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 usually 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*—Data utility maximization, energy harvesting sensor networks, temporally correlated sensing data.**

I. Introduction

Wireless sensor networks (WSNs) have wide applications in Internet of Things, 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 2:00 P.M. 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]–[7], [17], [19], [37]–[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]–[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

Manuscript received January 24, 2019; accepted February 22, 2019. Date of publication February 26, 2019; date of current version June 19, 2019. The work of W. Xu was supported in part by the National Natural Science Foundation of China under Grant 61602330, in part by the Sichuan Science and Technology Program under Grant 2018GZ0094, Grant 2018GZDZX0010, Grant 2018GZ0093, and Grant 2017GZDZX0003, and in part by the Fundamental Research Funds for the Central Universities under Grant 20822041B4104. The work of Z. Li was supported by the National Natural Science Foundation of China under Grant 61471250. *(Rongrong Zhang and Jian Peng contributed equally to this work.) (Corresponding author: Wenzheng Xu.)*

R. Zhang, J. Peng, W. Xu, and Z. Li are with the College of Computer Science, Sichuan University, Chengdu 610065, 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, Australian National University, Canberra, ACT 2601, Australia (e-mail: wliang@cs.anu.edu.au).

T. Wang is with the College of Computer Science and Technology, Huaqiao University, Xiamen 361000, China (e-mail: cs_tianwang@163.com).

Digital Object Identifier 10.1109/JIOT.2019.2901758

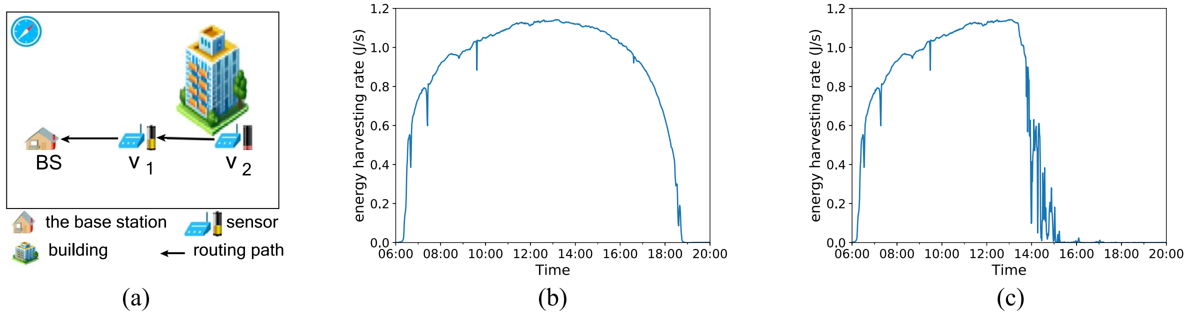


Fig. 1. Energy harvesting profiles in a renewable sensor network. (a) Sensor network. (b) Energy harvesting rates of sensor v_1 at different time points. (c) Energy harvesting rate of sensor v_2 at different time points.

function to measure the quality of the collected data D_{it} from a sensor v_i at time slot t.

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 kb/s, while the residual energy of v_2 can support its data transmission for only one time slot with the rate of 1 kb/s. 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 nonredundant information.

Notice that it is challenging to apply existing algorithms in [3], [5]–[7], [19], and [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.

1) 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.
2) 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 suboptimal solution to the former problem.

The main contributions in this paper are as follows.

1) 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.
2) We then propose a near-optimal algorithm for the data utility maximization problem.
3) 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 II describes the network model and formulates the problem. Section III devises a novel algorithm for the data utility maximization problem. Section IV evaluates the performance of the proposed algorithm. Section V further reviews the related work. Finally, Section VI concludes this paper.

II. Preliminaries

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

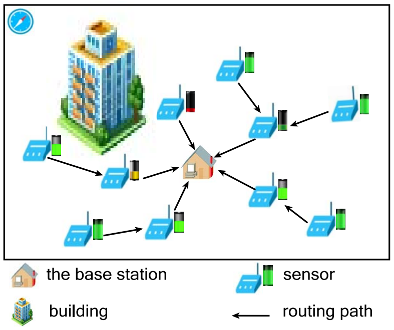


Fig. 2. Energy harvesting sensor network.

A. Network Model

We consider an energy harvesting sensor network $G = (V \cup \{s\}, E)$, where V is a set of n sensors $v_1, v_2, \ldots, v_n$ 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 period of T (e.g., 24 h). We divide the period T into consecutive time slots, which are indexed as $1, 2, \ldots, T$, and the duration of each time slot lasts for τ time units (e.g., 1 hr).

We assume that the maximum sensing rate of each sensor $v_i \in V$ is R_i (e.g., 10 kb/s), where the value of R_i can be obtained from the specification of the sensor. Since the energy in each sensor v_i is limited, it may not have enough energy to sense data at the maximum rate R_i. We thus assume that each sensor v_i is allowed to generate data at a rate of r_{it} at time slot t, and sends the data back to the base station directly, or via the relay of other sensors, where the value of r_{it} will be determined with $0 \le r_{it} \le R_i$.

Sensors transmit their sensing data to the base station in a multihop manner. That is, a sensor close to the base station can relay data from other sensors far from the base station. Let $N(v_i)$ be the set of neighboring sensors of a sensor v_i, i.e., $N(v_i) = \{v_j | (v_i, v_j) \in E\}$. For each neighboring sensor $v_j \in N(v_i)$, denote by f_{ijt} and f_{jit} the data transmission rates from sensor v_i to sensor v_j, and from v_j to v_i at time t, respectively. We assume that the amount $\sum_{v_j \in N(v_i)} f_{jit}$ of received data by sensor v_i plus the amount r_{it} of sensing data from sensor v_i itself at time slot t will be sent to the base station. That is,

$$\begin{aligned} \sum_{v_j \in N(v_i)} f_{jit} + r_{it} &= \sum_{v_j \in N(v_i)} f_{ijt} \\ \sum_{v_i \in N(s)} f_{ist} - \sum_{v_i \in N(s)} f_{sit} &= \sum_{v_i \in V} r_{it} \end{aligned} \tag{1}$$

where $\sum_{v_j \in N(v_i)} f_{ijt}$ is the amount of data sent to the base station from sensor v_i via the relay of its neighboring sensors, and constraint (1) ensures that the base station s receives the sensing data from all sensors.

B. Energy Consumption Model

Every sensor consumes its energy on data sensing, data reception, and data transmission. Let $P_s(v_i, t)$, $P_{rx}(v_i, t)$, and $P_{tx}(v_i, t)$ 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

$$P_s(v_i, t) = r_{it} \times \mu_s \times \tau \tag{2}$$

$$P_{rx}(v_i, t) = \sum_{v_j \in N(v_i)} f_{jit} \times \mu_{rx} \times \tau \tag{3}$$

$$P_{tx}(v_i, t) = \sum_{v_j \in N(v_i)} f_{ijt} \times \left(\beta_1 + \beta_2 d_{ij}^{\alpha}\right) \times \tau \tag{4}$$

μ_s and μ_{rx} are the energy consumption rates for sensing and receiving per unit data, respectively, τ is the duration of each time slot. $\sum_{v_j \in N(v_i)} f_{jit}$ and $\sum_{v_j \in N(v_i)} f_{ijt}$ are the data reception and transmission rates of sensor v_i at time t. Following [13], the term α 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/b/m^2 if $\alpha = 2$; or $\beta_2 = 0.001 \times 10^{-12}$ J/b/m^4 if $\alpha = 4$.

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

$$\begin{aligned} P(v_i, t) &= P_s(v_i, t) + P_{rx}(v_i, t) + P_{tx}(v_i, t) \\ &= a \cdot r_{it} + b \cdot \sum_{v_j \in N(v_i)} f_{jit} + \sum_{v_j \in N(v_i)} c_{ij} \cdot f_{ijt} \end{aligned} \tag{5}$$

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

C. 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 h). 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 $\overline{H}(v_i, d-1, t)$ 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:

$$\overline{H}(v_i, d, t) = \big(\omega \cdot \overline{H}(v_i, d-1, t) + (1-\omega) \times H(v_i, d-1, t)\big) \cdot \gamma(d, t) \quad (6)$$

where ω 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

$$RE(v_i, t+1) = \min\left\{RE(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1),\ B_i\right\} \quad (7)$$

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 $t+1$.

D. 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., 1 h) 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 $D_i = \sum_{t=1}^{T} r_{it}\tau$. 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 $\overline{H}(v_i, t)$ 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.,

$$\text{P1: } \max_{r_{it}, f_{ijt}} \sum_{v_i \in V} U\left(\sum_{t=1}^{T} r_{it}\tau\right) \quad (8)$$

subject to

$$r_{it} + \sum_{v_j \in N(v_i)} f_{jit} = \sum_{v_j \in N(v_i)} f_{ijt} \quad \forall v_i \in V,\ 1 \leq t \leq T \quad (9)$$

$$\sum_{v_i \in N(s)} f_{ist} - \sum_{v_i \in N(s)} f_{sit} = \sum_{v_i \in V} r_{it}, \quad 1 \leq t \leq T \quad (10)$$

$$P(v_i, t) \leq RE(v_i, t-1) + \overline{H}(v_i, t) \quad \forall v_i \in V, \quad 1 \leq t \leq T \quad (11)$$

$$RE(v_i, t) = \min\big\{RE(v_i, t-1) + \overline{H}(v_i, t) - P(v_i, t), B_i\big\} \quad \forall v_i \in V,\ 1 \leq t \leq T \quad (12)$$

$$0 \leq r_{it} \leq R_i \quad \forall v_i \in V,\ 1 \leq t \leq T \quad (13)$$

$$f_{ijt} \geq 0 \quad \forall\big(v_i, v_j\big) \in E,\ 1 \leq t \leq T \quad (14)$$

$$0 \leq RE(v_i, t) \leq B_i \quad \forall v_i \in V,\ 1 \leq t \leq T \quad (15)$$

where τ 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.

E. 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 concept into WSNs, 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.

III. Algorithm for the Data Utility Maximization Problem

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

A. Algorithm

The basic idea of the proposed algorithm is to first transform the data utility maximization problem P1 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 suboptimal solution to the original problem.

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 $\overline{H}(v_i, t)$ 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

$$\text{P2:} \quad \min_{r_{it}, f_{ijt}, RE'(v_i,t)} \quad -\sum_{v_i \in V} U(D_i) \tag{16}$$

subject to constraints (9), (10), (13)–(15), and

$$RE'(v_i, t) \le RE'(v_i, t-1) + \overline{H}(v_i, t) - P(v_i, t)$$
$$\forall v_i \in V, \ 1 \le t \le T \tag{17}$$

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_i 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.

1) One observation is that the optimal values of problem P1 and P2 are equal (see Lemma 1 in Section III-B).
2) The other is that, for each feasible solution $(r'_{i,t}, f'_{ijt}, RE'(v_i, t))$ to problem P2, a feasible solution $(r_{i,t}, f_{ijt})$ to problem P1 can be constructed by setting $r_{i,t} = r'_{i,t}$ and $f_{ijt} = f'_{ijt}$, and the objective value of solution $(r_{i,t}, f_{ijt})$ to problem P1 is equal to the objective value of $(r'_{i,t}, f'_{ijt}, RE'(v_i, t))$ to P2 (see Lemma 2 in Section III-B).

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 ϵ smaller than the optimal value, where ϵ is a 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 $(r'_{i,t}, f'_{ijt}, RE'(v_i, t))$ to problem P2 is first found by invoking an algorithm for convex optimization problem [2]. Then, let $r_{i,t} = r'_{i,t}$ and $f_{ijt} = f'_{ijt}$. Following observation 2), $(r_{i,t}, f_{ijt})$ is feasible to problem P1 and its objective value is equal to the objective value of solution $(r'_{i,t}, f'_{ijt}, RE'(v_i, t))$ to P2. Since the optimal values of problems P1 and P2 are equal and $(r'_{i,t}, f'_{ijt}, RE'(v_i, t))$ is near-optimal for problem P2, solution $(r_{i,t}, f_{ijt})$ 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:

$$\text{P3:} \quad \min_{r_{it}, f_{ijt}, RE'(v_i,t)} \quad -\sum_{v_i \in V} U(D_i) \tag{18}$$

subject to

$$r_{it} + \sum_{v_j \in N(v_i)} f_{jit} - \sum_{v_j \in N(v_i)} f_{ijt} = 0 \quad \forall v_i \in V, \ 1 \le t \le T \tag{19}$$

$$\sum_{v_i \in N(s)} f_{ist} - \sum_{v_i \in N(s)} f_{sit} - \sum_{v_i \in V} r_{it} = 0, \quad 1 \le t \le T \tag{20}$$

$$RE'(v_i, t) - RE'(v_i, t-1) - \overline{H}(v_i, t) + P(v_i, t) \le 0$$
$$\forall v_i \in V, \ 1 \le t \le T \tag{21}$$

$$-r_{it} \le 0, \quad r_{it} - R_i \le 0 \quad \forall v_i \in V, \ 1 \le t \le T \tag{22}$$

$$-f_{ijt} \le 0 \quad \forall (v_i, v_j) \in E, \ 1 \le t \le T \tag{23}$$

$$-RE(v_i, t) \le 0, \quad RE(v_i, t) - B_i \le 0 \quad \forall v_i \in V, \ 1 \le t \le T. \tag{24}$$

2) Solve the Convex Optimization Problem P3: We solve the convex optimization problem P3 by applying the barrier method [2] as follows.

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

$$\begin{aligned}\phi\big(\mathbf{r}, \mathbf{f}, \mathbf{RE}'\big) = &-\sum_{v_i\in V}\sum_{t=1}^{T}\log(O_{it}) - \sum_{v_i\in V}\sum_{t=1}^{T}\log(r_{it})\\ &-\sum_{v_i\in V}\sum_{t=1}^{T}\log(R_i - r_{it}) - \sum_{(v_i,v_j)\in E}\sum_{t=1}^{T}\log\big(f_{ijt}\big)\\ &-\sum_{v_i\in V}\sum_{t=1}^{T}\log\big(RE'_{it}\big) - \sum_{v_i\in V}\sum_{t=1}^{T}\log\big(B_i - RE'_{it}\big).\end{aligned} \tag{25}$$

Let $g(\mathbf{r, f, RE'}) = -p\cdot\sum_{v_i\in V}U(D_i) + \phi(\mathbf{r, f, RE'})$, 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

$$\text{P4: min} \quad g\big(\mathbf{r}, \mathbf{f}, \mathbf{RE}'\big) \tag{26}$$

subject to

$$r_{it} + \sum_{v_j\in N(v_i)} f_{jit} - \sum_{v_j\in N(v_i)} f_{ijt} = 0 \quad \forall v_i \in V,\ 1 \le t \le T \tag{27}$$

$$\sum_{v_i\in N(s)} f_{ist} - \sum_{v_i\in N(s)} f_{sit} - \sum_{v_i\in V} r_{it} = 0, \quad 1 \le t \le T. \tag{28}$$

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

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 $(r_{i,t}, f_{ijt}, RE'(v_i, t), O_{it})$ that satisfies $0 < r_{it} < R_i$, $f_{ijt} > 0$, $0 < RE'(v_i, t) < B_i$, $O_{it} > 0$, 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 $(r_{i,t}, f_{ijt}, RE'(v_i, t), O_{it})$. 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 δ (e.g., $\delta = 0.1$), let $r_{it} = \delta$, where $v_i \in V, 1 \le t \le T$. We construct positive, feasible $f_{ijt}, RE'(v_i, t)$, and O_{it} 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}s that satisfy constraints (9) and (10) from f'_{ijt}s as $f_{ijt} = f'_{ijt}$ if $f'_{ijt} > 0$; otherwise ($f'_{ijt} = 0$), we increase both flows f'_{ijt} and f'_{jit} by δ, i.e., $f_{ijt} = f'_{ijt} + \delta = \delta$ (as $f'_{ijt} = 0$) and $f_{jit} = f'_{jit} + \delta$. 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_{it}s and flows f_{ijt}s. 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_{it}s and flows f_{ijt}s, by applying (5) in Section II-B. Then, we let

$$RE'(v_i, t) = \min\big\{RE'(v_i, t-1) + \overline{H}(v_i, t) - P(v_i, t), B_i\big\} - \frac{\delta}{T} \quad \forall v_i \in V, \quad 1 \le t \le T. \tag{29}$$

Then, we have that

$$\begin{aligned} O_{it} &= RE'(v_i, t-1) + \overline{H}(v_i, t) - P(v_i, t) - RE'(v_i, t)\\ &\ge RE'(v_i, t) + \frac{\delta}{T} - RE'(v_i, t)\\ &= \frac{\delta}{T}\\ &> 0. \end{aligned} \tag{30}$$

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 δ is small. In this case, we can iteratively reduce the value of δ 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.

B. 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. We start with the following lemma.

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

Proof: Assume that (r_{it}, f_{ijt}) is an optimal solution to problem P1 and OPT is its optimal value. We also assume that $(r'_i(t), f'_{ijt}, RE'(v_i, t))$ is an optimal solution to problem P2, and the optimal value is OPT'. In the following, we show that: 1) $OPT \le OPT'$ and 2) $OPT' \le OPT$. Then, $OPT = OPT'$.

We first prove 1) $OPT \le OPT'$. Given an optimal solution (r_{it}, 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 (12). Notice that $(r_{it}, f_{ijt}, RE(v_i, t))$ forms a feasible solution to problem P2, as constraint (12) of problem P1 is stronger than constraint (17) of problem P2. Therefore, $OPT \le OPT'$.

We then show 2) $OPT' \le OPT$. Given an optimal solution $\big(r'_i(t), f'_{ijt}, RE'(v_i, t)\big)$ to problem P2, we can calculate the amount of residual energy $RE(v_i, t)$ in problem P1 by (12) with $r_{it} = r'_i(t)$ and $f_{ijt} = f'_{ijt}$. We show that $0 \le RE(v_i, t) \le B_i$, which means that $\big(r'_i(t), f'_{ijt}\big)$ is a feasible solution to problem P1. Therefore, $OPT' \le OPT$.

In the following, we show that $0 \le RE'(v_i, t) \le RE(v_i, t) \le B_i$ for every sensor $v_i \in V$ at each time slot $0 \le t \le T$, by an induction on t.

For $t = 0$, it is clear that $0 \le RE'(v_i, 0) \le RE(v_i, 0) \le B_i$, since $RE'(v_i, 0) = RE(v_i, 0)$ and $0 \le RE'(v_i, 0) \le B_i$.

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 $\overline{H}(v_i, t)$ of sensor v_i at each time slot t in the future T time slots.

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

Find shortest paths from the base station s to sensors in G, where the cost of each edge is one;
$\delta \leftarrow 0.1$; /* an initial small value */
flag $\leftarrow$ *false*;
while *flag* is equal to *false* **do**
  $r_{it} \leftarrow \delta$, $\forall v_i \in V, 1 \le t \le T$;
  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;
  for each flow f'_{ijt} **do**
    if $f'_{ijt} = 0$ **then**
      $f'_{ijt} \leftarrow \delta$;
      $f'_{jit} \leftarrow f'_{jit} + \delta$;
    end if
  end for
  Let $f_{ijt} \leftarrow f'_{ijt}$ with $(v_i, v_j) \in E$ and $1 \le t \le T$;
  Calculate the residual energy $RE'(v_i, t)$ of each sensor v_i at each time slot t, by applying Eq. (29);
  Calculate the value of O_{it} by applying Eq. (30), $\forall v_i \in V, 1 \le t \le T$;
  if the value of each residual energy $RE'(v_i, t)$ is strictly larger than zero **then**
    flag $\leftarrow$ *true*;
  else
    $\delta \leftarrow \frac{\delta}{2}$, /* reduce the value of δ by a half */
  end if
end while
return $(r_{it}, f_{ijt}, RE'(v_i, t), O_{it})$.

Assume that $0 \le RE'(v_i, t') \le RE(v_i, t') \le B_i$ for $t' = 0, 1, \ldots, t$, where $0 \le t \le T$. For the next time slot $t+1$, we distinguish into two cases for constraint (12) of problem P1. *Case 1*: $B_i \le RE(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1)$ and *Case 2*: $RE(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1) < B_i$.

For case 1, we have

$$\begin{aligned} &RE(v_i, t+1) \\ &\quad = \min\{RE(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1), B_i\} \\ &\quad = B_i \ \ge \ RE'(v_i, t+1) \ \ge \ 0. \end{aligned} \tag{31}$$

That is, $0 \le RE'(v_i, t+1) \le RE(v_i, t+1) \le B_i$.

On the other hand, for case 2, we have

$$\begin{aligned} &RE(v_i, t+1) \\ &\quad = \min\{RE(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1), B_i\} \\ &\quad = RE(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1) \\ &\quad \ge RE'(v_i, t) + \overline{H}(v_i, t+1) - P(v_i, t+1), \\ &\qquad\quad \text{as } RE(v_i, t) \ge RE'(v_i, t) \\ &\quad \ge RE'(v_i, t+1) \ge 0, \quad \text{by (17).} \end{aligned} \tag{32}$$

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 $\overline{H}(v_i, t)$ 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

Transform the original data utility maximization problem P1 into another convex optimization problem P2; and obtain its canonical form **P3**;
Construct a strictly feasible solution $(r_{i,t}, f_{ijt}, RE'(v_i, t), O_{it})$ to problem **P3**, by invoking `Algorithm` 1;
Obtain an optimal data rate r'_{it}, transmission rate f'_{ijt} and residual energy $RE'(v_i, t)$ for problem **P3** by applying the barrier method in [2];
Let $r_{it} = r'_{it}$ for $v_i \in V$ and $1 \le t \le T$ and $f_{ijt} = f'_{ijt}$ for $(v_i, v_j) \in E$ and $1 \le t \le T$;
return the data rate r_{it} and the transmission rate f_{ijt}.

Then, we still have $0 \le RE'(v_i, t+1) \le RE(v_i, t+1) \le B_i$.

Therefore, the inequality $0 \le RE'(v_i, t) \le RE(v_i, t) \le B_i$ holds for each sensor $v_i \in V$ at each time slot $0 \le t \le T$. Thus, $(r'_i(t), f'_{ijt})$ is a feasible solution to problem P1, then $OPT' \le OPT$. Recall that $OPT \le OPT'$, we know $OPT = OPT'$, which means that the optimal values for problems P1 and P2 are equal. The lemma then follows. ∎

Lemma 2: Given a feasible solution $(r'_{it}, f'_{ijt}, RE'(v_i, t))$ to problem P2, let $r_{it} = r'_{it}$ and $f_{ijt} = f'_{ijt}$. Then, (r_{it}, f_{ijt}) forms a feasible solution to problem P1, and its objective value in P1 is equal to the objective value of solution $(r'_{it}, f'_{ijt}, RE'(v_i, t))$ 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 $O((m + nT)\log(nT))$, where $n = |V|$ and $m = |E|$.

Proof: We first shows that Algorithm 1 terminates after $O(\log(nT))$ iterations of the *while* loop. It can be seen that both the sums of in-coming flows $\sum_{v_j \in N(v_i)} f_{jit}$ and out-coming flows $\sum_{v_j \in N(v_i)} f_{ijt}$ for each sensor v_i at each time slot t are only $O(\delta n)$. Then, the total energy consumption of sensor v_i in period T is $O(\delta n T)$. Assume that Algorithm 1 terminates after k iterations, which means that $([O(\delta n T)]/2^k) < RE(v_i, 0)$, 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 $O((m + nT)\log(nT))$. ∎

Theorem 1: Given 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 $\overline{H}(v_i, t)$ of sensor v_i 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 ϵ smaller than the maximum utility, where ϵ 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 ϵ smaller than the maximum utility, where ϵ is a given additive error with $0 < \epsilon < 1$. The theorem then follows. ■

IV. Performance Evaluation

In this section, we evaluate the performance of the proposed algorithm with existing algorithms for the data utility maximization problem.

A. Simulation Settings

We consider an energy harvesting sensor network consisting of from 50 to 200 sensors, which are randomly deployed in a 1000 m× 1000 m square area. The base station is located at the center of the area. The transmission range between any two nodes is 100 m. Assume that the battery capacity of each sensor is $B = 10.8$ kJ and its maximum sensing rate is $R_i = 1$ kb/s.

Each sensor is equipped with a 37×33 mm^2 solar panel. We adopt the real solar energy data collected by the Baseline Measurement System at the NREL Solar Radiation Research Laboratory from June 1, 2017 to August 31, 2017 (92 energy harvesting profiles in total) [22]. We allocate the harvested energy of each sensor in a day from the 92 energy profiles randomly. Specifically, the energy harvesting rate $H(v_i, t)$ of each sensor v_i at time slot t is $H(v_i, t) = \lambda \cdot H_{\text{sample}}(t)$, where λ is given constant with $0 \leq \lambda \leq 1$, and $H_{\text{sample}}(t)$ is the real energy harvesting rate of a randomly chosen energy profile at time slot t. The default value of λ is 1. Recall that we adopted the energy profiles in summer (from June 1 to August 31), to consider the energy harvesting profiles in other seasons, we can adopt a smaller value of λ. For example, we can set $\lambda = 0.1$ to obtain the energy profiles in winter. λ is called as *the energy scaling coefficient*.

The monitoring period T is one day, which consists of 24 time slots and the duration of each time slot thus is one hour, i.e., $\tau = 1$ h. The initial residual energy $RE(v_i, 0)$ of each sensor v_i at the beginning of the period T is randomly selected from 0 to 0.01 B, where B is the energy capacity of each sensor. The utility function is $U(D_i) = \log_2(D_i + 1)$, where D_i is the total amount of collected data from sensor v_i in period T.

To evaluate the performance of the proposed algorithm `maxUtility`, we consider existing five algorithms for the benchmark purpose.

1) Algorithm `maxThroughput` [21] aimed to maximize the total amount of collected data in a period by allocating harvested sensor energy.
2) Algorithm `utilityTimeSlot` [18] allocated sensor data rates and routing per time slot to maximize data utility of collected data in the time slot, based on the harvested energy of sensors.
3) Similar to algorithm `utilityTimeSlot`, algorithm `utilityTimeSlot+` [3] maximized data utility in each time slot. But the energy budget of each sensor in algorithm `utilityTimeSlot+` is equal to the minimum value among its residual energy and its past average harvested energy per time slot, rather than just its harvested energy in algorithm `utilityTimeSlot`.
4) Algorithm `utilityCorrelation` [6] focused on maximizing the total utility of collected data in a period T, assuming that the sensing data from each sensor in consecutive time slots is temporally independent, i.e., $\max \sum_{v_i \in V} \sum_{t=1}^{T} U(D_{it})$. It assumed that data routing is given.
5) Algorithm `utilityCorrelation+` [5] extended algorithm `utilityCorrelation`, by allowing data routing to change over time.

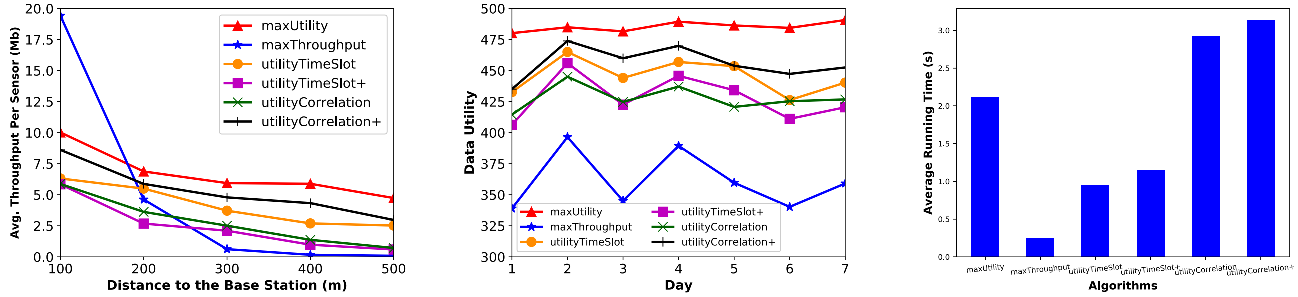
All experiments were performed on a server with an Intel Core i5-3330S CPU (2.70 GHz) and 8 GB RAM.

B. Algorithm Performance

We first study the performance of algorithms `maxUtility`, `maxThroughput`, `utilityTimeSlot`, `utilityTimeSlot+`, `utilityCorrelation`, and `utilityCorrelation+` within one week in a network with 100 sensors. Fig. 3(a) illustrates the distribution of average throughput per sensor in the seventh day with the distance between a sensor and the base station, from which it can be seen that, in algorithm `maxThroughput`, 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 m 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 m. 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 m. 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.

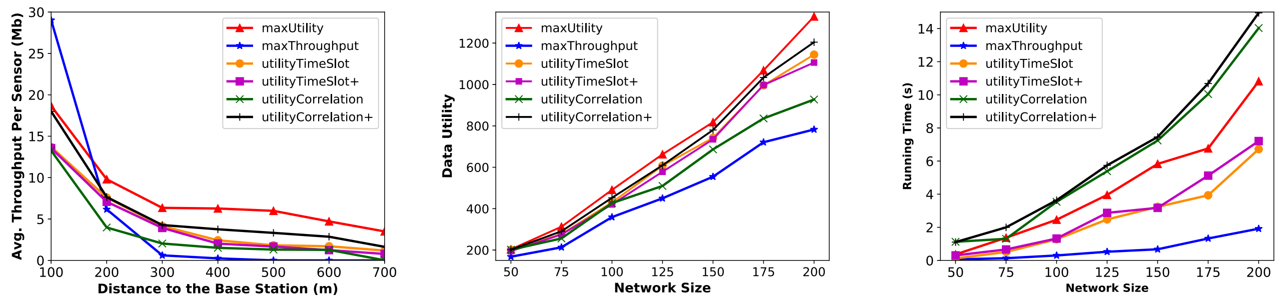
Fig. 3(b) plots the utility sums of collected data by different algorithms during the week. It can be seen that the utility sum by algorithm `maxUtility` is the largest one among the six algorithms, while the one by algorithm `maxThroughput` is the smallest, as it does not consider data correlation. 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`,



(a) (b) (c)

Fig. 3. Performance of different algorithms during a week with 100 sensors. (a) Average throughput per sensor in the seventh day with 100 sensors. (b) Utility sum of collected data in each day during one week. (c) Average running times of different algorithms.



(a) (b) (c)

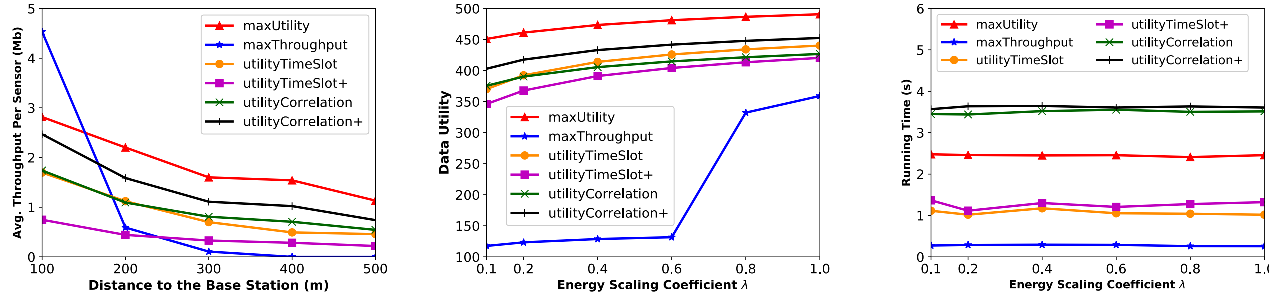
Fig. 4. Performance of different algorithms by varying the network size. (a) Average throughput per sensor with 200 sensors. (b) Utility sums of collected data with different network sizes. (c) Running times of different algorithms.

and `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 s.

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 m, 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 m. 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.

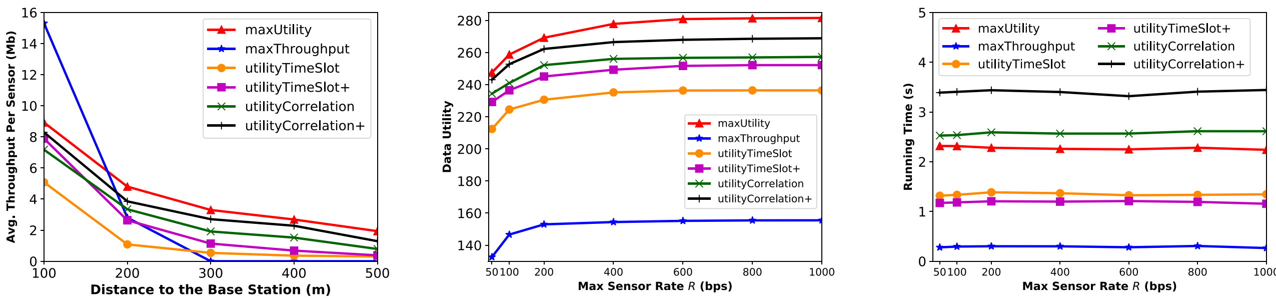
We also evaluate the impact of amounts of harvested energy on the data utility, by increasing the energy scaling efficient λ from 0.1 to 1 in a network with 100 sensors. Recall that we adopted real solar energy profiles in summer and the energy harvesting rate $H(v_i, t)$ of each sensor v_i at time slot t in our simulation is $H(v_i, t) = \lambda \cdot H_{\text{sample}}(t)$, where $H_{\text{sample}}(t)$ 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 λ. 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 m. On the other hand, in algorithm `maxThroughput`, only sensors close to the base station upload sensing data, while the sensors with 300 m 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 λ.

We finally study the impact of the maximum data rate R on the data utility, by increasing R from 50 b/s to 1000 b/s in a sensor network with 200 sensors. It can be seen from Fig. 6(a) that the average throughput by the proposed algorithm `maxUtility` is the highest for the sensors at least 200 m away from the base station, while the throughput by algorithm `maxThroughput` approaches to zero, when the distance is longer than 300 m, 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 b/s. 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



(a) (b) (c)

Fig. 5. Performance of different algorithms by varying the energy scaling coefficient λ from 0.1 to 1 in a network with 100 sensors. (a) Average throughput per sensor with $\lambda = 0.1$. (b) Utility sum by varying the energy scaling coefficient λ. (c) Running times of different algorithms.



(a) (b) (c)

Fig. 6. Performance of different algorithms by varying the maximum data rate R from 50 b/s to 1000 b/s with 200 sensors. (a) Average throughput per sensor with $R = 1000$ b/s. (b) Utility sum by varying the maximum data rate R. (c) Running times of different algorithms.

transmission for such a high data rate. Also, Fig. 6(b) demonstrates that the sum of data utility by algorithm `maxUtility` is about 4.7% higher than that by the second best algorithm `utilityCorrelation+`. Fig. 6(c) plots that the running time of algorithm `maxUtility` is only about 2.2 s.

V. 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 and Tucci [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. Shafieirad *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 satisfies 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 predefined path to collect sensor data in an energy harvesting sensor network. For example, studies [23] and [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, this paper is promising to improve the data collection performance.

VI. 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.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[2] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[3] S. Chen, P. Sinha, N. B. Shroff, and C. Joo, "A simple asymptotically optimal joint energy allocation and routing scheme in rechargeable sensor networks," *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1325–1336, Aug. 2014.

[4] A. Dehwaha, J. S. Shamma, and C. G. Claudel, "A distributed routing scheme for energy management in solar powered sensor networks," *Ad Hoc Netw.*, vol. 67, pp. 11–23, Dec. 2017.

[5] R. Deng, H. Liang, J. Yong, B. Chai, and T. Yang, "Distributed rate control, routing, and energy management in dynamic rechargeable sensor networks," *Peer-to-Peer Netw. Appl.*, vol. 10, no. 3, pp. 425–439, 2017.

[6] R. Deng, Y. Zhang, S. He, J. Chen, and X. Shen, "Maximizing network utility of rechargeable sensor networks with spatiotemporally coupled constraints," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1307–1319, May 2016.

[7] Y. Dong, J. Wang, B. Shim, and D. I. Kim, "DEARER: A distance-and-energy-aware routing with energy reservation for energy harvesting wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3798–3813, Dec. 2016.

[8] Q. Guo *et al.*, "Towards low-cost yet high-performance sensor networks by deploying a few ultra-fast charging battery powered sensors," *Sensors*, vol. 18, no. 9, p. 2771, 2018.

[9] G. Han, Y. Dong, H. Guo, L. Shu, and D. Wu, "Cross-layer optimized routing in wireless sensor networks with duty cycle and energy harvesting," *Wireless Commun. Mobile Comput.*, vol. 15, no. 16, pp. 1957–1981, 2015.

[10] *Internet of Things: Wireless Sensor Networks*, Geneva, Switzerland, IEC, White Paper, 2014. [Online]. Available: http://www.iec.ch/whitepaper/internetofthings/

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

[12] S. Kim *et al.*, "Health monitoring of civil infrastructures using wireless sensor networks," in *Proc. ACM 6th Int. Conf. Inform. Process. Sensor Netw. (IPSN)*, 2007, pp. 254–263.

[13] J. Li and P. Mohapatra, "Analytical modeling and mitigation techniques for the energy hole problem in sensor networks," *Pervasive Mobile Comput.*, vol. 3, no. 3, pp. 233–254, 2007.

[14] K. Li *et al.*, "Fair scheduling for data collection in mobile sensor networks with energy harvesting," *IEEE Trans. Mobile Comput.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8423120

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

[16] Y. Liang *et al.*, "Non-redundant information collection in rescue applications via an energy-constrained UAV," *IEEE Internet Things J.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8502870

[17] W. Liu, X. Zhou, S. Durrani, H. Mehrpouyan, and S. D. Blostein, "Energy harvesting wireless sensor networks: Delay analysis considering energy costs of sensing and transmission," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4635–4650, Jul. 2016.

[18] R. Liu, P. Sinha, and C. E. Koksal, "Joint energy management and resource allocation in rechargeable sensor networks," in *Proc. 30th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Diego, CA, USA, 2010, pp. 1–9.

[19] T. Lu, G. Liu, W. Li, S. Chang, and W. Guo, "Distributed sampling rate allocation for data quality maximization in rechargeable sensor networks," *J. Netw. Comput. Appl.*, vol. 80, pp. 1–9, Feb. 2017.
[20] S. Manfredi and E. D. Tucci, "Decentralized control algorithm for fast monitoring and efficient energy consumption in energy harvesting wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1513–1520, Aug. 2017.
[21] S. Mao, M. H. Cheung, and V. W. S. Wong, "Joint energy allocation for sensing and transmission in rechargeable wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2862–2875, Jul. 2014.
[22] *Measurement and Instrumentation Data Center*. Accessed: Oct. 20, 2017. [Online]. Available: http://midcdmz.nrel.gov/
[23] A. Mehrabi and K. Kim, "General framework for network throughput maximization in sink-based energy harvesting wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 7, pp. 1881–1896, Jul. 2017.
[24] C. Qiu, Y. Hu, Y. Chen, and B. Zeng, "Lyapunov optimization for energy harvesting wireless sensor communications," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1947–1956, Jun. 2018.
[25] X. Ren, W. Liang, and W. Xu, "Data collection maximization in renewable sensor networks via time-slot scheduling," *IEEE Trans. Comput.*, vol. 64, no. 7, pp. 1870–1883, Jul. 2015.
[26] H. Shafieirad, R. S. Adve, and S. Shahbazpanahi, "Max-SNR opportunistic routing for large-scale energy harvesting sensor networks," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 506–516, Jun. 2018.
[27] Q. Tan *et al.*, "Energy harvesting aware topology control with power adaptation in wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2014, pp. 2722–2727.
[28] L. Tan and S. Tang, "Energy harvesting wireless sensor node with temporal death: Novel models and analyses," *IEEE/ACM Trans. Netw.*, vol. 25, no. 2, pp. 896–909, Apr. 2017.
[29] C. Wang, S. Guo, and Y. Yang, "An optimization framework for mobile data collection in energy-harvesting wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 12, pp. 2969–2986, Dec. 2016.
[30] T. Wang, G. Zhang, A. Liu, M. Bhuiyan, and Q. Jin, "A secure IoT service architecture with an efficient balance dynamics based on cloud and edge computing," *IEEE Internet Things J.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/8464241
[31] W. Xu *et al.*, "Maximizing sensor lifetime with the minimal service cost of a mobile charger in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 11, pp. 2564–2577, Nov. 2018.
[32] W. Xu, W. Liang, X. Lin, and G. Mao, "Efficient scheduling of multiple mobile chargers for wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7670–7683, Sep. 2016.
[33] S. Yang, Y. Tahir, P.-Y. Chen, A. Marshall, and J. McCann, "Distributed optimization in energy harvesting sensor networks with dynamic in-network data processing," in *Proc. IEEE Int. Conf. Comput. Commun.*, San Francisco, CA, USA, 2016, pp. 1–9.
[34] J. Yang, X. Wu, and J. Wu, "Optimal scheduling of collaborative sensing in energy harvesting sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 512–523, Mar. 2015.
[35] Z. Yang, W. Xu, Y. Pan, C. Pan, and M. Chen, "Energy efficient resource allocation in machine-to-machine communications with multiple access and energy harvesting for IoT," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 229–245, Feb. 2018.
[36] D. Zeng *et al.*, "Energy minimization in multi-task software-defined sensor networks," *IEEE Trans. Comput.*, vol. 64, no. 11, pp. 3128–3139, Nov. 2015.
[37] D. Zhang *et al.*, "Utility-optimal resource management and allocation algorithm for energy harvesting cognitive radio sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3552–3565, Dec. 2016.
[38] D. Zhang, Z. Chen, H. Zhou, L. Chen, and X. Shen, "Energy-balanced cooperative transmission based on relay selection and power control in energy harvesting wireless sensor network," *Comput. Netw.*, vol. 104, pp. 189–197, Jul. 2016.
[39] Y. Zhang, S. He, and J. Chen, "Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1632–1646, Jun. 2016.
[40] Y. Zhang, S. He, and J. Chen, "Near optimal data gathering in rechargeable sensor networks with a mobile sink," *IEEE Trans. Mobile Comput.*, vol. 16, no. 6, pp. 1718–1729, Jun. 2017.
[41] Q. Zhang *et al.*, "An improved algorithm for dispatching the minimum number of electric charging vehicles for wireless sensor networks," *Wireless Netw.*, vol. 25, no. 3, pp. 1371–1384, 2019.
[42] N. Zlatanov, R. Schober, and Z. Hadzi-Velkov, "Asymptotically optimal power allocation for energy harvesting communication networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7286–7301, Aug. 2017.

Rongrong Zhang received the B.Sc. degree in computer science from Sichuan University, Chengdu, China, in 2016, where she is currently pursuing the master's degree.

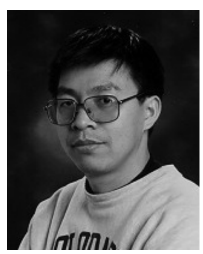
Her current research interests include wireless sensor networks and mobile computing.

Jian Peng received the B.A. and Ph.D. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1992 and 2004, respectively.

He is a Professor with the College of Computer Science, Sichuan University, Chengdu. His current research interests include wireless sensor networks, big data, and cloud computing.

Wenzheng Xu (M'15) received the B.Sc., M.E., and Ph.D. degrees in computer science from Sun Yat-sen University, Guangzhou, China, in 2008, 2010, and 2015, respectively.

He is currently an Associate Professor with Sichuan University, Chengdu, China. He was also a Visitor with the Australian National University, Canberra, ACT, Australia, and the Chinese University of Hong Kong, Hong Kong. His current research interests include wireless ad hoc and sensor networks, mobile computing, approximation algorithms, combinatorial optimization, online social networks, and graph theory.



Weifa Liang (M'99–SM'01) received the B.Sc. degree in computer science from Wuhan University, Wuhan, China, in 1984, the M.E. degree in computer science from the University of Science and Technology of China, Hefei, China, in 1989, and the Ph.D. degree in computer science from Australian National University, Canberra, ACT, Australia, in 1998.

He is currently a Full Professor with the Research School of Computer Science, Australian National University. His current research interests include design and analysis of energy-efficient routing protocols for wireless ad hoc and sensor networks, mobile edge computing and cloud computing, software-defined networking, online social networks, design and analysis of parallel and distributed algorithms, approximation algorithms, combinatorial optimization, and graph theory.

Zheng Li received the B.Sc. and M.E. degrees in computer science and the Ph.D. degree in applied mathematics from Sichuan University, Chengdu, China, in 1997, 2000, and 2009, respectively.

He is currently a Full Professor with Sichuan University. His current research interests include computer vision, wireless sensor networks, approximation algorithms, combinatorial optimization, and parallel and distributed algorithms.

Tian Wang received the B.Sc. and M.Sc. degrees in computer science from Central South University, Changsha, China, in 2004 and 2007, respectively, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2011.

He is currently a Professor with Huaqiao University, Xiamen, China. His current research interests include wireless sensor networks, social networks, and mobile computing.