Delay-Tolerant Data Gathering in Energy Harvesting Sensor Networks With a Mobile Sink

Xiaojiang Ren   Weifa Liang

Research School of Computer Science

Australian National University

Canberra, ACT 0200, Australia

Email: richard.rxj@anu.edu.au, wliang@cs.anu.edu.au

Abstract—In this paper we consider data collection in an energy harvesting sensor network with a mobile sink, where a mobile sink travels along a trajectory for data collection subject to a specified tolerant delay constraint T. The problem is to find an optimal close trajectory for the mobile sink that consists of sojourn locations and the sojourn time at each location such that the network throughput is maximized, assuming that the mobile sink can only collect data from one-hop sensors, for which we first show that the problem is NP-hard. We then devise novel heuristic algorithms. We finally conduct extensive experiments to evaluate the performance of the proposed algorithms. We also investigate the impact of different parameters on the performance. The experimental results demonstrate that the proposed algorithms are efficient. To the best of our knowledge, this is the first kind of work of data collection for energy harvesting sensor networks with mobile sinks.

I. INTRODUCTION

Wireless sensor network has emerged as a key technology for many monitoring applications such as habitat monitoring and environmental sensing. However, the limited lifetime of batteries has hampered the large-scale deployment of such networks. A viable solution to this problem is to allow sensor nodes to harvest ambient energy such as solar energy, wind energy, vibration energy, and so on, from their surroundings [5]. 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 replacement. However, since the time-varying characteristics of energy harvesting sources, it raises numerous challenges in the routing protocol design for such networks.

Conventional sensor networks consist of a fixed sink and hundreds of sensors powered by batteries. The sensing data generated by sensors is transmitted to the sink through multi-hop relays. Since the sensors near to the sink have to relay data for others, they usually bear disproportionate amounts of traffic and thus deplete their energy much faster than others. Such an unbalanced energy consumption among sensors will shorten the network operational time, data delivery reliability, and other network performance. To mitigate this uneven energy consumption among sensors, the concept of mobile sinks has been exploited, and extensive studies have shown that mobile sinks can significantly improve various network performance including network lifetime, connectivity, data

delivery reliability, throughput, etc [1], [9], [10], [8], [12], [11], [17], [19], [18], [20], [21], [22]. Most of these studies focus on the network lifetime maximization due to the sensors are powered by energy-limited batteries. However, these data gathering algorithms for conventional sensor networks are not applicable to energy harvesting sensor networks since the sensors in energy harvesting sensor networks can be recharged periodically, and virtually such networks can be operate perpetually. Thus, this creates a shift in research focus from energy efficient to energy neutral approaches [6]. That is, the energy consumed by each node must be no greater than the energy harvested by the node. However, due to the time-varying nature of harvested energy, designing algorithm of data gathering in energy harvesting sensor networks with maximizing the network throughput poses great challenges.

In this paper we consider data collection in an energy harvesting sensor network with a mobile sink, by formulating a novel constrained optimization problem consisting of finding an optimal close trajectory for the mobile sink and scheduling the sojourn time at each sojourn location such that the network throughput (*the quantity of data collected*) is maximized, subject to a specified tolerant delay constraint. Specifically, we assume that the mobile sink traverses along a close trajectory and stops at each sojourn location in the trajectory for a certain amount of time to collect data from one-hop sensors, and each sensor that has only sufficient energy during the sojourn time can perform its data transmission. The mobile sink finally will return to its starting point within the given time bound.

Our major contribution in this paper are as follows. We consider data collection in an energy harvesting sensor networks with a mobile sink. We first formulate this mobile data collection as a joint optimization problem consisting of finding a close trajectory and sojourn time scheduling with an aim of maximizing the network throughput. We then show that the problem is NP-hard, and devise heuristic algorithms for it. The proposed algorithms exhibit low computational complexity and high scalability. Finally, we evaluate the performance of the proposed algorithms through experimental simulations. We also investigate the impact of different network parameters on the performance. The experimental results demonstrate that the proposed algorithms are efficient. To the best of our knowledge, we are not aware of any similar work on this issue in energy harvesting sensor networks with mobile sinks. This

is the first kind of work of data collection in such networks.

The rest of the paper is organized as follows. Section II summaries related works. Section III introduces the network model and problem definition, and shows the problem is NP-hard. Section IV is devoted to devising algorithms for the problem, while Section V evaluates the performance of the proposed algorithms through experimental simulations. Section VI concludes the paper.

II. RELATED WORK

As mentioned, extensive studies of data gathering in conventional sensor networks with mobile sinks have been conducted in the past several years. Most of these studies focus on network lifetime maximization, since the sensors are powered by limited-energy batteries. However, this is no longer an issue in energy harvesting sensor networks as the sensors can be continuously recharged by renewable energy. As we employ the mobile sink for data collection, it is unavoidable ti incur the data delivery delay, which usually represents *the time duration* of a sampling reading from its generation to its collection by the mobile sink. In terms of data collection with a specified tolerant delay, the closely related work in conventional sensor networks is briefly described as follows. Yun and Xia [23] considered the network lifetime maximization problem. With the assumption that each sensor does not require to send its data immediately when they are generated, instead the data can be stored at the sensor temporarily and be transmitted when the mobile sink is at the most favorable location to achieve the maximum network lifetime. They formulated the problem as an optimization problem subject to the delay bound constraint, and proposed a flow-based optimization framework. Xu *et al.* [21] addressed a delay-tolerant data collection problem for event-detection with a guaranteed collection rate. They formulated the problem as a sensor selection problem and solved the problem by incorporating the spatial-temporal correlation of the event so that the network lifetime can be significantly extended. Liang *et al.* [10], [9] incorporated the travel distance of the mobile sink into the problem formulation and proposed heuristics to find a feasible trajectory for each mobile sink so that the network lifetime can be maximized.

In contrast, very little attention has been paid to data collection in energy harvesting sensor networks with mobile sinks. Most existing studies on data collection in energy harvesting sensor networks assumed that there is a fixed sink, and the collected data is routed to the sink through multi-hop relays through tree structures [24] or flows [2], [13], [7], and focused on fair data rate allocation in response to dynamic changes of harvesting energy of sensors and the fairness of rate allocation among the sensors. The data rate allocated to each node varies, depending on the energy harvesting rate of the node [13], [7]. For example, Liu *et al.* [13], [7] formulated the problem as the lexicographic maxmin rate allocation problem, and provided a centralized algorithm for the problem by solving an integer program. They also proposed a distributed solution for the problem in special networks like DAGs. Zhang *et al.* [24] 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 of all sensors in tree networks.

III. PRELIMINARIES

A. Network model

We consider an energy harvesting sensor network $G = (V \cup S, E)$ with a set of stationary sensor nodes V which are not necessarily homogeneous, a mobile sink, and a set of potential sojourn locations S at which the mobile sink can sojourn. There is an edge between $s_j \in S$ and $v_i \in V$ in E if the mobile sink located at s_j is within the maximum transmission range of v_i. Each sensor $v \in V$ is powered by a rechargeable battery with capacity C_v and its energy is harvested from its surrounding environment (e.g. solar energy). Each sensor v senses its vicinity with sampling data generation rate g_v. Assume that the maximum transmission range of each sensor is R, and there is a link $l_{i,j}$ between a sensor $v_i \in V$ and the mobile sink located at $s_j \in S$ if the sink is within the maximum transmission range of v_i, i.e. the Euclidean distance $|l_{i,j}|$ between v_i and s_j is no greater than R. Furthermore, we assume that the mobile sink has identical transmission range as the sensors, and denote by $N(s_j)$ the neighboring set of sensors within the maximum transmission range of s_j. The mobile sink could be a mobile robot or a vehicle equipped with multiple transceivers and has unlimited energy supply in comparison with energy-constrained sensor nodes. Thus, it can collect data successfully from multiple sensors when it is within their transmission ranges.

B. Energy replenishment and consumption models

We assume that time is slotted into equal slots. Following a widely adopted assumptions: the energy replenishment rate of each sensor node is much slower than its energy consumption rate at each time slot; and the amount of energy harvested in a future time period is uncontrollable but predictable based on the source type and harvesting history [7]. Denote by C_v energy storage capacity and $B_v(t)$ the amount of energy stored at each node $v \in V$ at the beginning of time slot t, where $B_v(t) = \min\{B_v(t-1) + Q_v(t-1) - s_v(t-1), C_v\}$, $Q_v(t-1)$ and $s_v(t-1)$ are the amount of energy harvested and consumed at time slot $t-1$ and $0 \leq B_v(t) \leq C_v$. To estimate the energy budget $\hat{B}_v(t)$ of sensor v at time slot t, there are many estimation techniques available [6], [14]. For example, a simple estimation method is to use the weighted average of harvested energy in previous time slots and historical information of harvesting sources to predict the harvested energy at time slot t, i.e.

$$\hat{B}_v(t) = \min\{\sum_{j=t-q}^{t-1} \alpha_j B_v(j), C_v\}, \tag{1}$$

where α_j is a constant with $\sum_{j=q}^{t-1} \alpha_j = 1$, and q is the number of previous time slots.

As extensive studies on energy prediction have been conducted in the past several years, we here assume that a specific energy replenishment model is available. We assume that each sensor node only consumes its energy on wireless communication, while its other energy consumptions including sensing and computation will be ignored [16]. For each link $l_{i,j}$, denote by $e_{i,j}$ the energy consumption of node v_i by transmitting a unit-length data to the mobile sink located at s_j, then $e_{i,j}$ can be represented by the following equation.

$$e_{i,j} = \alpha + \beta \cdot |l_{i,j}|^{\gamma} \qquad (2)$$

where α is a constant that represents the energy consumption to run the transmitter circuitry which is negligibly small, β is a constant that represents the transmitter amplifier, and $|l_{i,j}|$ is the Euclidean distance between sensor v_i and the sojourn location s_j. The exponent γ is determined by the field measurements, which is typically a constant between 2 and 4 [4]. In this paper we assume that $\gamma = 2$.

If the mobile sink located at $s_j \in S$ is within the maximum transmission range of a sensor $v_i \in V$, then, *the survival time* of sensor v_i prior to its energy depletion $t_{i,j}$ will be determined by its residual energy $RE(v_i)$ and its distance to the mobile sink $|l_{i,j}|$, which can be expressed as $t_{i,j} = \frac{RE(v_i)}{e_{i,j} \cdot g_{v_i} \cdot L_R}$, where L_R is the length of single reading and g_v is the data generation rate of sensor v.

C. Problem definition

Given an energy harvesting sensor network $G(V \cup S, E)$ with the sensor set V and a set of potential sojourn locations S for a mobile sink, *the throughput maximization problem* in G is to find an optimal close trajectory for the mobile sink consisting of sojourn locations in S and time scheduling at each sojourn location such that the network throughput is maximized, subject to a given tolerant delay T, assuming that the mobile sink only collects data from one-hop sensor nodes, where the tolerant delay T is the total amount of time spent by the mobile sink per tour. In other words, let $S' = s_0, s_1, s_2, \cdots, s_m$ be the sequence of sojourn locations in a trajectory of the mobile sink, where for all $s_j \in S'$ with $1 \le j \le m$, t'_j is the travel time from s_{j-1} to s_j, t_j is the sojourn time at location s_j, and s_0 is the depot where the mobile sink starts from and ends at. The volume of data generated by all sensors in the network for a time period of T is $D_{total} = \sum_{v \in V}(T \cdot g_v \cdot L_R)$, while the volume of data collected by the mobile sink during this period is $D_{mobile} = \sum_{j=1}^{m} \sum_{v \in N_s(s_j)}(t_j \cdot g_v \cdot L_R)$, where $N_s(s_j)$ is the set of sensors within one-hop distance from the mobile sink located at s_j with $N_s(s_j) \subseteq N(s_j)$ and their survival time is at least t_j, and $\sum_{j=1}^{m}(t'_j + t_j) + t'_{j,0} \le T$. When the sensors in the network are homogeneous sensors, they all have the same data generation dates, and let $r_g = g_v \cdot L_R$ for all sensors. Let $n_j = |N_s(s_j)|$, then $D_{mobile} = \sum_{j=1}^{m}(t_j \cdot n_j \cdot r_g)$. *The ratio of the network throughput thus is*

$$\eta = \frac{D_{mobile}}{D_{total}} = \frac{\sum_{j=1}^{m}(t_j \cdot n_j \cdot r_g)}{T \cdot |V| \cdot r_g} = \frac{\sum_{j=1}^{m} t_j \cdot n_j}{T \cdot |V|}. \qquad (3)$$

The throughput maximization problem thus is to find a close trajectory for the mobile sink and the sojourn time scheduling at each sojourn location such that η is *maximized*, subject to a given tolerant delay constraint T.

D. NP-Hardness

In this section we show the throughput maximization problem is NP-hard by the following theorem.

Theorem 1: The decision version of the throughput maximization problem in energy harvesting sensor networks with a mobile sink is NP-hard.

Proof: We show the claim by a polynomial reduction from a well known NP-complete problem – the set cover problem [3], as follows. An instance of the set cover problem is: given a set of n elements $U = \{a_1, a_2, \ldots, a_n\}$ and a family of m subsets $\mathcal{F} = \{S_1, S_2, \ldots, S_m\}$ with $S_j \subseteq U$ and $\bigcup_{j=1}^{m} S_j = U$. Now, given a positive integer K ($K \le m$), the decision version of the instance is to determine whether there is a collection $\mathcal{C}$ of K sets from $\mathcal{F}$ such that $\bigcup_{S_i \in \mathcal{C}} S_i = U$, where $\mathcal{C} \subseteq \mathcal{F}$.

We now reduce this instance to an instance of the throughput maximization problem in an energy harvesting sensor network $G(V \cup S, E)$ with a mobile sink as follows. Each element $a_i \in U$ corresponds to a sensor $v_i \in V$ with initial energy capacity of $c_v = 1$, assume that the data generation rate of each sensor is 1. Each subset $S_j \in \mathcal{F}$ corresponds to a potential sojourn location $s_j \in S$, and its elements correspond to the sensors that are within the transmission range of the mobile sink located at s_j, i.e. $N(s_j) = S_j$ if there is no distinction between a sensor v_i and its corresponding element $e_i \in S_j$, there is an edge in E between v_i and s_j. When a location $s_j \in S$ is chosen as the sojourn location of the mobile sink, then, all sensor nodes in $N(s_j)$ that have not yet sent their data to the mobile sink will have their initial energy capacity of 1 following our energy consumption assumption (we ignore the energy consumption in data sensing). They all will consume the same amount of energy $e = 1$ by transmitting a packet from each of them to the mobile sink because they have identical transmission ranges, assuming that the energy consumption per packet transfer is 1. We further assume that the energy replenishment rate of each sensor is very slow during the given time period of T, and the total amount of energy harvested during this period is no greater than $\frac{1}{n^2}$. In other words, during the tolerant delay period of T, each sensor $v \in V$ can send a packet to the mobile sink at most once, since its survival time is $\frac{C_v}{e} = \frac{1}{1} = 1$. Once its data has been sent to the mobile sink, the accumulative harvested energy of sensor v is not enough for it to send its other sensing data to the mobile sink again. Furthermore, we assume that the time spent on the traveling by the mobile sink from one sojourn location to another is a small fraction of 1 , e.g. the time for each movement of the mobile sink is no more than $\frac{1}{n+1}$, thus, the total amount of time spent on traveling of the mobile sink $\zeta \le \frac{n}{n+1}$ (< 1) is strictly less than 1.

Having constructed an instance of the throughput maximization problem in G with a given tolerant delay $T = K + 1$, the decision version of the throughput maximization instance is to

decide whether there is a close trajectory for the mobile sink and sojourn time scheduling such that the throughput ratio is $\frac{1}{K}$ (i.e. one packet from each sensor will be collected), subject to the tolerant delay T. If there is such a solution to this instance, there is a corresponding solution to the set cover instance as follows.

It is easy to verify that the found trajectory contains exactly K sojourn locations; otherwise, the total time spent per tour will be larger than $K + 1 > T$, since the sojourn time at each sojourn location is exactly 1. Each sojourn location s_j in the trajectory corresponds to a set in $\mathcal{F}$, and all sensors within the maximum transmission range of the mobile sink located at s_j are the elements in the corresponding set $S_j \in \mathcal{F}$. Given a sojourn location s_j, only these sensors within the transmission range of the mobile sink located at s_j that have the initial energy capacities can transmit their data to the mobile sink, and the volume of data collected by the mobile sink is equal to the number of such sensors. During the time period of T, it is known that each sensor has sent exactly one packet to the mobile sink when the mobile sink is at one of the K sojourn locations, i.e. the corresponding element in U of each sensor will be in a corresponding set of the sojourn location. This means that the union of all elements in U will be covered by the K sets. As the amount of time spent on travelling ζ is less than 1, the amount of time spent by the mobile sink per tour is $K < K + \zeta < T$. Thus, given the tolerant delay $T = K + 1$, there is a solution for the instance of the throughput maximization problem with the throughput ratio of $\frac{n}{(T-1) \cdot n \cdot r_g} = 1/K$ if and only if there is a solution to the instance of the set cover problem with the cardinality of K. Note that it takes a fraction of unit-time on mobile sink traveling and we assume that a single packet per time unit will be generated. The total volume of data generated by all sensors per tour is $(T-1) \cdot r_g \cdot |V| = (T-1) \cdot 1 \cdot n$. However, it is well known that the set cover problem is NP-complete [3]. Thus, the throughput maximization problem is NP-hard. ∎

IV. HEURISTIC ALGORITHM

In this section, we deal with the throughput maximization problem by devising a scalable heuristic algorithm. The proposed algorithm proceeds iteratively. Within each iteration a new sojourn location as well as the sojourn time at the location is added to the tail of the trajectory. The procedure continues until the specified tolerant delay is no longer met.

A. Algorithm

Assume that s_i is the current location of the mobile sink in the found trajectory, we consider its next sojourn location s_j. Notice that a visited sojourn location can be visited multiple times. A location $s_j \in S$ is *a feasible sojourn location* if the time spent on all previous sojourn locations and the traveling plus the time t'_j from s_i to s_j, the sojourn time t_j at s_j, and the time $t'_{j,0}$ from s_j to s_0 is no more than T, i.e. $\sum_{l=1}^{i}(t'_l + t_l) + t'_j + t_j + t'_{j,0} \leq T$. Consider a sojourn location s_j, the volume of data collected by the mobile sink at s_j will be jointly determined by the sojourn time t_j and the number of

neighboring sensors $|N_s(s_j)|$ with survival time no less than t_j. To maximize the volume of data collected at s_j, ideally, the mobile sink should move to a location a bit far from its current location s_i, thus, all neighboring sensors there have enough energy to transmit their data to the mobile sink and the expected volume of data collected will be maximized, because most nearby sensors in $N_s(s_i)$ within the current location s_i have transmitted their data to the mobile sink already, and their energy is low which may not be enough for a longer survival time if continuing to sending their data to the mobile sink when it moves to location s_j. On the other hand, no data is collected when the mobile sink travels from location s_i to location s_j. To eliminate such data loss, we should shorten the travel distance of the mobile sink from its current sojourn location. Thus, there is a non-trivial trade-off between the travel distance (or the time spent for traveling), the number of sensor nodes, and the amount of sojourn time when we choose the next sojourn location for the mobile sink. In the following we detail the choice of the next sojourn location $s_j \in S$.

Recall that $N(s_j)$ is the set of sensors that location s_j is within their maximum transmission ranges. If $v_i \in N(s_j)$, the survival time of sensor v_i is $t_{i,j} = \frac{RE(v_i)}{r_g \cdot e_{i,j}}$. Assume that $v_{i_l} \in N(s_j)$ with $1 \leq l \leq |N(s_j)|$. Let $v_{i_1}, v_{i_2}, \ldots, v_{i_{|N(s_j)|}}$ be the sensor sequence sorted by their survival time in decreasing order, then, $t_{i_{k'},j} \geq t_{i_k,j}$ if $1 \leq k' < k \leq |N(s_j)|$. Denote by $N(s_j, i_k) = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$, clearly $N(s_j, i_k) \subseteq N(s_j)$ and $|N(s_j, i_k)| = k$ with $1 \leq k \leq |N(s_j)|$. If just the sensors in $N(s_j, i_k)$ send their sensing data to the mobile sink, then the sojourn time of the mobile sink at s_j is $t(s_j, i_k) = \min\{t_{i_l,j} \mid 1 \leq l \leq k\} = t_{i_k,j}$, and the amount of data collected by the mobile sink is $D(s_j, i_k) = t(s_j, i_k) \cdot r_g \cdot |N(s_j, i_k)| = t_{i_k,j} \cdot r_g \cdot k$. As our objective is to identify a s_j with the maximum value of $D(s_j, i_k)$, this actually is to determine an index i_k from *the value sequence*: $(t_{i_1,j} \cdot 1 \cdot r_g), \ldots, (t_{i_k,j} \cdot k \cdot r_g), \ldots, (t_{i_{|N(s_j)|},j} \cdot |N(s_j)| \cdot r_g)$. Let i_k be the index of the maximum term with value of $D(s_j, i_k)$. The time cost associated with this volume of data collected thus is $\Delta t(s_j, i_k) = t'_j + t_{i_k,j} + \frac{|l_{j,0}|}{r_m} - \frac{|l_{i,0}|}{r_m}$, assuming that the speed of the mobile sink r_m is fixed. If the location s_j will be chosen as the next sojourn location of the mobile sink, the sojourn time of the mobile sink at s_j is $t_j = t(s_j, i_k)$ and the amount of data collected is $D(s_j) = D(s_j, i_k)$. Thus, given all feasible sojourn locations, to maximize the network throughput, one such location s_j with the maximum value of $D(s_j)$ will be identified as the next sojourn location of the mobile sink.

With the addition to the trajectory by more and more sojourn locations, it will reach a point where no location in S will become a feasible sojourn location any more. In other words, consider a location s_j that is not a feasible sojourn location, it must have $T_i + t_j + t'_j + t'_{j,0} > T$ but $T_i + t'_j + t'_{j,0} < T$, where $t_j = t(s_j, i_k)$ defined as the above. For this case, if s_j is chosen as a sojourn location of the mobile sink, it must be the last sojourn location in the trajectory, and the sojourn time at it should be no more than $\Delta t_j = T - (T_i + t'_j + t'_{j,0})$. To

find an appropriate sojourn time at location s_j, we re-examine the value sequence by identifying the terms whose survival times are no greater than Δt_j and choosing a term with the maximum value (the maximum volume of data) and put its survival time as a "candidate" sojourn time of the mobile sink at s_j, and the associated set of sensors that send their data to the mobile sink as the candidate set of sensors. Meanwhile, we also identify another term with the maximum index from the value sequence that its survival time is greater than Δt_j. Let i_k be the index of the term, then, the volume of data collected is $k \cdot r_g \cdot \Delta t_j$ if its survival time is used as the sojourn time at s_j, i.e. the sojourn time is Δt_j for all the sensors in $N(s_j, i_k)$. We finally choose the larger volume one from the candidate term and the term indexed by i_k and use the survival time of the candidate item or Δt_j as the sojourn time of the mobile sink at location s_j. If there are multiple such locations, we choose the one from the locations that results in the maximum volume of data collected. The detailed description is given by the following routine `Last_Location`.

Algorithm 1: `Last_Location`

Input: A trajectory S' with the tail s_i, sink speed r_m
Output: The last sojourn location $next\_soj\_location$ of the trajectory, its sojourn time $next\_soj\_time$, and the set of sensors $next\_sensor\_set$.

begin
  $max\_data \leftarrow 0$;
  for *each location $s_j \in S$* **do**
    compute Δt_j;
    if $\Delta t_j > 0$ **then**
      Generate the value sequence:
      $t_{i_1,j} \cdot 1 \cdot r_g, t_{i_2,j} \cdot 2 \cdot r_g, \ldots, t_{i_{|N(s_j)|},j} \cdot |N(s_j)| \cdot r_g$;
      Identify a maximum term from the terms whose survival times are less than Δt_j, and let i_l be the index of the maximum term; identify the terms from the sequence whose survival times are strictly greater than Δt_j, and let i_k be the maximum index of all such terms;
      if $(t_{i_l,j} \cdot |N(s_j, i_l)| \cdot r_g \geq \Delta t_j \cdot k \cdot r_g)$ **then**
        $t_j \leftarrow t_{i_l,j}; N_s(s_j) \leftarrow N(s_j, i_l)$;
        $D(s_j) \leftarrow t_{i_l,j} \cdot l \cdot r_g$;
      else
        $t_j \leftarrow \Delta t_j; N_s(s_j) \leftarrow N(s_j, i_k)$;
        $D(s_j) \leftarrow \Delta t_j \cdot k \cdot r_g$;
      if $max\_data < D(s_j)$ **then**
        $next\_soj\_location \leftarrow s_j$;
        $max\_data \leftarrow D(s_j)$;
        $next\_sensor\_set \leftarrow N_s(s_j)$;
        $next\_soj\_time \leftarrow t_j$;
  return $next\_soj\_location$, $next\_soj\_time$ and $next\_sensor\_set$;
end

In summary, the proposed algorithm proceeds iteratively. Initially, the trajectory contains only the depot s_0. Within each iteration, a feasible sojourn location $s_j \in S$ is added to the found trajectory if $D(s_j)$ at it is the maximum one among all such locations. This procedure continues until there is no feasible location found. In the end, the last sojourn location is added to the trajectory if the tolerant delay is still met. The detailed algorithm `Max_Throu` is described as follows.

Algorithm 2: `Max_Throu`

Input: A set of potential sojourn locations $S \cup \{s_0\}$, the tolerant delay T, and the sink speed r_m
Output: The trajectory of the sink S', the sojourn time t_j and the sensor set $N_s(s_j)$ at each sojourn location $s_j \in S'$.

begin
  $S' \leftarrow \langle s_0 \rangle$ /* the location sequence in trajectory */ ;
  $max\_data \leftarrow 0$;
  /* Assume that s_i is the current sojourn location*/; /* we aim to find the next sojourn location $next\_soj\_location$ with the maximum volume */; /* of data collected $D(next\_soj\_location)$ */;
  repeat
    for *each feasible sojourn location $s_j \in S$* **do**
      Compute $t_{i,j}$ for each sensor $v_i \in N(s_j)$;
      Sort the survival time sequence in decreasing order. Let $t_{i_1,1}, t_{i_2,j}, \ldots, t_{i_{|N(s_j)|},j}$ be the sorted sequence and $v_{i_1}, v_{i_2}, \ldots, v_{i_{|N(s_j)|}}$ the corresponding sensor sequence;
      Find the maximum term from the value sequence
      $t_{i_1,j} \cdot 1 \cdot r_g, t_{i_2,j} \cdot 2 \cdot r_g, \ldots, t_{i_{|N(s_j)|},j} \cdot |N(s_j)| \cdot r_g$;
      Let i_k be the index of the maximum term;
      $D(s_j) \leftarrow t_{i_k,j} \cdot k \cdot r_g$;
      if $max\_data < D(s_j)$ **then**
        $next\_soj\_location \leftarrow s_j$;
        $next\_sensor\_set \leftarrow \{ v_{i_l} \mid 1 \leq l \leq k\}$;
        $next\_soj\_time \leftarrow t_{i_k,j}$;
        $max\_data \leftarrow D(s_j)$;
    $S' \leftarrow S' \cup \{next\_soj\_location\}$;
    $max\_data \leftarrow 0$;
    Update the energy of sensors;
  until *there is not any feasible sojourn location* ;
  call Routine `Last_Location` (S') by adding the last sojourn location to the trajectory;
  $S' \leftarrow S' \cup \{next\_soj\_location\}$;
  return S', $N_s(s_j)$ and t_j for each $s_j \in S'$.
end

Theorem 2: Given an energy harvesting sensor network $G(V \cup S, E)$, a mobile sink and a specified tolerant delay T, there is an algorithm `Max_Throu` for the throughput maximization problem in G, which takes $O(n \log n \cdot |S| \cdot T)$ time, where $n = |V|$.

    Proof: Clearly, it is easy to verify that the solution

delivered by algorithm `Max_Throu` is a feasible solution as the tolerant delay constraint is met. In the following we analyze the time complexity of the proposed algorithm.

Within each iteration, a new sojourn location will be added to the found trajectory, while finding such a new location takes $O(|N_{max}(S)| \log |N_{max}(S)| \cdot |S|) = O(n \log n \cdot |S|)$ time due to sorting and $|S|$ feasible sojourn location to be examined, where $N_{max}(S) = \{N(s) \mid |N(s)| \geq |N(s')|$ for all $s, s' \in S\}$. The number of iterations is determined by T. The algorithm takes $O(n^2 \log n \cdot n')$ where n' is the number of sojourn locations in the trajectory, If the sojourn time spent at each sojourn location is at least one-unit time, then, the algorithm takes $O(n^2 \log n \cdot T)$ time since $|N_{max}(S)| \leq |V| = n$, $n' \leq T$, and $|S| << n$. ∎

B. Further improvement

The proposed algorithm `Max_Throu` can be further improved. Recall that within each iteration of algorithm `Max_Throu`, a sojourn location s_j with the sojourn time t_j is added to the trajectory if this leads to the maximum volume of data collected $D(s_j)$. However, the time cost associated with this data collection may not be economical. As the total time per tour is bounded by T, we here introduce another metric by incorporating the time cost while choosing the next sojourn location, and we refer to this metric as *the data gain per time-unit* metric, which is defined as follows.

$$gain(s_j) = \max\{\frac{D(s_j, i_k)}{\Delta t(s_j, i_k)} \mid 1 \leq k \leq |N(s_j)|\}, \quad (4)$$

where $\Delta t(s_j, i_k) = t'_j + t_{i_k, j} + t'_{j,0} - t'_{i,0}$.

We thus have another algorithm `Impro_Max_Throu` for the throughput maximization problem, which is essentially identical to algorithm `Max_Throu`. The only difference between them lies in the choice of the next sojourn location of the mobile sink. Instead of choosing the one with the maximum volume of data collected, algorithm `Impro_Max_Throu` will choose a location s_j with the maximum data gain per time-unit $gain(s_j)$. The algorithm description is similar to the one for algorithm `Max_Throu`, omitted.

V. PERFORMANCE EVALUATION

In this section we study the performance of the proposed algorithms through experimental simulation. We also investigate the impact of related parameters: the tolerant delay T and the number of sensors n on the network performance.

A. Simulation environment

We consider an energy harvesting sensor network consisting of 100 to 600 sensors randomly deployed in a $100m \times 100m$ square region. For the sake of convenience, we assume that both the mobile sink and the sensor nodes have identical maximum transmission ranges of 30 meters. The potential sojourn locations in S are randomly generated with the default setting of $|S| = 50$. Each sensor has a data generation rate $r_g = 1Kbps$ and its battery capacity is $1,000 Jules$. We set $\beta = 150nJ/bit/m^2$ for the energy consumption

model. Each sensor is powered by a solar panel with a dimension $(10mm \times 10mm)$. The solar power harvesting profile is obtained from the Baseline Measurement System at the National Renewable Energy Laboratory [15], in which the total amount of energy collected from a $37mm \times 33mm$ solar cell over a 48-hour period is $655.15mWh$ in a sunny day and $313.70mWh$ in a partly cloudy day. Thus, we assume the energy replenishment rate of each sensor is a random value ranged in $[0.4, 0.9]mJ/s$. We assume the speed of the mobile sink is $r_m = 2m/s$. Each value in figures is the mean of the results by applying each mentioned algorithm to 15 different network topologies of the same network size.

B. Throughput performance evaluation of different algorithms

We first evaluate the performance of algorithms `Max_Throu` and `Impro_Max_Throu` against that of another heuristic `Random_Throu`, which is a variant of algorithm `Max_Throu` by randomly selecting a sojourn location in S in each iteration. We evaluate these algorithms by varying the network size from 100 to 600 with an increment of 100 while fixing the tolerant delay at $100s$ and $800s$. Fig. 1 shows that algorithm `Impro_Max_Throu` always outperforms the others. Specifically, when $T = 100s$, the throughput ratio of algorithm `Impro_Max_Throu` is at least 5% and 23% higher in comparison with these of algorithms `Max_Throu` and `Random_Throu`, while $T = 800s$ the throughput ratio of algorithm `Impro_Max_Throu` is at least 13% and 46% higher compared with these of algorithms `Max_Throu` and `Random_Throu`.

We then study the impact of the tolerant delay T and network size n on the network throughput ratio, by varying n from 100 to 600 with an increment of 100 and setting T as 100, 200, 400, 800, 3,200, and 6,400 seconds, respectively. From Fig. 2, it can be seen that with the growth of T, the throughput ratios of algorithms `Impro_Max_Throu` and `Max_throughput` increase too. However, when $T \geq 3,200$, the gap of the throughput ratio between them becomes smaller, which is around 0.09. This is because that all potential sojourn locations in S have almost been visited. It can also be seen that with the growth of network size, the throughput ratios of all three algorithms decrease slightly when $T = 800$. Specifically, when the network size n increases from 100 to 600, the throughput ratio of algorithm `Impro_Max_Throu` is 0.0794, 0.0763, 0.0726, 0.0717, 0.0708, and 0.0698, respectively, the throughput ratio of algorithm `Max_Throu` is 0.0702, 0.0628, 0.0620, 0.0601, 0.0582, and 0.0595, respectively, and the throughput ratio of algorithm `Random_Throu` is 0.0525, 0.0465, 0.0497, 0.0478, 0.0457, and 0.0458, respectively. In general, the throughput ratio of algorithm `Random_Throu` is about 70% of either algorithm `Max_Throu` and algorithm `Impro_Max_Throu`.

VI. CONCLUSION

In this paper we studied mobile data collection in an energy harvesting sensor network with a mobile sink. We first formulated the problem as a joint optimization problem consisting
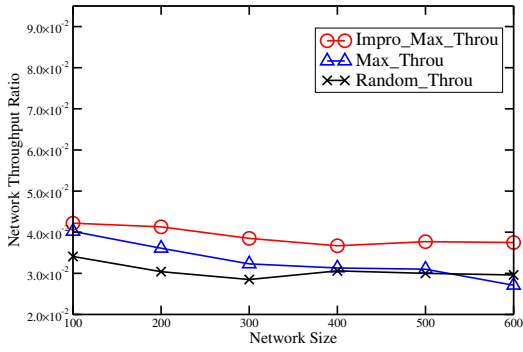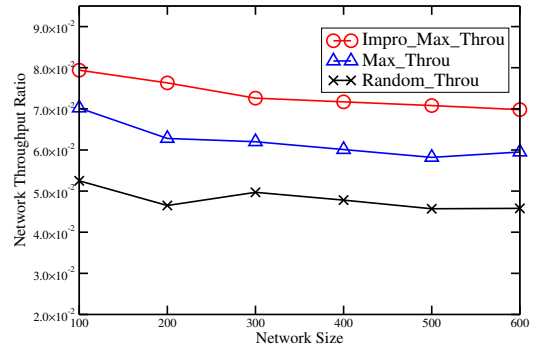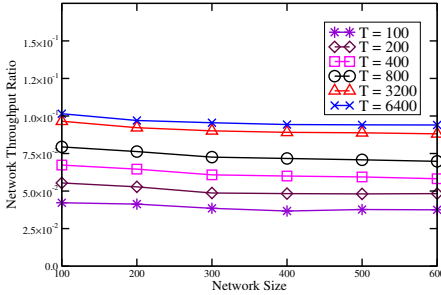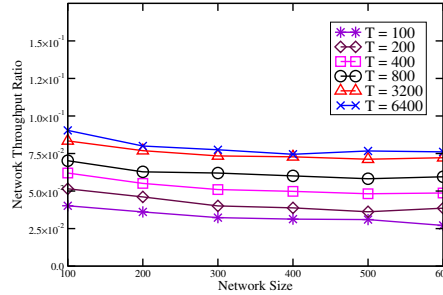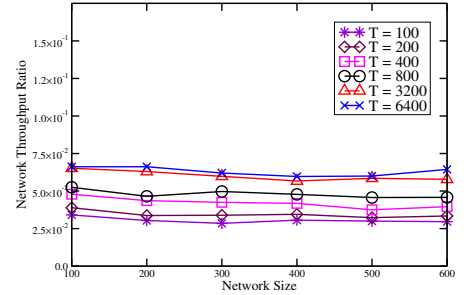
(a) $T = 100s$ (b) $T = 800s$

Fig. 1. The throughput ratio performance of different algorithms by varying the network size n and setting the tolerant delay $T = 100$ and 800

(a) Algorithm `Impro_Max_Throu`   (b) Algorithm `Max_Throu`   (c) Algorithm `Random_Throu`

Fig. 2. Impact of parameters T and n on the throughput ratios

of trajectory finding and sojourn time scheduling with an aim to maximize the network throughput. We then showed the problem is NP-hard and devised heuristic algorithms. Finally, we conducted experiments by simulations to evaluate the performance of the proposed algorithms. The experimental results demonstrate that the proposed algorithms are efficient.

REFERENCES

[1] S. Basagni, A. Carosi, E. Melachrinoudis, C. Petrioli, and Z. M. Wang. Controlled sink mobility for prolonging wireless sensor networks lifetime. *Wireless Networks*, Vol. 14, pp.831–858, 2008.

[2] K.-W Fan, Z. -Z Zheng, and P. Sinha. Steady and fair rate allocation for rechargeable sensors in perpetual sensor networks. *Proc. of SenSys*, ACM, 2008.

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

[4] W. Heinzelman. Application-specific protocol architectures for wireless networks. *Ph.D. Thesis*, Massachusetts Institute of Technology, 2000.

[5] X. Jiang, J. Polastre, and D. Culler. Perpetual environmentally powered sensor networks. *Proc. of IPSN*, ACM, 2005.

[6] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava. Power management in energy harvesting sensor networks. *ACM Trans. Embed. Comput. Sys.*, Vol.6, Sep. 2007.

[7] 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, pp.947–960, 2011.

[8] J. Luo and J-P Hubaux. Joint mobility and routing for lifetime elongation in wireless sensor networks. *Proc. of INFOCOM*, IEEE, 2005.

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

[10] W. Liang, J. Luo, and X. Xu. Prolonging network lifetime via a controlled mobile sink in wireless sensor networks. *Proc. of GLOBECOM*, IEEE, 2010.

[11] W. Liang, P. Schweitzer, and Z. Xu. Approximation algorithms for capacitated minimum spanning forest problems in wireless sensor networks with a mobile sink. To appear in *IEEE Trans. on Computers*, June, 2012.

[12] J. Luo, J. q, M. Piorkowski, M. Grossblauser, and J-P Hubaux. Mobiroute: routing towards a mobile sink for improving lifetime in sensor networks. *Proc. of DCOSS*, IEEE, 2006.

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

[14] C. Moser, J. Chen, and L. Thiele. Reward maximization for embedded systems with renewable energies. *Proc. of RTCSA*, IEEE, 2008.

[15] 'National Renewable Energy Laboratory'. http://www.nrel.gov.

[16] G. J. Pottie. Wireless sensor networks. In *Information Theory Workshop*, ACM, pp.139–140, 1998.

[17] Y. Shi and Y. T. Hou. Theoretical results on base station movement problem for sensor network. *Proc. of INFOCOM*, IEEE, 2008.

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

[19] X. Xu, W. Liang, and T. Wark. Data quality maximization in sensor networks with a mobile sink. *Proc. of DCOSS*, IEEE, 2011.

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

[21] X. Xu, J. Luo, and Q. Zhang. Delay tolerant event collection in sensor networks with mobile sink. *Proc. of INFOCOM*, IEEE, 2010.

[22] G. Xing, T. Wang, W. Jia, and M. Li. Rendezvous design algorithms for wireless sensor networks with a mobile base station. *Proc. of MobiHoc*, ACM, 2008.

[23] Y. Yun and Y. Xia. Maximizing the lifetime of wireless sensor networks with mobile sink in delay-tolerant applications. *IEEE Trans. Mobile Computing*, Vol. 9, pp.1308–1318, 2010.

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