
for $l \leftarrow 1$ to n **do**
  /* Assume $A(v_l) = \{l_s, \cdots l_e\}$ */
  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 energy budget 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 solution delivered by the approximation algorithm, where $\overline{S_l} \subseteq A(v_l)$;
  /*decompose the profit function into two profit functions $D_{i,j}^{(l+1)}$ and $T_{i,j}^{(l+1)}$ */
  $D_{i,j}^{(l+1)} \leftarrow T_{i,j}^{(l+1)}$;
end for;
$S_n \leftarrow \overline{S_n}$;
for $l \leftarrow n-1$ downto 1 **do**
  $S_l \leftarrow \overline{S_l} \setminus \cup_{j=l+1}^{n} S_j$;
end for;
return S_l for all l with $1 \leq l \leq n$.

B. Complexity analysis

Theorem 2: Given an energy harvesting sensor network $G(V \cup \{s\}, E)$, there is an approximation algorithm for the data collection maximization problem with an approximation ratio of $\frac{1}{2+\epsilon}$, where ϵ is a constant with $0 < \epsilon < 1$. The time complexity of the proposed approximation algorithm is $O(n^2)$.

Proof: Cohen *et al.* [3] have showed that algorithm `Offline_Appro` is a $\frac{1}{1+\beta}$-approximation algorithm, where β is the approximation ratio of an approximation algorithm for the single knapsack problem. Obviously, the approximation ratio of the approximation algorithm for the single knapsack problem is $\beta = 1 + \epsilon$ [13], 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 ϵ is a constant with $0 < \epsilon < 1$ and $t_{max} = \max\{|A(v)| \mid v \in V\}$. 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. Thus, the running time of allocating all time slots into the n sensors 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 $\frac{1}{1+\beta} = \frac{1}{2+\epsilon}$. ■

V. Online Distributed Algorithm

So far we have provided an offline approximation algorithm with a provable approximation ratio for the data collection maximization problem. However, the solution obtained by this algorithm is based an assumption that the global knowledge of the network topology and the profiles of sensors including their physical locations, power levels, 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 and communicates with the sensor. Thus, although the proposed off-line approximation algorithm can guarantee an approximation ratio, it lacks of scalability and may not be applicable to large-scale sensor networks.

In the following we focus on developing a fast, scalable online distributed algorithm that is more suitable to real distributive sensor networks, by removing the global knowledge assumptions. We shall make use of the solution obtained by the off-line algorithm as the benchmark to evaluate the effectiveness and efficiency of the proposed online distributed algorithm.

A. General framework of online distributed algorithms

The framework of the proposed online distributed algorithm proceeds as follows. The mobile sink periodically broadcasts a 'Probe' message with a 'Registration' timer, announcing its presence while traveling along the predefined path once per time interval, 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, etc. 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 scheduling result to the registered sensors and each registered sensor then sets its transmission time slots.

In 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. 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 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 framework of the online distributed algorithm is given in **Algorithm** 2.

B. GAP-based time slot scheduling

In the following we devise a *GAP-based time-slot scheduling algorithm* as algorithm $\mathcal{A}$. 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. We refer to this GAP-based online distributed algorithm as `Online_Appro`, and have the following lemma and theorem.

Lemma 1: Within the proposed framework of the online distributed algorithm 2, each sensor is within at most two consecutive broadcasting regions (or two consecutive time intervals).

Proof: We show the claim by contradiction. Consider

Algorithm 2 `Framework_Distributed_Algorithm`

$continue \leftarrow$ '$true$'; /* the 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;
  if the timer expires **then**
    if the mobile sink received 'Ack' messages **then**
      Call a time-slot scheduling algorithm $\mathcal{A}$ in the mobile sink to allocate the time slots in time interval j 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 updates their energy profiles when their received the 'Finish' messages. That is, registered sensor v_i updates its energy budget:
      $P(v_i) \leftarrow P(v_i) - \sum_{j \in S_i} \cdot P_{i,j} \cdot \tau$, where S_i is the set of time slots assigned to v_i by algorithm $\mathcal{A}$ in the current time interval and $S_i \subseteq A(v_i)$;
    else
      $continue \leftarrow$ '$false$'; /* finish the tour */
    end if
  else
    Waiting for replies;
  end if
end while

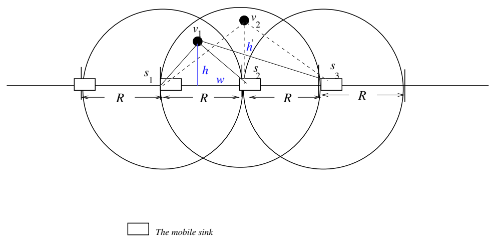


Figure 1. A sensor v_1 (or v_2) cannot be in three consecutive time intervals.

Fig. 1, 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 harvesting sensor network $G = (V \cup \{s\}, E)$, there is an online GAP-based 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 take 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 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 at most in 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. ■

VI. Special data collection maximization problem

In this section we deal with a special case of the data collection maximization problem where the transmission power at each sensor is fixed and there is only one single transmission power P'. For this special case, the problem becomes polynomially solvable and a fast, scalable online distributed algorithm is devised through a reduction to the maximum weight matching problem. We shall adopt the proposed framework of online distributed algorithms in the previous section. The detailed description of the proposed algorithm is as follows.

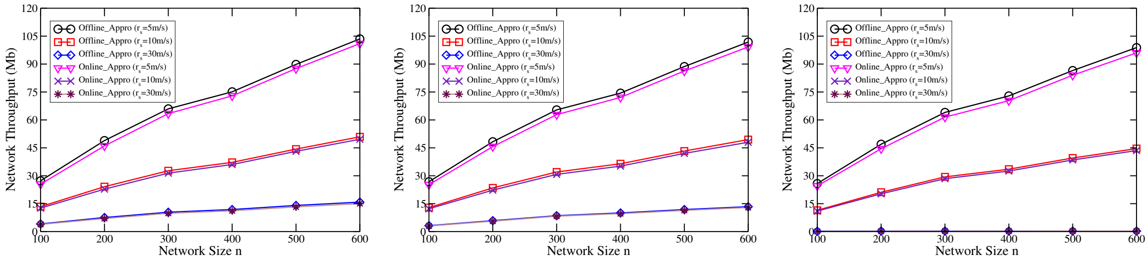
We reduce the special data collection maximization problem to the maximum weight matching problem in another auxiliary graph, which is a bipartite graph $G = (X \cup Y, E_{XY})$, where X is the set of sensors that acknowledged the probing message of the mobile sink in the beginning of time interval j, 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 amount of data received by the mobile sink from sensor v_i at time slot y_j, $D^{(0)}_{i,j} = r_{i,j} \cdot \tau$, 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 the special data collection maximization problem by reducing it to a maximum weight matching problem in another bipartite graph $G' = (\{x_i^{(k)} \mid x_i \in X, 1 \leq k \leq n'_i\} \cup Y, E')$, 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 $n'_i = \min\{\lfloor \frac{R}{r_s \cdot \tau} \rfloor, |i'_s - i'_e| + 1, \lfloor \frac{P(v_i)}{P' \cdot \tau} \rfloor\}$, where P' is the fixed transmission power of sensors.

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 $D^{(0)}_{i,j}$. Thus, finding a solution to 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 the 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 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 harvesting sensor network $G = (V \cup \{s\}, E)$, 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, identical transmission power for all sensors. 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



(a) $\tau = 1s$ (b) $\tau = 2s$ (c) $\tau = 4s$

Figure 2. Network throughput delivered by algorithms `Offline_Appro` and `Online_Appro` through varying the sink speed r_s and the network size n.

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 $G = (V, E)$ [20]. 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 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 knowledge of the entire network and the profiles of all sensors are given, an offline algorithm based on maximum weight matching for the special data collection maximization problem can also be obtained, which can deliver an exact solution in polynomial time. We refer to this offline algorithm as `Offline_MaxMatch`.

VII. Performance Evaluation

In this section we study the performance of the proposed algorithms through experimental simulation. We also 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 network throughput.

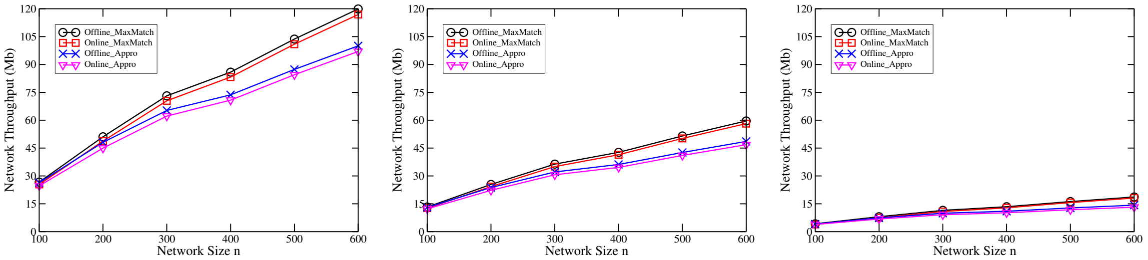
A. Experimental environment setting

We consider an energy harvesting sensor network consisting of 100 to 600 homogeneous sensor nodes randomly deployed along a pre-defined path, and a mobile sink s travels along the path at constant speed r_s to collect sensing data from one-hop sensors. We further assume that the length of the pre-defined path is $10,000m$ and the maximum distance between the location of any sensor and the path is $180m$. All sensors have identical maximum transmission ranges of 200 meters. Each sensor is powered by a $10mm \times 10mm$ square solar panel with the battery capacity of $10,000 Joules$. The solar power harvesting profile is built upon real solar radiation measurements [14], in which the total amount of energy collected from a $37mm \times 37mm$ solar panel over a 48-hour period is $655.15mWh$ in a sunny day and $313.70mWh$ in a partly cloudy day. Without loss of generality, we here adopt a 4-pairwise communication parameters setting, where the transmission parameters and corresponding distances are: $250Kbps$ with the transmission power being $170mW$ between 0 and 20 meters, $19.2Kbps$ with the transmission power being $220mW$ between 20 and 50 meters, $9.6Kbps$ with the transmission power being $300mW$ between 50 and 120 meters, and $4.8Kbps$ with the transmission power being $330mW$ between 120 and 200 meters. In the default setting the duration of each time slot τ is 1 second. Each value in figures is the mean of the results by applying each mentioned algorithm to 50 different network topologies of the same network size.

B. Performance evaluation of different algorithms

We first evaluate the performance of algorithms `Offline_Appro` and `Online_Appro` by varying the network size n from 100 to 600 and setting the mobile sink speed r_s at $5m/s$, $10m/s$, and $30m/s$, while the duration of time slot τ is fixed at $1s$, $2s$, and $4s$, respectively.

Fig. 2 demonstrates that algorithm `Offline_Appro` always outperforms algorithm `Online_Appro` slightly. For example, when $r_s = 5m/s$ and $\tau = 1s$, the network throughput of algorithm `Online_Appro` is no less than 93% of that of algorithm `Offline_Appro`. The reason behind this is that algorithm `Online_Appro` only has the local rather than the 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 derived from 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 higher speed.



(a) $r_s = 5m/s$ and $\tau = 1s$ (b) $r_s = 10m/s$ and $\tau = 1s$ (c) $r_s = 30m/s$ and $\tau = 1s$

Figure 3. Network throughput delivered by different algorithms for special case through varying the mobile sink speed r_s and the network size n.

C. Performance of different algorithms for special data collection maximization problem

Assuming that each sensor always transmits its data with one identical transmission power $300mW$, we now investigate the performance of algorithms `Offline_MaxMatch`, `Online_MaxMatch`, `Offline_Appro`, and `Online_Appro` and also study the impact of the network size n and the mobile sink speed r_s by varying n from 100 to 600 and setting r_s at $5m/s$, $10m/s$, and $30m/s$, respectively, while the duration of time slot τ is fixed at $1s$.

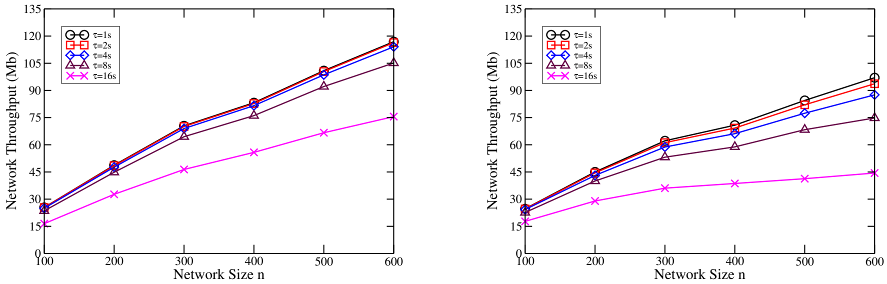
When the mobile sink speed is fixed at $5m/s$, Fig. 3(a) clearly shows that algorithm `Offline_MaxMatch` outperforms the other three algorithms. Moreover, it is observed that algorithm `Online_MaxMatch` is inferior to algorithm `Offline_MaxMatch`, as algorithm `Online_MaxMatch` only has the local rather than global knowledge of the network. However, the performance gap between them is marginal. It is also noticed that algorithm `Online_MaxMatch` outperforms the other two algorithms, and the performance gaps between 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 $n = 600$, the performance of algorithm `Online_MaxMatch` is 16% and 19% higher than that of algorithms `Offline_Appro` and `Online_Appro`, respectively. When the mobile sink speed is fixed at $10m/s$ and $30m/s$ respectively, Fig. 3(b) and 3(c) exhibit the similar performance behaviors, omitted. Fig. 3 implies that when the network size is fixed, the network throughput delivered by all mentioned algorithms decreases, with the increase of the mobile sink speed. Specifically, the network throughput delivered by algorithm `Offline_MaxMatch` when $r_s = 5m/s$ is at least 101%, 540% higher than that by itself when $r_s = 10m/s$ and $30m/s$, respectively. This is because when the mobile sink travels at a higher speed, the duration of the mobile sink travels the entire path will be shortened, while the data transmission rate does not change with the mobile sink speed, thus, the amount of data uploaded from sensors will be reduced. Although a higher speed leads to a shorter delay on data delivery, it will result in a less amount of data collected per tour too.

We then 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 600 and setting τ as $1s$, $2s$, $4s$, $8s$, and $16s$, respectively, while the mobile sink speed r_s is fixed at $5m/s$.

Fig. 4 illustrates that for each mentioned algorithm, the network throughput decreases with the increase of the duration of each time slot, and the performance gap grows bigger with the growth of network size. Specifically, in Fig. 4(a), the network throughput delivered by algorithm `Online_MaxMatch` with $\tau = 1s$ is at least 0.5%, 1%, 2%, 8%, and 50% higher than that by itself when $\tau = 2s$, $4s$, $8s$, and $16s$, respectively. In Fig. 4(b), the network throughput delivered by algorithm `Online_Appro` with $\tau = 1s$ is at least 1%, 1.5%, 2%, 9%, and 56% higher than that by itself when $\tau = 2s$, $4s$, $8s$, and $16s$, respectively. The reason behind is that with shorter time slot, the registered sensors can utilize their energy more efficiently. In detail, with a longer duration of time slot, sensors that are close to the mobile sink are more likely to lose their chances to transmit their data as they do not have enough energy. However, in realistic scenario, a proper time slot duration should be set by taking the energy consumption to run the transmitter circuitry into consideration.

VIII. CONCLUSIONS

In this paper we studied mobile data collection in an energy harvesting sensor network, using a mobile sink travelling along a pre-defined path. 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 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 propose a polynomial solution to the problem. Finally, we conducted 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.



(a) Algorithm `Online_MaxMatch`

(b) Algorithm `Online_Appro`

Figure 4. Impact of network size n and the time slot duration τ on the network throughput delivered by algorithms `Online_MaxMatch` and `Online_Appro`.

References

[1] R. Bar-Yehuda and S. Even. A local-ratio theorem for approximating the weighted vertex cover problem. *Annuals of Discrete Mathematics*, vol.25, pp.27–46, 1985.

[2] A. Chakrabarti, A. Sabharwal, and B. Aazhang. Communication power optimization in a sensor network with a path-constrained mobile observer. *ACM Trans. Sen. Netw.*, vol.2(3), pp.297–324, Aug., 2006.

[3] R. Cohen, L. Katzir, and D. Raz. An efficient approximation for the generalized assignment problem. *Information Processing Letters*, Vol.100(4), pp.162–166, Nov., 2006.

[4] M. Di Francesco, S. K. Das, and G. Anastasi. Data collection in wireless sensor networks with mobile elements: a survey. *ACM Trans. Sen. Netw.*, vol.8(1), article no.7, Aug., 2011.

[5] M. L. Fisher, R. Jaikumar, and L.-N. Wassenhove. A multiplier adjustment method for the generalized assignment problem. *Manage. Sci.*, vol.32(9), pp.1095–1103, Sep., 1986.

[6] L. Fleischer, M. X. Goemans, V. S. Mirrokni, and M. Sviridenko. Tight approximation algorithms for maximum general assignment problems. *Proc. of SODA'06*, ACM, 2006.

[7] S. Guo and Y. Yang. A distributed optimal framework for mobile data gathering with concurrent data uploading in wireless sensor networks. *Proc. of INFOCOM'12*, IEEE, 2012.

[8] S. Gao, H .Zhang, and S. K. Das. Efficient data collection in wireless sensor networks with path-constrained mobile sinks. *IEEE Trans. Mobi. Comput.*, vol.10(4), pp.592–608, Apr., 2011.

[9] L. He, J. Pan, and J. Xu. A progressive approach to reducing data collection latency in wireless sensor networks with mobile elements. *IEEE Trans. Mobi. Comput.*, http://doi.ieeecomputersociety.org/10.1109/TMC.2012.105, 2012.

[10] S. Jain, R. C. Shah, W. Brunette, G. Borriello, and S. Roy. Exploiting mobility for energy efficient data collection in wireless sensor networks. *Mob. Netw. Appl.*, vol.11(3), pp.327–339, Jun., 2006.

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

[12] A. Kansal, A. A. Somasundara, D. D. Jea, M. B. Srivastava, and D. Estrin. Intelligent fluid infrastructure for embedded networks. *Proc. of MobiSys'04*, ACM, 2004.

[13] E. Lawler. Fast approximation algorithms for knapsack problems. *Math of Operations Research*, vol.4(4), pp.339–356, Nov., 1979.

[14] 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(4), pp.947–960, Aug., 2011.

[15] W. Liang and J. Luo. Network lifetime maximization in sensor networks with multiple mobile sinks. *Proc. of LCN'11*, IEEE, 2011.

[16] R.-S. Liu, P. Sinha, and C. E. Koksal. Joint energy management and resource allocation in rechargeable sensor networks. *Proc. of INFOCOM'10*, IEEE, 2010.

[17] 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(5), pp.1021–1034, Oct., 2007.

[18] 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.*, http://doi.ieeecomputersociety.org/10.1109/TC.2012.124, 2012.

[19] S. Lin, J. Zhang, G. Zhou, L. Gu, J. A. Stankovic, and T. He. ATPC: adaptive transmission power control for wireless sensor networks. *Proc. of SenSys'06*, ACM, 2006.

[20] S. Micali and V. V. Vazirani. An $O(\sqrt{|V|}|E|)$ algorithm for finding maximum matching in general graphs. *Proc. of 21st Annual Symposium on Foundations of Computer Science*, IEEE, 1980.

[21] X. Ren, and W. Liang. Delay-tolerant data gathering in energy harvesting sensor networks with a mobile sink. *Proc. of GLOBECOM'12*, IEEE, 2012.

[22] X. Ren, and W. Liang. The use of a mobile sink for quality data collection in energy harvesting sensor networks. *Proc. of WCNC'13*, IEEE, 2013.

[23] A. A. Somasundara, A. Kansal, D. D. Jea, D. Estrin, and M. B. Srivastava. Controllably mobile infrastructure for low energy embedded networks. *IEEE Trans. Mobi. Comput.*, vol.5(8), pp.958–973, Aug., 2006.

[24] Q. Wang, D. O. Wu, and P. Fan. Delay-constrained optimal link scheduling in wireless sensor networks. *IEEE Tran. Vehi. Tech.*, vol.59(9), pp.4564–4577, Nov., 2010.

[25] X. Xu and W. Liang. Monitoring quality optimization in wireless sensor networks with a mobile sink. *Proc. of MSWiM'11*, ACM, 2011.

[26] Z. Xu, W. Liang, and Y. Xu. Network lifetime maximization in delay-tolerant sensor networks with a mobile sink. *Proc. of DCOSS'12*, IEEE, 2012.

[27] B. Zhang, R. Simon, and H. Aydin. Maximum utility rate allocation for energy harvesting wireless sensor networks. *Proc. of MSWiM'11*, ACM, 2011.