Utility Maximization of Temporally-correlated Sensing Data in Energy Harvesting Sensor Networks

Rongrong Zhang, Jian Peng, Wenzheng Xu, Member, IEEE, Weifa Liang, Senior Member, IEEE, Zheng Li, and Tian Wang

Abstract—Sensing data collection in energy harvesting sensor networks poses great challenges, since energy generating rates of different sensors vary significantly. Most existing studies on efficient data collection assumed that the sensing data from a sensor is temporally independent. We however notice that such sensing data is highly temporally correlated, rather than independent. In this paper we study the problem of allocating energy and data rates to sensors, and performing sensing data routing in an energy harvesting sensor network for a given monitoring period, such that the utility sum of temporally-correlated data collected from sensors in the period is maximized, subject to the temporally-spatially varying harvesting energy constraint on each sensor. We then propose a near-optimal algorithm for the data utility maximization problem. We finally evaluate the performance of the proposed algorithm with real solar energy data. Experimental results show that the proposed algorithm is very promising and the utility sum of collected sensing data is up to 10% larger than that by the state-of-the-art.

Index Terms—energy harvesting sensor networks, temporally correlated sensing data, data utility maximization

1 INTRODUCTION

Wireless sensor networks (WSNs) have wide applications in Internet of Things (IoT), including environmental monitoring, electric grid network monitoring, water network monitoring, precision agriculture, target tracking, etc [1], [8], [10], [12], [16], [30], [31]. The limited battery capacities of sensors however greatly restrict the large-scale deployment of WSNs. In recent years, a promising technique to replenishing sensor energy is proposed, that is, sensors are enabled to harvest renewable energy from their surrounding environment, such as solar energy, wind energy, thermal energy, etc [24], [28], [32], [35], [41], [42].

Sensing data collection in energy harvesting sensor networks poses great challenges, since energy harvesting rates of sensors are not only temporally-varying, but also spatially-varying. For example, assume there are two sensors $v_1$ and $v_2$ powered by solar energy, see Fig. 1(a), from which it can be seen that sensor $v_1$ can harvest solar energy in daytime, while sensor $v_2$ barely generates any energy in the afternoon, as it will lose its exposure to sunlight after 2:00pm due to the shadowing of a nearby building. Fig. 1(b) demonstrates the energy harvesting rates of sensor $v_1$ at different time points in a day, while Fig. 1(c) plots the energy rates of sensor $v_2$, where these two energy harvesting profiles are adopted from [6].

Extensive studies on the efficient data collection in energy harvesting sensor networks have been conducted in past years [3], [5], [6], [7], [17], [19], [37], [38], [39]. For example, Dong et al. [7] proposed a cluster-based routing protocol and selected cluster head nodes by considering harvested sensor energy as well as the distance between each sensor and the base station. Zhang et al. [39] focused on maximizing data utility by jointly optimizing energy allocation, data sensing, and data routing, given the amount of harvested energy of each sensor in a given monitoring period. Lu et al. [19] investigated the problem of maximizing the total utility of collected data from sensors. They developed a distributed algorithm to allocate an optimal data rate for each sensor. Chen et al. [3] considered the utility maximization problem, where the energy budget of each sensor at each time slot is equal to the minimum value among its battery residual energy and its past average harvested energy per time slot. Deng et al. [6] also studied the data utility maximization problem in an energy harvesting sensor network, where data routing paths are given and fixed. They also extended their study to the case where the routing paths are allowed to change over time [5].

We notice that most existing studies [3], [5], [6], [7], [19], [39] assumed that the sensing data collected from each sensor in consecutive time slots is independent, where a time slot is a period of short time, e.g., an hour. Their objective usually is to maximize the total utility of collected data in a period of $T$, i.e., $\max \sum_{v_i \in V} \sum_{t=1}^{T} U(D_{it})$, where $U(D_{it})$ is a utility function to measure the quality of the collected data $D_{it}$ from a sensor $v_i$ at time slot $t$. 

R. Zhang, J. Peng, W. Xu (Corresponding author), and Z. Li are with College of Computer Science, Sichuan University, Chengdu, 610065, P.R. China. E-mail: rongrongzhang@stu.scu.edu.cn, jianpeng@scu.edu.cn, wenzheng.xu3@gmail.com, lizheng@scu.edu.cn

W. Liang is with the Research School of Computer Science, The Australian National University, Canberra, ACT 2601, Australia. E-mail: wliang@cs.anu.edu.au

T. Wang is with College of Computer Science and Technology, Huazhao University, P.R. China. E-mail: cs_tianwang@163.com

Notice that the first two authors (R. Zhang and J. Peng) contributed equally to this work, i.e., co-primary authors.
The sensing data from each sensor however usually is temporally correlated in a real sensor network. For example, consider a sensor network for environmental monitoring, which consists of two sensors $v_1$ and $v_2$ (see Fig. 1 (a)). Assume that sensor $v_1$ monitors the temperature of its surrounding environment and sensor $v_2$ measures the air moisture. We also assume that the residual energy of sensor $v_1$ can support its data transmission for three time slots with a data rate of 1 kbps, while the residual energy of $v_2$ can support its data transmission for only one time slot with the rate of 1 kbps. For simplicity, we further assume the energy consumptions on data transmission and data reception are equal. Following the existing studies, sensor $v_1$ transmits its data to the base station for three time slots, while sensor $v_2$ does not upload any data to the base station, as the sensing data from each sensor at different time slots are assumed to be independent in these studies, and the relaying for sensor $v_2$ by sensor $v_1$ will consume its energy on data reception and data transmission. We however note that the sensing temperature data from sensor $v_1$ will not change too much in a short time. Then, in the existing studies, the collected data from sensor $v_1$ in the three consecutive time slots is highly correlated, while the air moisture data from sensor $v_2$ has not been collected at all. However, it is important to collect sufficient data from both the temperature sensor $v_1$ and the air moisture sensor $v_2$, such that the environment are better monitored.

Unlike existing studies that assumed independent sensing data from each sensor in consecutive time slots, in this paper we consider that the sensing data from each sensor is temporally correlated. We aim to allocate sensor energy, sensor data rates, and data routing for a given period of $T$ such that the utility sum $\sum_{v_i \in V} U(\sum_{t=1}^{T} D_{it})$ of temporally-correlated sensing data collected is maximized, where the energy harvesting rates of sensors are not only temporally-varying, but also spatially-varying. Then, both the temperature sensor $v_1$ and the air moisture sensor $v_2$ can transmit their sensing data to the base station for one time slot, and the collected sensing data has more non-redundant information.

Notice that it is challenging to apply existing algorithms in [3], [5], [6], [7], [19], [39] for the problem in this paper, as these algorithms highly depend on the assumption that the sensing data from each sensor is temporally independent. Therefore, new algorithms must be designed to tackle the temporal correlation.

The novelties of this paper are twofold. (i) First, we consider the utility data collection of temporally-correlated sensing data in a given long period $T$, by formulating a novel data utility maximization problem, which takes into account not only the different residual energy of different sensors, but also the both temporally-varying and spatially-varying energy generating rates in the future. (ii) Second, we smartly transform the challenging data utility maximization problem to another convex optimization problem, and show how to convert a near-optimal solution to the latter problem to a sub-optimal solution to the former problem.

The main contributions in this paper are as follows.

- We first consider a novel optimization problem of allocating sensor energy, data rates, and data routing for a given period in an energy-harvesting sensor network, such that the utility of temporally-correlated sensing data during the period is maximized, subject to the changing energy constraint on each sensor.

- We then propose a near-optimal algorithm for the data utility maximization problem.

- We finally evaluate the performance of the proposed algorithm, using real solar energy data. Experimental results show that the proposed algorithm is very promising. Especially, the sensing data from different sensors is more fairly collected in the proposed algorithm, and the utility sum by the algorithm is up to 10% larger than that by the state-of-the-art.

The remainder of this paper is organized as follows. Section 2 describes the network model and formulates the problem. Section 3 devises a novel algorithm for the data utility maximization problem. Section 4 evaluates the performance of the proposed algorithm. Section 5 further reviews the related work. Finally, Section 6 concludes this paper.

2 Preliminaries

In this section, we first introduce the network model, energy consumption model and energy harvesting model, we then define the problem.

2.1 Network model

We consider an energy harvesting sensor network $G = (V \cup \{s\}, E)$, where $V$ is a set of $n$ sensors $v_1, v_2, \cdots v_n$ that are randomly deployed in a monitoring area, and $s$ is a base station for collecting data from sensors. Notice that the sensors may be heterogeneous, such as temperature sensors,
air moisture sensors, pressure sensors, etc. There is an edge in E between a sensor and the base station or two sensors if they are within the transmission range of each other. Each sensor $v_i \in V$ is powered by a rechargeable battery with an energy capacity $B_i$, and the battery can be charged by harvesting energy from its surrounding environment (e.g., solar energy). Fig. 2 illustrates an energy harvesting sensor network. In this paper, we consider the data collection in a multi-hop manner. That is, a sensor close to the base station is the relay of its neighboring sensors. There is an edge between sensors $v_i$ and $v_j$, and each sensor $v_i \in V$ can be reached by the base station $s$ via the relay of its neighboring sensors, and constraint (1) ensures that the base station $s$ receives the sensing data from all sensors.

2.2 Energy consumption model

Every sensor consumes its energy on data sensing, data reception and data transmission. Let $P_s(v_i, t)$, $P_{rx}(v_i, t)$, $P_{tx}(v_i, t)$ be the energy consumptions of each sensor $v_i$ for its data sensing, data reception and data transmission at time slot $t$, respectively. We adopt the real energy consumption model from [13], where

$$P_s(v_i, t) = r_{it} \times \mu_s \times \tau,$$

$$P_{rx}(v_i, t) = \sum_{v_j \in N(v_i)} f_{jit} \times \mu_{rx} \times \tau,$$

$$P_{tx}(v_i, t) = \sum_{v_j \in N(v_i)} f_{ijt} \times (\beta_1 + \beta_2 d_{ij}^2) \times \tau,$$

$\mu_s$ and $\mu_{rx}$ are the energy consumption rates for sensing and receiving per unit data, respectively, $\tau$ is the duration of each time slot. $\sum_{v_j \in N(v_i)} f_{ijt}$ and $\sum_{v_j \in N(v_i)} f_{jit}$ are the data reception and transmission rates of sensor $v_i$ at time $t$. Following [13], the term $\alpha$ is the path-loss exponent whose typical value is 2 or 4, $d_{ij}$ is the Euclidean distance between sensors $v_i$ and $v_j$. Specifically, the values of these parameters are: $\mu_s = 60 \times 10^{-9} J/b$, $\mu_{rx} = 135 \times 10^{-9} J/b$, $\beta_1 = 45 \times 10^{-9} J/b$, $\beta_2 = 10 \times 10^{-12} J/m^2$ if $\alpha = 2$; or $\beta_2 = 0.001 \times 10^{-12} J/m^4$ if $\alpha = 4$.

The energy consumption $P(v_i, t)$ of sensor $v_i$ at time slot $t$ then is

$$P(v_i, t) = P_s(v_i, t) + P_{rx}(v_i, t) + P_{tx}(v_i, t) = a \cdot r_{it} + b \cdot \sum_{v_j \in N(v_i)} f_{jit} + \sum_{v_j \in N(v_i)} c_{ij} \cdot f_{ijt},$$

where $a = \mu_s \tau$, $b = \mu_{rx} \tau$, and $c_{ij} = (\beta_1 + \beta_2 d_{ij}^2) \tau$ are constants.

2.3 Energy harvesting model

As sensors are powered by renewable energy (e.g., solar energy), the amount of harvested energy by a sensor in future is uncontrollable, but is predictable based on historical records and weather forecast. It is well-known that the weather condition changes significantly during a day. For example, it may be sunny at noon, but there is no sunlight at night. We assume that the energy harvesting rate of each sensor is predictable in a period $T$ (e.g., 24 hours). We further assume that the rate does not change within one time slot, or such changes are negligible. To estimate the amount of harvested energy of each sensor at each time slot $t$, we extend the Exponentially Weighted Moving-Average algorithm [11], by taking the predicted weather condition at different time slots into account. The weather condition at time slot $t$ on the next day $d$ can be obtained by weather forecast. Specifically, denote by $H(v_i, d, t)$ and $H(v_i, d-1, t)$ the predicted value and the actual value of the harvested energy at time slot $t$ on day $d-1$.

It is well-known that the amount of harvested energy $H(v_i, d, t)$ of sensor $v_i$ highly depends on the weather condition at time slot $t$, and the amount of harvested energy is high if it is sunny at time slot $t$, whereas the energy is very low if it is rainy. We thus use a factor $\gamma(d, t)$ to indicate the goodness of the weather condition at time $t$ on day $d$, and the larger the value of $\gamma(d, t)$ is, the better the weather is, where $0 \leq \gamma(d, t) \leq 1$. For example, $\gamma(d, t) = 1$ if it is sunny; $\gamma(d, t) = 0.8$ if it is partly cloudy; $\gamma(d, t) = 0.5$ if it is cloudy; $\gamma(d, t) = 0.2$ if it is rainy; and $\gamma(d, t) = 0.1$ if it is stormy.
We predict the amount of harvested energy of sensor \( v_i \) at time slot \( t \) on day \( d \) as follows.

\[
\overline{H}(v_i, d, t) = (\omega \cdot \overline{H}(v_i, d - 1, t) + (1 - \omega) \cdot H(v_i, d - 1, t)) \cdot \gamma(d, t)
\]  

(6)

where \( \omega \) is a given weight with \( 0 \leq \omega \leq 1 \). For the sake of convenience, we abbreviate \( \overline{H}(v_i, d, t) \) by \( \overline{H}(v_i, t) \).

Denote by \( RE(v_i, t) \) the amount of residual energy of sensor \( v_i \) at the end of time slot \( t \). Then, the amount \( RE(v_i, t + 1) \) of residual energy of sensor \( v_i \) at the end of next time slot \( t + 1 \) is

\[
RE(v_i, t + 1) = \min \{ RE(v_i, t) + \overline{H}(v_i, t + 1) - P(v_i, t + 1), B_i \}
\]  

(7)

where \( B_i \) is the battery capacity of sensor \( v_i \). \( \overline{H}(v_i, t + 1) \) is the amount of harvested energy at time slot \( t + 1 \), and \( P(v_i, t + 1) \) is the energy consumption at time slot \( t + 1 \).

### 2.4 Problem definition

It is ideal that each sensor can transmit all its sensing data back to the base station. However, the amounts of harvested energy in sensors usually are limited. Furthermore, the energy harvesting rate of each sensor may experience dramatic changes during the whole period \( T \). It is very critical to make full use of the harvested energy to collect as much sensing data as possible.

We note that the sensing data from each sensor in consecutive time slots may be highly correlated. For example, in a sensor network for monitoring temperature, the temperature in a short time (e.g., one hour) does not change too much. We thus introduce a utility function \( U(D_i) \) to measure the quality of collected data \( D_i \) from each sensor \( v_i \), where \( U(D_i) \) is an increasing, twice-differentiable and strictly concave function with respect to the total amount \( D_i \) of data collected from sensor \( v_i \) during period \( T \), and \( D_i = \sum_{t=1}^{T} r_{it} \). For example, we can adopt the widely used logarithmic function as the utility function, i.e., \( U(D_i) = \log_2(D_i + 1) \).

Given an energy harvesting sensor network \( G = (V \cup \{s\}, E) \), and the harvested energy \( \overline{H}(v, t) \) of each sensor at each time slot \( t \) in a period \( T \), the data utility maximization problem in \( G \) is to allocate data rate \( r_{it} \) for each sensor \( v_i \in V \) and the data transmission rate \( f_{ijt} \) for each link \((v_i, v_j) \in E\) so that the utility sum of data collected from sensors in period \( T \) is maximized, i.e.,

\[
P_1: \max_{r_{it}, f_{ijt}} \sum_{v_i \in V} U(\sum_{t=1}^{T} r_{it}),
\]

subject to

\[
r_{it} + \sum_{v_j \in N(v_i)} f_{ijt} = \sum_{v_j \in N(v_i)} f_{ijt}, \quad \forall v_i \in V, 1 \leq t \leq T
\]

(9)

\[
\sum_{v_i \in N(s)} f_{ist} - \sum_{v_i \in N(s)} f_{ist} = \sum_{v_i \in V} r_{it}, \quad 1 \leq t \leq T
\]

(10)

\[
P(v_i, t) \leq RE(v_i, t - 1) + \overline{H}(v_i, t), \quad \forall v_i \in V, 1 \leq t \leq T
\]

(11)

\[
RE(v_i, t) = \min \{ RE(v_i, t - 1) + \overline{H}(v_i, t) - P(v_i, t), B_i \},
\]

(12)

\[
0 \leq r_{it} \leq R_i, \quad \forall v_i \in V, 1 \leq t \leq T
\]

(13)

\[
f_{ijt} \geq 0, \quad \forall (v_i, v_j) \in E, 1 \leq t \leq T
\]

(14)

\[
0 \leq RE(v_i, t) \leq B_i, \quad \forall v_i \in V, 1 \leq t \leq T
\]

(15)

where \( \tau \) is the duration of each time slot \( t \), and \( B_i \) is the battery capacity of sensor \( v_i \). Constraint (9) indicates that each sensor should forward the received data and the sensing data from itself to the base station via the relay of its neighboring sensors. Constraint (10) ensures that the base station \( s \) receives the sensing data from all sensors. Constraint (11) implies that the energy consumption \( P(v_i, t) \) of each sensor \( v_i \) at each time slot \( t \) cannot exceed its energy budget \( RE(v_i, t - 1) + \overline{H}(v_i, t) \). Constraint (12) demonstrates the energy relationship between the amounts of residual energy \( RE(v_i, t) \) and \( RE(v_i, t - 1) \) of sensor \( v_i \) at each time slot \( t \) and its previous time slot \( t - 1 \).

Notice that we do not consider the spatial correlation of data collected from different sensors [15], as we assume that the sensor network consists of different types of sensors, e.g., temperature sensor, air moisture sensor, pressure sensor, etc. Moreover, the data utility maximization problem will become intractable if we consider both the temporal correlation and spatial correction of sensor data at the same time, and we will put this as our future study.

### 2.5 Software defined sensor network framework

Existing studies on the data utility maximization problem assumed that the allocation of sensor energy, sensor data rates, and data routing are performed in a distributed way. However, the distributed way is inflexible to deal with a multiple tasks sensor network, since sensors in a WSN may be heterogeneous. In contrast, in this paper we incorporate the software-defined network (SDN) concept into wireless sensor networks, which decouples the control plane from the data plane. Then, the base station is the center controller, and every software-defined sensor is able to perform different sensing tasks [36]. By doing so, the base station computes the data rates and data routing in a resource-rich server, since the computing needs intensive computing resources, while sensors only perform sensing tasks, and transmit data by following the routing. We briefly describe the software defined sensor network framework as follows.

Each sensor \( v_i \) measures its energy harvesting rate \( H(v_i, t) \) and sends its energy information to the base station \( s \) at each time slot, by piggybacking the information in its sensing data. Having received the energy information from each sensor, the base station then predicts the future energy harvesting rate of each sensor.

The base station calculates the data rates \( r_{it} \), and data routing \( f_{ijt} \) in a period \( T \), based on the predicted future energy information, then sends the data rate \( r_{it} \) and routing \( f_{ijt} \) to each sensor \( v_i \) at the beginning of each time slot \( t \), via a given fixed routing path, e.g., a path from the base station \( s \) to \( v_i \) with the minimum energy consumption. We assume that the energy consumption for sending energy information, data rates, and routing information of each sensor is significantly smaller than the energy consumed for receiving and transmitting sensing data in period \( T \), thus is negligible. In case the energy consumption cannot be ignored, we can take the energy consumption into account, by simply reducing the initial residual energy \( RE(v_i, 0) \) of sensor \( v_i \) by an amount of the energy consumption.

### 3 Algorithm for the Data Utility Maximization Problem

In this section, we propose a near-optimal algorithm for the data utility maximization problem.

#### 3.1 Algorithm

The basic idea of the proposed algorithm is to first transform the data utility maximization problem \( P_1 \) into another convex optimization problem, and then solve the convex optimization problem. Then, a near-optimal solution to the convex optimization problem in turn returns a sub-optimal solution to the original problem.
3.1.1 Transform the original problem

Given an energy harvesting sensor network \( G = (V \cup \{s\}, E) \), a time period \( T \), the initial residual energy \( RE(v_i, 0) \) of each sensor \( v_i \) and the harvested energy \( \overline{P}(v_i, t) \) of sensor \( v_i \) at each time slot \( t \) in period \( T \), notice that the original data utility maximization problem P1 is not a convex optimization problem, since both the objective function and Constraint (12) are not convex. We transform problem P1 into another convex optimization problem P2.

\[
P_2 : \min_{r_{it}, f_{ijt}, RE(v_i, t)} \sum_{v_i \in V} U(D_i),
\]
subject to constraints (9), (10), (13)–(15), and

\[
RE(v_i, t) \leq RE(v_i, t - 1) + \overline{P}(v_i, t) - P(v_i, t), \quad \forall v_i \in V, 1 \leq t \leq T,
\]

where the variables are the data rate \( r_{it} \) of each sensor \( v_i \) at time slot \( t \), the transmission rate \( f_{ijt} \) of each link \((v_i, v_j)\), the residual energy \( RE(v_i, t) \) of sensor \( v_i \) at the end of each time slot \( t \), and \( RE(v_i, 0) = RE(v_i, 0) \) for each sensor \( v_i \in V \).

Notice that in problem P2, we negative the objective function of problem P1, remove constraint (11) of problem P1, and transform constraint (12) of problem P1 into constraint (17). Notice that we do not explicitly consider the battery capacity \( B_t \), constraint in constraint (17), since this constraint has already been included in constraint (15).

Denote by \( F_1 \) and \( F_2 \) the sets of feasible solutions to problems P1 and P2, respectively. It can be seen that each feasible solution to problem P1 is also a feasible solution to problem P2, since constraint (12) of problem P1 is stronger than constraint (17) of problem P2. Then, the set \( F_1 \) of feasible solutions to problem P1 is a subset of set \( F_2 \), i.e., \( F_1 \subseteq F_2 \). However, a feasible solution to problem P2 may not be feasible to problem P1.

Although the sets \( F_1 \) and \( F_2 \) of feasible solutions to problems P1 and P2 are different, we have the following two important observations.

(i) One observation is that the optimal values of problem P1 and P2 are equal (see Lemma 1 in Section 3.2).

(ii) The other is that, for each feasible solution \((r'_{it}, f'_{ijt}, RE'(v_i, t))\) to problem P2, a feasible solution \((r_{it}, f_{ijt})\) to problem P1 can be constructed by setting \( r_{it} = r'_{it} \) and \( f_{ijt} = f'_{ijt} \), and the objective value of solution \((r_{it}, f_{ijt})\) to problem P1 is equal to the objective value of \((r'_{it}, f'_{ijt}, RE'(v_i, t))\) to P2 (see Lemma 2 in Section 3.2).

Notice that problem P2 can be cast as a convex optimization problem and there are efficient algorithms for finding near-optimal solutions to a convex optimization problem [2], where ‘near-optimal’ means that the objective value of an identified solution is only \( \epsilon \) smaller than the optimal value, where \( \epsilon \) is given additive error with \( 0 < \epsilon < 1 \).

A near-optimal solution to the original problem P1 can be found by leveraging the two important observations as follows.

A near-optimal solution \((r'_{it}, f'_{ijt}, RE'(v_i, t))\) to problem P2 is first found by invoking an algorithm for convex optimization problem [2]. Then, let \( r_{it} = r'_{it} \) and \( f_{ijt} = f'_{ijt} \). Following observation (ii), \((r_{it}, f_{ijt})\) is feasible to problem P1 and its objective value is equal to the objective value of solution \((r'_{it}, f'_{ijt}, RE'(v_i, t))\) to P2. Since the optimal values of problems P1 and P2 are equal and \((r'_{it}, f'_{ijt}, RE'(v_i, t))\) is near-optimal for problem P2, solution \((r_{it}, f_{ijt})\) is also near-optimal to the original problem P1.

The rest is to solve problem P2. We consider its canonical form in terms of the convex optimization as follows.

\[
P_3 : \min_{r_{it}, f_{ijt}, RE'(v_i, t)} \sum_{v_i \in V} U(D_i),
\]
subject to

\[
\begin{align*}
& r_{it} + \sum_{v_j \in N(v_i)} f_{ijt} - \sum_{v_j \in N(v_i)} f_{ijt} = 0, \quad \forall v_i \in V, 1 \leq t \leq T, \\
& \sum_{v_j \in N(s)} f_{ijt} - \sum_{v_j \in N(s)} f_{ijt} - \sum_{v_j \in V} r_{it} = 0, \quad 1 \leq t \leq T, \\
& RE'(v_i, t) - RE'(v_i, t - 1) - \overline{P}(v_i, t) + P(v_i, t) \leq 0, \quad \forall v_i \in V, 1 \leq t \leq T, \\
& -r_{it} \leq 0, \quad r_{it} - R_t \leq 0, \quad \forall v_i \in V, 1 \leq t \leq T, \\
& -f_{ijt} \leq 0, \quad (v_i, v_j) \in E, \quad 1 \leq t \leq T, \\
& -RE(v_i, t) \leq 0, \quad RE(v_i, t) - B_t \leq 0, \quad \forall v_i \in V, 1 \leq t \leq T.
\end{align*}
\]

3.1.2 Solve the convex optimization problem P3

We solve the convex optimization problem P3 by applying the barrier method [2] as follows.

Let \( O_t = RE'(v_i, t - 1) + \overline{P}(v_i, t) - P(v_i, t) - RE'(v_i, t) \), where \( v_i \in V, 1 \leq t \leq T \). The barrier method first introduces a logarithmic barrier function \( \phi(r, f, RE') \) as

\[
\phi(r, f, RE') = -\sum_{v_i \in V} \sum_{t=1}^{T} \log(O_t) - \sum_{v_i \in V} \sum_{t=1}^{T} \log(r_{it}) - \sum_{v_i \in V} \sum_{t=1}^{T} \log(f_{ijt}) - \sum_{v_i \in V} \sum_{t=1}^{T} \log(RE'_{it}) - \sum_{v_i \in V} \sum_{t=1}^{T} \log(B_t - RE'_{it}). \]

Let \( g(r, f, RE') = p \cdot \sum_{v_i \in V} U(D_i) + \phi(r, f, RE') \), where \( p > 0 \) is a parameter and can be considered as a constant (please refer to [2] for the choice of the value of \( p \)). The barrier method then approximates problem P3 with another convex optimization problem P4 with only equality constraints and the accuracy of the approximation is determined by the value of \( p \), where

\[
P_4 : \min \quad g(r, f, RE')
\]
subject to

\[
\begin{align*}
& r_{it} + \sum_{v_j \in N(v_i)} f_{ijt} - \sum_{v_j \in N(v_i)} f_{ijt} = 0, \quad \forall v_i \in V, 1 \leq t \leq T, \\
& \sum_{v_j \in N(s)} f_{ijt} - \sum_{v_j \in N(s)} f_{ijt} - \sum_{v_j \in V} r_{it} = 0, \quad 1 \leq t \leq T, \\
& -RE'(v_i, t) \leq 0, \quad RE'(v_i, t) - B_t \leq 0, \quad \forall v_i \in V, 1 \leq t \leq T.
\end{align*}
\]

Notice that problem P4 can be solved by applying Newton’s method [2].

3.1.3 Construct a strictly feasible solution

The barrier method for the convex optimization problem P3 requires that a strictly feasible solution for problem P3 is given as an input, i.e., a solution \((r_{it}, f_{ijt}, RE'(v_i, t), O_t)\) that satisfies \( 0 < r_{it} < R_t, f_{ijt} > 0, 0 < RE'(v_i, t) < B_t, O_t > 0 \), and the two flow constraints (9) and (10). Then, the barrier method finds the optimal solution to problem P3 in an iterative way, starting from the initial solution \((r_{it}, f_{ijt}, RE'(v_i, t), O_t)\). It is however not easy to construct such a solution. In the following, we propose a novel algorithm to quickly find a strictly feasible solution.

Given a small positive number \( \delta \) (e.g., \( \delta = 0.1 \)), let \( r_{it} = \delta \), where \( v_i \in V, 1 \leq t \leq T \). We construct positive, feasible \( f_{ijt}, RE'(v_i, t), \) and \( O_t \) as follows.

We first construct positive, feasible flows \( f_{ijt} \). We find shortest paths from the base station to the sensors, where the cost of each edge \((v_i, v_j)\) is one. Assume that each sensor sends its data to the base station via its shortest path to the base.
station. Then, we can derive flows $f'_{ijt}$ that satisfy the flow constraints (9) and (10) from the shortest paths. Note that the value of flow $f_{ijt}$ is strictly greater than zero, if edge $(v_i, v_j)$ is contained in the shortest paths and node $v_j$ is the parent of node $v_i$. Otherwise, the value of flow $f_{ijt}$ is 0. We then construct positive, feasible flows $f_{ijt}$ that satisfy constraints (9) and (10) from $f'_{ijt}$ as $f_{ijt} = f'_{ijt}$ if $f'_{ijt} > 0$; otherwise $f_{ijt} = 0$, we increase both flows $f_{ijt}$ and $f_{ijt}$ by $\delta$, i.e., $f_{ijt} = f_{ijt} + \delta$ (as $f'_{ijt} = 0$) and $f_{ijt} = f_{ijt} + \delta$. It can be validated that every flow $f_{ijt}$ is strictly larger than zero and the flows meet constraints (9) and (10), respectively.

We then derive $RE(v_i,t)$ and $O_{it}$ from the data rates $r_{ht}$ and flows $f_{ijt}$. We can calculate the energy consumption $P(v_i,t)$ of each sensor $v_i$ at each time slot $t$ with the data rates $r_{ht}$ and flows $f_{ijt}$, by applying Eq. (5) in Section 2. Then, we let

$$RE(v_i,t) = \min \{RE(v_i,t) + P(v_i,t), B_i\} - \frac{\delta}{T}, \quad \forall v_i \in V, \ 1 \leq t \leq T. \quad (29)$$

Then, we have that

$$O_{it} = RE(v_i,t-1) + \bar{H}(v_i,t) - P(v_i,t) - RE(v_i,t) \geq \frac{RE(v_i,t) + \delta}{T} - RE(v_i,t) = \frac{\delta}{T} > 0. \quad (30)$$

However, the residual energy $RE(v_i,t)$ of some sensor $v_i$ at some time slot $t$ may not be positive, which violates the requirement of the strict feasibility on each $RE(v_i,t)$, though the value of $\delta$ is small. In this case, we can iteratively reduce the value of $\delta$ by a fraction, until the value of each residual energy $RE(v_i,t)$ is strictly greater than zero.

The algorithm for finding a strictly feasible solution is presented in Algorithm 1. Also, the algorithm for the data utility maximization problem is shown in Algorithm 2.

### 3.2 Algorithm analysis

In the following, we analyze the performance of the proposed algorithm. To this end, we first show that the optimal values of problems $P1$ and $P2$ are equal. We then prove that Algorithm 1 can quickly find a strictly feasible solution to problem $P2$. We finally show that Algorithm 2 delivers a near-optimal solution to the data utility maximization problem.

#### Lemma 1. The optimal values of problems $P1$ and $P2$ are equal.

*Proof:* Assume that $(r_{ht}, f_{ijt})$ is an optimal solution to problem $P1$ and $OPT$ is its optimal value. We also assume that $(r'_{ht}, f'_{ijt}, RE(v_i,t))$ is an optimal solution to problem $P2$, and the optimal value is $OPT'$. In the following, we show that (i) $OPT \leq OPT'$ and (ii) $OPT' \leq OPT$. Then, $OPT = OPT'$.

We first prove (i) $OPT \leq OPT'$. Given an optimal solution $(r_{ht}, f_{ijt})$ to problem $P1$, we can calculate the amount of residual energy $RE(v_i,t)$ of each sensor at each time slot $t$ by Eq.(12). Notice that $(r_{ht}, f_{ijt}, RE(v_i,t))$ forms a feasible solution to problem $P2$, as constraint (12) of problem $P1$ is stronger than constraint (17) of problem $P2$. Therefore, $OPT \leq OPT'$.

We then show (ii) $OPT' \leq OPT$. Given an optimal solution $(r'_{ht}, f'_{ijt}, RE(v_i,t))$ to problem $P2$, we can calculate the amount of residual energy $RE(v_i,t)$ in problem $P1$ by Eq.(12) with $r_{ht} = r'_{ht}$ and $f_{ijt} = f'_{ijt}$. We show that $0 \leq RE(v_i,t) \leq B_i$, which means that $(r'_{ht}, f'_{ijt})$ is a feasible solution to problem $P1$. Therefore, $OPT' \leq OPT$.

In the following, we show that $0 \leq RE(v_i,t) \leq B_i$ for every sensor $v_i \in V$ at each time slot $0 \leq t \leq T$, by an induction on $t$.

For $t = 0$, it is clear that $0 \leq RE(v_i,0) \leq RE(v_i,0) \leq B_i$, since $RE(v_i,0) = RE(v_i,0)$ and $0 \leq RE(v_i,0) \leq B_i$.

Assume that $0 \leq RE(v_i,t') \leq RE(v_i,t') \leq B_i$ for $t' = 0, 1, \ldots, t$, where $0 \leq t \leq T$. For the next time slot $t+1$, we distinguish into two cases for constraint (12) of problem $P1$.

#### Algorithm 1 find a strictly feasible solution to problem $P3$

**Input:** an energy harvesting sensor network $G = (V \cup \{s\}, E)$, $T$ time slots, the initial residual energy $RE(v_i,0)$ of each sensor $v_i$, the amount of harvested energy $\bar{H}(v_i,t)$ of sensor $v_i$ at each time slot $t$ in the future $T$ time slots.

**Output:** a strictly feasible solution $(r_{ht}, f_{ijt}, RE'(v_i,t), O_{it})$ to problem $P3$

1. Find shortest paths from the base station $s$ to sensors in $G$, where the cost of each edge is one;
2. $\delta \leftarrow 0.1$; /* an initial small value */
3. $flag \leftarrow false$;
4. while flag is equal to false do
5. $r_{it} \leftarrow \delta$, $\forall v_i \in V, 1 \leq t \leq T$;
6. Find flows $f_{ijt}$ from the shortest paths, assuming that each sensor $v_i$ sends its data $r_{it}$ to the base station $s$ via its shortest path;
7. for each flow $f_{ijt}$ do
8. if $f_{ijt} = 0$ then
9. $f_{ijt} \leftarrow \delta$;
10. $f_{ijt} \leftarrow f_{ijt} + \delta$;
11. end if
12. end for
13. Let $f_{ij} \leftarrow f_{ij}$ with $(v_i, v_j) \in E$ and $1 \leq t \leq T$;
14. Calculate the residual energy $RE'(v_i,t)$ of each sensor $v_i$ at each time slot $t$, by applying Eq. (29); 
15. Calculate the value of $O_{it}$ by applying Eq. (30), $\forall v_i \in V, 1 \leq t \leq T$;
16. if the value of each residual energy $RE'(v_i,t)$ is strictly larger than zero then
17. $flag \leftarrow true$;
18. else
19. $\delta \leftarrow \frac{\delta}{2}$, /* reduce the value of $\delta$ by a half */
20. end if
21. end while
22. return $(r_{it}, f_{ijt}, RE'(v_i,t), O_{it})$.

#### Algorithm 2 maxUtility

**Input:** an energy harvesting sensor network $G = (V \cup \{s\}, E)$, $T$ time slots, the initial residual energy $RE(v_i,0)$ of each sensor $v_i$, the amount of harvested energy $\bar{H}(v_i,t)$ of sensor $v_i$ at each time slot $t$ in the future $T$ time slots.

**Output:** an optimal data rate $r_{it}$ for each sensor $v_i \in V$ and an optimal transmission rate $f_{ijt}$ for each link $(v_i, v_j) \in E$ in period $T$, such that utility sum of collected data during $T$ is maximized

1. Transform the original data utility maximization problem $P1$ into another convex optimization problem $P2$; and obtain its canonical form $P3$;
2. Construct a strictly feasible solution $(r_{ht}, f_{ijt}, RE'(v_i,t), O_{it})$ to problem $P3$, by invoking Algorithm 1;
3. Obtain an optimal data rate $r'_{it}$ for each sensor $v_i \in V$ and an optimal transmission rate $f'_{ijt}$ and residual energy $RE'(v_i,t)$ for problem $P3$ by applying the barrier method in [2];
4. Let $r_{it} = r'_{it}$ for $v_i \in V$ and $1 \leq t \leq T$ and $f_{ijt} = f'_{ijt}$ for $(v_i, v_j) \in E$ and $1 \leq t \leq T$;
5. return the data rate $r_{it}$ and the transmission rate $f_{ijt}$.
Case 1: $B_t \leq RE(v_t, t) + \overline{P}(v_t, t + 1) - P(v_t, t + 1); \text{ and Case 2:}$

$$RE(v_t, t) + \overline{P}(v_t, t + 1) - P(v_t, t + 1) < B_t.$$ 

For Case 1, we have

$$RE(v_t, t + 1) = \min \{RE(v_t, t) + \overline{P}(v_t, t + 1) - P(v_t, t + 1), B_t\}$$

$$B_t \geq RE(v_t, t + 1) \geq 0.$$ (31)

That is, $0 \leq RE(v_t, t + 1) \leq RE(v_t, t + 1) \leq B_t.$

On the other hand, for Case 2, we have

$$RE(v_t, t + 1) = \min \{RE(v_t, t) + \overline{P}(v_t, t + 1) - P(v_t, t + 1), B_t\}$$

$$= RE'(v_t, t + 1) + \overline{P}(v_t, t + 1) - P(v_t, t + 1);$$

$$\geq RE'(v_t, t + 1) \geq 0,$$ by Eq. (17). (32)

Then, we still have $0 \leq RE'(v_t, t + 1) \leq RE(v_t, t + 1) \leq B_t.$

Therefore, the inequality $0 \leq RE'(v_t, t + 1) \leq RE(v_t, t + 1) \leq B_t$ holds for each sensor $v_t \in V$ at each time slot $0 \leq t \leq T.$ Thus, $(r'_t(t), f'_jt)$ is a feasible solution to problem P1, and $OPT' \leq OPT.$ Recall that $OPT \leq OPT',$ we know $OPT' \leq OPT,'$ which means that the optimal values for problems P1 and P2 are equal. The lemma then follows.

**Lemma 2.** Given a feasible solution $(r'_t(t), f'_jt, RE'(v_t, t))$ to problem P2, let $r_a = r'_t$ and $f_{jt} = f'_jt.$ Then, $(r_t, f_{jt})$ forms a feasible solution to problem P1, and its objective value in P1 is equal to the objective value of solution $(r'_t(t), f'_jt, RE'(v_t, t))$ to problem P2.

**Proof:** The proof is similar to that in Lemma 1, omitted.

**Lemma 3.** Given an energy harvesting sensor network $G = (V \cup \{s\}, E)$ and $T$ time slots, Algorithm 1 delivers a strictly feasible solution to problem P3 in time $O((m + n)T)log(\frac{n}{\epsilon}))$, where $n = |V|$ and $m = |E|.$

**Proof:** We first show that Algorithm 1 terminates after $O(\log(nT))$ iterations of the while loop. It can be seen that both the sums of incoming flows $\sum_{v_t \in V(n)} f_{jt}$ and outcoming flows $\sum_{v_t \in V(n)} f_{jt}$ for each sensor $v_t$ at each time slot $t$ are only $O(\delta n).$ Then, the total energy consumption of sensor $v_t$ in period $T$ is $O(\delta nT).$ Assume that Algorithm 1 terminates after $k$ iterations, which means that $\frac{1}{2^k} < RE(v_t, 0),$ i.e., $k = O(\log(nT)).$ On the other hand, in each iteration, Algorithm 1 finds a solution to problem P3 in time $O(m + n) + O(nT) = O(m + nT).$ Then, the running time of Algorithm 1 is $O((m + nT)\log(\frac{n}{\epsilon})).$

**Theorem 1.** Given an energy harvesting sensor network $G = (V \cup \{s\}, E)$, $T$ time slots, the initial residual energy $RE(v_t, 0)$ of each sensor $v_t$, the amount of harvested energy $\overline{P}(v_t, t)$ of sensor $v_t$ at each time slot $t$ in the future $T$ time slots, Algorithm 2 delivers a solution to the data utility maximization problem in $G$, such that the total utility of the solution is only $\epsilon$ smaller than the maximum utility, where $\epsilon$ is a given additive error with $0 < \epsilon < 1.$

**Proof:** Following Lemma 1, the data utility maximization problem P1 is equivalent to problem P2, where the latter is equivalent to problem P3. Also, Lemma 3 shows that Algorithm 1 delivers a strictly feasible solution to problem P3. Following the work [2], the barrier method in Algorithm 2 can find a solution with its total utility only $\epsilon$ smaller than the maximum utility, where $\epsilon$ is a given additive error with $0 < \epsilon < 1.$ The theorem then follows.

### 4.2 Algorithm performance

We first study the performance of algorithms maxUtility, maxThroughput, utilityTimeSlot, utilityTimeSlot+, utilityCorrelation and...
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2901758, IEEE Internet of Things Journal

Fig. 3. The performance of different algorithms during a week with 100 sensors

Fig. 4. The performance of different algorithms by varying the network size

utilityCorrelation+ within one week in a network with 100 sensors. Fig. 3(a) illustrates the distribution of average throughput per sensor in the 7th day with the distance between a sensor and the base station, from which it can be seen that, in algorithm maxUtility, the sensors close to the base station upload much more data than the sensors far from the base station. For example, the average throughput for sensors with 300 meters away from the base station is only 0.6 Mb. In contrast, the average throughput per sensor by the proposed algorithm maxUtility is larger than that by the other five algorithms when the sensor distance to the base station is longer than 200 meters. For instance, the average throughput per sensor by algorithms maxUtility, maxThroughput, utilityTimeSlot, utilityTimeSlot+, utilityCorrelation and utilityCorrelation+ are 4.7 Mb, 0.08 Mb, 2.5 Mb, 0.6 Mb, 0.7 Mb, and 2.9 Mb, respectively, when the sensor distance to the base station is 500 meters. Therefore, Fig. 3(a) shows that the sensing data from different sensors is more fairly collected in algorithm maxUtility than that in the other five algorithms.

Fig. 3(b) plots the utility sums of collected data by different algorithms during the week. It can be seen that the utility sum by algorithm maxUtility is the largest one among the six algorithms, while the one by algorithm maxThroughput is the smallest, as it does not consider data correlation. Also, the utility sum by algorithm maxUtility is about from 2.5% to 10% higher than that by the second best algorithm utilityCorrelation+.

Fig. 3(c) compares the running times of different algorithms. It can be seen that the running times by the three algorithms maxThroughput, utilityTimeSlot, utilityTimeSlot+ are much shorter than the other three algorithms maxUtility, utilityCorrelation and utilityCorrelation+, since the former three algorithms consider the allocation of data rates and routing one time slot by one time slot, but the latter three take the allocation for a period with multiple time slots into account. In spite of it, the running time of algorithm maxUtility is about 25% shorter than that of algorithms utilityCorrelation and utilityCorrelation+, which is about only 2.1 seconds.

We then investigate the performance of the different algorithms by varying the network size from 50 sensors to 200 sensors. Fig. 4(a) shows that the average throughput per sensor by algorithm maxUtility is the largest one among the six algorithms when the distance to the base station is longer than 200 meters, and its throughput is even larger than 3.5 Mb, while the throughput by the other five algorithms are no more than 1.6 Mb, when the distance is 700 meters. Fig. 4(b) demonstrates the utility sums by different algorithms. We can see that the utility sum by each of the six algorithms increases with the increase of the network size, and the utility sum by algorithm maxUtility is about from 0.8% to 10% larger than that by the other five algorithms. Fig. 4(c) illustrates that the running time of each algorithm increases for a larger network.

We also evaluate the impact of amounts of harvested energy on the data utility, by increasing the energy scaling efficient λ from 0.1 to 1 in a network with 100 sensors. Recall that we adopted real solar energy profiles in summer and the energy harvesting rate $H(v_i, t)$ of each sensor $v_i$ at time slot $t$ in our simulation is $H(v_i, t) = \lambda \cdot H_{sample}(t)$, where $H_{sample}(t)$ is the real energy harvesting rate of a randomly chosen energy profile at time slot $t$. We thus can consider energy harvesting profiles in other seasons by varying the value of $\lambda$. Fig. 5(a) shows the average throughput per sensor by algorithm maxUtility is the largest one among the six algorithms when the distance is longer than 200 meters. On the other hand, in algorithm maxThroughput, only sensors close to the base station upload sensing data, while the sensors with 300 meters away barely send their data to the base station. Fig. 5(b) plots that the utility sum by algorithm maxUtility is about from 4.5% to 11.5% higher than that by the second best algorithm utilityCorrelation+. Fig. 5(c) demonstrates the running times of different algorithms almost do not change with the increase of $\lambda$. 

2327-4662 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
We finally study the impact of the maximum data rate $R$ on the data utility, by increasing $R$ from 50 bps to 1,000 bps in a sensor network with 200 sensors. It can be seen from Fig. 6(a) that the average throughput by the proposed algorithm $\text{maxUtility}$ is the highest for the sensors at least 200 meters away from the base station, while the throughput by algorithm $\text{maxThroughput}$ approaches to zero, when the distance is longer than 300 meters, since it does not take the data correlation into consideration. Fig. 6(b) shows that the sum of data utility by each of the algorithms increases with the growth of the maximum data rate $R$. However, the data utility by each algorithm only slightly grows when the maximum data rate $R$ is larger than 600 bps. The rationale behind is that, although each sensor is allowed to sense at a faster rate, the energy harvested by sensors cannot support the data transmission for such a high data rate. Also, Fig. 6(b) demonstrates that the sum of data utility by algorithm $\text{maxUtility}$ is about 4.7% higher than that by the second best algorithm $\text{utilityCorrelation+}$. Fig. 6(c) plots that the running time of algorithm $\text{maxUtility}$ is only about 2.2 seconds.

5 Related Work

With development of energy harvesting technology, more and more research is concentrated on energy harvesting sensor networks.

A lot of attentions have been paid to the data collection in energy-harvesting sensor networks. Most of the works focus on the data throughput maximization problem. For example, Mao et al. [21] studied the problem of allocating energy for data sensing and data transmitting, such that the amount of collected data is maximized. They first formulated the problem as an infinite horizon Markov decision process with finite data buffer, and solved the problem by the value iteration algorithm. They also considered the case with infinite data buffer, and proposed a near-optimal algorithm. Manfredi et al. [20] studied the problem of controlling sensor transmission radii so as to prolong the network lifetime, by increasing the transmission radii of sensors with high energy availability, while decreasing the radii of sensors with low energy availability. Li et al. [14] aimed at maximizing the amount of data collected from sensors by scheduling the transmission per time slot according to the harvested sensor energy and link quality. They proposed a heuristic algorithm for the problem with linear running time. Shafieeirad et al. [26] also aimed at maximizing the total data throughput in large-scale energy harvesting sensor network. They proposed a routing algorithm which prioritizes the relay sensors with sufficient energy and short distances to the base station. The algorithm however ignored the energy consumption on data receiving and did not considered the limited battery capacity constraint on each sensor node. Furthermore, some studies investigated the energy-away routing problem, by finding routing paths with the smallest energy consumption [4], [9], [27].

There are several studies on the data utility maximization problem in energy harvesting sensor networks. For instance, Dong et al. [7] proposed a cluster-based routing protocol and selected cluster head nodes by considering harvested sensor energy as well as the distance between each sensor and the base station. Liu et al. [18] studied the data utility maximization problem and jointly optimized data sensing and routing, provided the harvested energy in sensors. Chen et al. [3] also investigated the utility maximization problem, but the energy budget of each sensor at each time slot is equal to the minimum value among its battery residual energy and its past average harvested energy per time slot. Lu et al. [19] proposed a distributed algorithm to allocate data rates for all sensors according to the amount of harvested energy. Deng et al. [6] investigated the problem of allocating a sensing rate for each sensor, such that the data utility is maximized and the amount of the energy consumed by each sensor is no more than its residual energy, assuming that routing paths are fixed. They later extended their work in [6] to the case where routing paths are allowed to change over time [5]. Zhang et al. [39] considered that the sensors having larger amounts of residual energy and higher energy harvesting rates are able to relay...
more data for other sensors, as the amount of harvested energy in such a sensor cannot exceed its battery capacity. Yang et al. [34] studied the problem of selecting a subset of sensors to be active and perform sensing tasks so that the average data utility per time slot is maximized, while satisfying the energy consumption constraints on each sensor and each time slot. They first relaxed the constraints by assuming that the battery capacity of each sensor is infinite and the energy harvested by each time slot is given. They also assumed that the energy harvesting processes at each sensor is independent. They then proposed an algorithm to maximize the average data utility per time slot and proved that the proposed algorithm is optimal. Yang et al. [33] incorporated the data aggregation and data compression with sensing and routing in energy harvesting sensor network, so that the network data utility is maximized while ensuring that no sensors run out of their energy. They then formulated the optimization problem as a network utility maximization problem, and designed an algorithm by applying the Lyapunov optimization framework, which is shown to achieve the asymptotical optimality with limited data buffer and sustainable networking operations.

We also note that some other studies employed a mobile sink to travel along a pre-defined path to collect sensor data in an energy harvesting sensor network. For example, studies [23], [25] considered the problem of collecting as much data as possible during the period that the sink travels along the path. Zhang et al. [40] considered the data utility maximization problem where sensors can only consume the energy harvested during last period. They proposed a distributed near-optimal solution to the data gathering. Wang et al. [29] also employed a mobile sink but they developed an anchor selection algorithm for the mobile sink to stop for data collection. In order to maximize the data utility, they proposed a distributed algorithm to optimize the data rates, data routing for each sensor and sojourn time at each anchor location.

It can be seen that some of the aforementioned studies considered only making use of the energy that has already been harvested but ignored the future harvested energy. On the other hand, although the other studies jointly considered the current residual sensor energy and future to-be-harvested energy, they usually assumed that the data from each sensor is temporally independent. Unlike the existing studies, in this paper we consider the temporal correlation of data from each sensor, and allocate data rates, data routing, and sensor energy for a given period, such that the utility of collected temporally-correlated data in the period is maximized. Therefore, our work is promising to improve the data collection performance.

6 Conclusion

Unlike the existing studies assumed that the data from each sensor is temporally independent, in this paper we studied the problem of allocating sensor energy, sensor data rates, and data routing in an energy harvesting sensor network for a given monitoring period $T$, such that the utility sum of temporally-correlated sensing data collected during the period $T$ is maximized. We then proposed a near-optimal algorithm for the data utility maximization problem, by reducing the problem to another equivalent optimization problem, and devising an efficient algorithm for the latter. A feasible solution to the latter in turn returns a feasible solution to the formal with guaranteed approximation ratio. We finally evaluated the performance of the proposed algorithm with extensive simulation experiments, using real solar energy data. Experimental results showed that the sensing data from different sensors by the proposed algorithm is more fairly collected and the utility sum by the algorithm is up to 10% larger than that by the state-of-the-art.

ACKNOWLEDGEMENT

The work by Wenzheng Xu was supported by the National Natural Science Foundation of China (NSFC) with grant number 61602330, Sichuan Science and Technology Program (Grant No. 2018GZ0094, 2018GZDDX0010, 2018GZ0093, 2017GZDDX0003), and the Fundamental Research Funds for the Central Universities (Grant No. 20822041B1040). Also, the work by Zheng Li was supported by the National Natural Science Foundation of China (Grant No. 61471250).

REFERENCES

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2901758, IEEE Internet of Things Journal


