

Maximizing Information Transmission for Energy Harvesting Sensor Networks by an Uneven Clustering Protocol and Energy Management

Yujia Ge^{1,2*}, Yurong Nan¹ and Yi Chen³

¹ College of Information Engineering, Zhejiang University of Technology
Hangzhou, Zhejiang - China
[E-mail: nyr@zjut.edu.cn]

² School of Computer and Information Engineering, Zhejiang Gongshang University
Hangzhou, Zhejiang - China
[E-mail: yge@mail.zjgsu.edu.cn]

³ Zhijiang College of Zhejiang University of Technology
Shaoxing, Zhejiang - China
[E-mail: eeyzchen@zjut.edu.cn]

*Corresponding author: Yujia Ge

*Received May 23, 2019; revised December 19, 2019; accepted January 27, 2020;
published April 30, 2020*

Abstract

For an energy harvesting sensor network, when the network lifetime is not the only primary goal, maximizing the network performance under environmental energy harvesting becomes a more critical issue. However, clustering protocols that aim at providing maximum information throughput have not been thoroughly explored in Energy Harvesting Wireless Sensor Networks (EH-WSNs). In this paper, clustering protocols are studied for maximizing the data transmission in the whole network. Based on a long short-term memory (LSTM) energy predictor and node energy consumption and supplement models, an uneven clustering protocol is proposed where the cluster head selection and cluster size control are thoroughly designed for this purpose. Simulations and results verify that the proposed scheme can outperform some classic schemes by having more data packets received by the cluster heads (CHs) and the base station (BS) under these energy constraints. The outcomes of this paper also provide some insights for choosing clustering routing protocols in EH-WSNs, by exploiting the factors such as uneven clustering size, number of clusters, multiple CHs, multihop routing strategy, and energy supplementing period.

Keywords: Wireless sensor networks; clustering; solar radiation prediction; routing protocols; energy balance

This research was partly supported by the National Social Sciences Foundation Projects [Grant No. 17BTQ069, 18BGL101].

1. Introduction

Internet of things (IoT) techniques could be widely applied in household devices and industrial elements by sensors. The traditional power supply of sensor nodes is usually batteries with a finite lifetime and may run out quickly after operating a period of time. Promising energy harvesting techniques provide a perpetual power supply for energy-limited networks, such as radio frequency energy harvesting techniques [1, 2]. Recent energy harvesting technologies from sources, such as solar irradiance, wind, thermoelectric, and piezoelectric, have made the possibility of more extensive wireless sensor networks applications and eliminates node dependence on battery power limit and maintenance [3, 4, 5]. Among these energy-harvesting sources, solar irradiance is most commonly used since it is reliable, predictable and has the highest power density. Recent researches on solar energy harvesting systems include sensor node architectures, system and performance modeling, and implementations. Energy harvesting techniques offer new opportunities for protocol and algorithm design. In a traditional Wireless Sensor Network (WSN), the network goal is to extend the whole network lifetime by reducing the energy consumption of the nodes, and now it is shifted to consider performance metrics, for example, energy usage, scalability, and network throughput on Energy Harvesting Wireless Sensor Networks (EH-WSNs).

There are still many challenges of making EH-WSNs useful in real applications, such as universal harvester design, energy-harvesting modeling, and network energy optimization [6]. Green communications utilize energy-efficient communications and networking technologies and protocols to minimize resource use in the whole network [7, 8]. A routing protocol is essential for network energy optimization in any energy-efficient WSNs. For traditional WSNs, extensive researches on routing protocols in traditional WSNs have been carried out. Three main types of routing protocols are flat routing, cluster-based routing and location-based routing protocol [9]. Among these protocols, clustering protocols have advantages in both efficiency and scalability in energy. In a typical cluster-based routing protocol, nodes in a WSN is clustered into several clusters. Each cluster is composed of a cluster head (CH) and some cluster members (CMs). CHs collect data sent by the CMs nearby within their clusters, then aggregate and forward these data to a possible far away base station (BS) which helps reduce energy consumption. But the CHs may be heavily burdened and nodes close to the BS have potentially more energy consumption. The objectives of the clustered protocols in traditional WSNs are usually to lower the energy consumption and extend the lifetime of sensor networks. Since the energy harvesting technology could make the sensor nodes acquire additional environmental energy continuously, the network lifetime is no more the primary goal. The ultimate goal of the EH-WSNs is to achieve energy balance by optimizing the energy consumption and configuration, which leads to a higher network throughput under long-term operation. A status known as Energy Neutral Operation (ENO) [10, 11, 12] in EH-WSN is the energy balance state of the whole network where the energy consumed by a node is always less than or equal to the energy harvested from the environment. However, clustering protocols that aim at providing maximum information throughput have not been thoroughly explored in EH-WSNs.

Due to the difference between traditional battery-powered sensor networks and energy harvesting sensor networks and more exhaustive research needs for EH-WSNs, this paper studies clustering protocols for maximizing information gathering under the energy-harvesting environment. Some insights for designing clustering routing protocols in

EH-WSNs are also provided and factors such as uneven clustering size, number of clusters, multiple CHs, multihop routing strategy, and energy supplementing period are explored. The proposed method is suitable for situations with the needs for a very long operational duration, such as environmental monitoring, structural monitoring, etc. In these situations, more data transmission and collection will provide better references for the entire application. The main contributions of our work are summarized as follows.

- 1) This paper proposes a carefully designed Uneven Clustering protocol for EH-WSNs (UCEH) where different cluster size is determined according to the distance from the sensor node to the BS and a harvesting energy index in an energy-harvesting environment. Extensive simulations demonstrate that better network performance referring to the network throughput and failure node percentage of the network could be achieved than classical clustering protocols.
- 2) An energy management framework is proposed with an LSTM-based energy predictor which increases solar radiation prediction accuracy and then helps for energy utilization and achieve energy balance in the whole network.
- 3) Experiments on factors such as uneven clustering size, number of clusters, multiple CHs in one cluster, multihop routing strategy, and energy supplementing period for a node are designed to give better thoughts on designing protocols in EH-WSNs.

2. Related Work

Clustering routing protocols of WSNs have been extensively studied in the literature. They are represented by the Low Energy Adaptive Clustering Hierarchy (LEACH) [13] and its following improved algorithms, such as Threshold Sensitive Energy Efficient Sensor Network (TEEN) [14], Energy Efficient Heterogeneous Cluster-based routing protocol (EEHC) [15] and Hybrid Energy Efficient Distributed Clustering Approach (HEED) [16]. Their main goals are to achieve energy balance among the whole networks. The operations of the clustering protocol are usually divided into three phases: the selection of CHs, the formation of clusters, and the transmission of data. The central part of the protocol is the CH selection algorithm because it determines the energy efficiency of the network. On the issue of the CH election, LEACH is the earliest proposed WSN clustering routing protocol, which uses random probability to select CHs and uniform clustering strategy and periodic CH rotation. HEED [16] is a multihop protocol using a fixed cluster radius and the selection of CHs in HEED depends mainly on the sensor residual energy and node density. LEACH with Sliding Window and Dynamic Number of Nodes (LEACH-SWDN) algorithm [17] generate a sliding window according to the initial energy of the node and the average energy of the non-CHs in this round. It adjusts the threshold and optimal cluster numbers accordingly. However, multihop protocols may lead to the “hot spot” problem in WSNs [18], i.e., the nodes nearing the BS die faster than nodes far from the BS. Uneven clustering algorithm could solve part of this problem. The Energy Efficient Uneven Clustering Algorithm (EEUC) [19] determines the uneven competition radius according to the distance between the nodes and the BS, so that the size of clusters close to the BS is smaller, saving energy for communication and transmission between clusters. Moreover, in the selection of the relay node, the remaining energy of the node and the distance from the CH to the BS are comprehensively considered.

Although clustering methods have been studied widely in traditional WSNs, clustering algorithms proposed for EH-WSNs are still not adequate. Among the existing limited work, ENC [20] is proposed to achieve energy neutrality by creating cluster head groups (CHGs), i.e., several nodes in a cluster take turns to be the CH. Traditional WSNs uses residual energy as an

essential index, while EH-WSNs could use the energy supplementing rate for routing decisions. In [21], the clustering protocol considers the energy harvesting nodes to get a more prominent possibility to be the CH.

Under the energy harvesting circumstances, power management systems and routing decisions heavily on the accuracy of environmental energy prediction models. The solar irradiation prediction algorithms are categorized into three major classes: statistical, stochastic, and machine learning based models [22]. Statistical models include the classic Exponential Weighted Moving Average (EWMA) [11], the Weather Conditioned Moving Average (WCMA) [23], and the Profile-Energy (Pro-Energy) model [24]. Other models such as Autoregressive Integrated Moving Average (ARIMA) and Linear Regression (LR) are also in this category. Stochastic models use stochastic processes, such as Markov Chains to represent signals. Machine learning prediction schemes include neural networks, such as Recurrent Neural Network (RNN) [25] and reinforcement learning [26]. Machine learning prediction schemes are shown to outperform the traditional models by achieving increased accuracy but with a more substantial computational burden [22].

Although there are already works on maximizing network throughput in WSNs and EH-WSNs, most of the work is focusing on formulating the mathematical problems in these situations and more investigations under different protocols need to be done more thoroughly. The mathematical formula to get the optimal transmitting power of nodes and the algorithm for select the CH to maximizing the data gathering in WSNs are investigated in [27]. Under EH-WSNs, a mixed-integer linear programming (MILP) optimization model for maximizing network throughput and designing algorithms to solve the problems in the scenario is introduced in [28].

3. System Model

3.1 Overall Framework

A typical clustered WSN as depicted in Fig. 1 has N energy harvesting nodes randomly located in the sensor network field which are grouped into several clusters. The field shape is a square with length L and thus the area of the field is L^2 square meters. According to the research that the best performance is achieved when a BS is located in the center of the field and the worst when a BS is in the corner [29], a BS in our system model is employed in the middle of the WSN field. All of the wireless sensor nodes are divided into different clusters periodically. Section 4 describes the details of an uneven clustering routing protocol. Each cluster is composed of a CH and some CMs. CHs of all the clusters collect data sent by the CMs within their clusters, then aggregate and forward these data to the BS. The following assumptions are made adhering to the existing literature [13, 19, 20]:

- The BS could get access to an unlimited amount of energy and have strong computational power, and all the node sensors use solar panel devices to complement energy.
- All of the sensor nodes, including the BS, are stationary and can get the localization by feasible techniques, such as GPS.
- Nodes can adjust their transmission power according to the distances between nodes.

For randomly located nodes, a clustered routing protocol needs to be carefully designed to achieve good performance, i.e., the maximized overall throughput and the minimum number of failure nodes, during a period of operation time. Due to the more complex situation in a harvesting energy environment, a framework of power management is designed at each sensor

node illustrated in Fig. 2. The power management framework relies heavily on energy predictor. The energy predictor is to predict the amount of energy that a sensor node can harvest in the next time slot. Power management of the whole network utilizes the periodical predicted energy information from the energy predictor, and the network information from the EH-WSN, such as the number of clusters, the number of alive nodes in this round. Then according to the current network status, the power management regulates the network parameters as the number of clusters and cluster sizes correspondingly which makes the better utilization of the energy and achieve energy neutralization in the whole network.

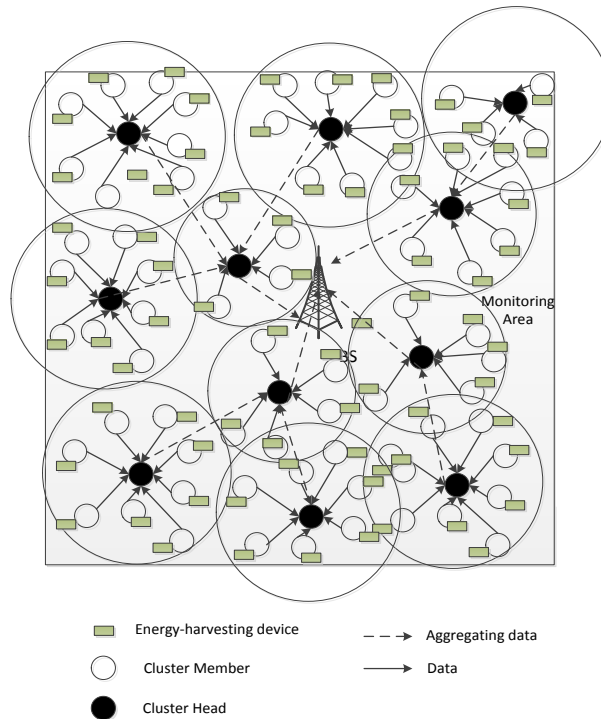


Fig. 1. Clustering protocol for EH-WSNs

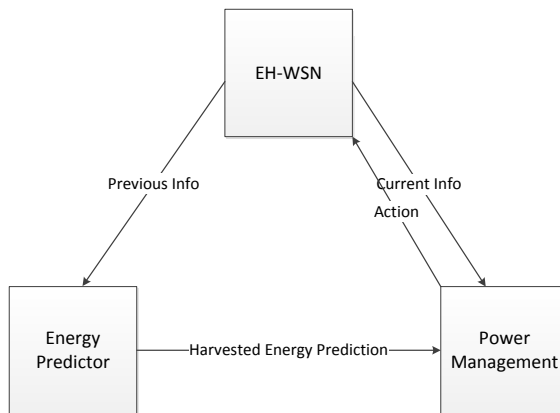


Fig. 2. The framework of power management at each node

3.2 Energy Predictor

The energy predictor is designed to provide an estimation of the energy which could be harvested for each node in a certain amount of time. Time series prediction in solar radiation refers to a sequence of numbers in which successive observations of the solar radiation are arranged and there is a statistical dependence on each other. As the most popular and widely used algorithm, the advantage of EWMA is to make full use of the solar cycle and adapt to seasonal changes. When the weather has been in a stable state, such as continuous sunny days and cloudy days, the prediction error of the algorithm is extremely small. At the same time, the main disadvantage of EWMA is its vulnerability to changing weather conditions. In particular, EWMA produces significant prediction errors when mixed on sunny and cloudy days. To improve the EWMA algorithm, WCMA was proposed. The dramatic error of the WCMA algorithm appears around time slots around sunset and sunrise which is because that the WCMA takes into account the preceding time slots. Due to the dramatic changes in sunshine conditions during sunrise and sunset, high error rates are caused. To achieve more accurate results, our harvesting energy predictor uses a long short-term memory (LSTM) neural network. The LSTM network was proposed by Hochreiter et al. in 1997 [30] which is an RNN architecture and suitable for time series sequence prediction. It was designed to eliminate the vanishing gradient problem and capture the long-term dependencies in the time series effectively which has been extensively applied in various fields. Fig. 3 shows the LSTM network structure with the embedding layer, LSTM layer, and output layer. Our LSTM networks use a multilayer architecture consisting of two LSTM layers and one fully connected layer. The Empirical Mode Decomposition (EMD) method [31] is used to decompose the time sequence data into a series of relatively stable component sequences before trained by LSTM networks. We finished experiments in LSTM, EWMA, and WCMA on the data from the US national solar radiation database [32] containing comprehensive solar and meteorological related data in more than 1000 locations of the United States. Each solar radiation historical data has a one-hour sample rate over a whole year period. Fig. 4 shows the prediction results by EWMA, WCMA, and LSTM compared with the actual data in two consecutive days (Oct 20 and 21) at the site of Alabama in 1999. For measuring prediction accuracy, Root Mean Square Error (RMSE) as the most commonly used metric is chosen to evaluate the performance between these prediction models in our experiments. Overall, the accuracy of LSTM outperforms other algorithms by 8.2-25.3%. The accuracy of energy prediction affects the overall performance of the sensor networks. Overestimates in prediction can lead to the energy depletion in some nodes and decreased the overall number of packages sent to the BS whereas underestimates could reduce throughput directly.

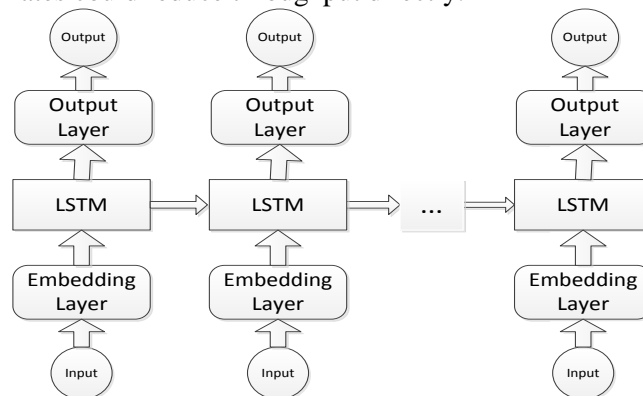


Fig. 3. LSTM structure

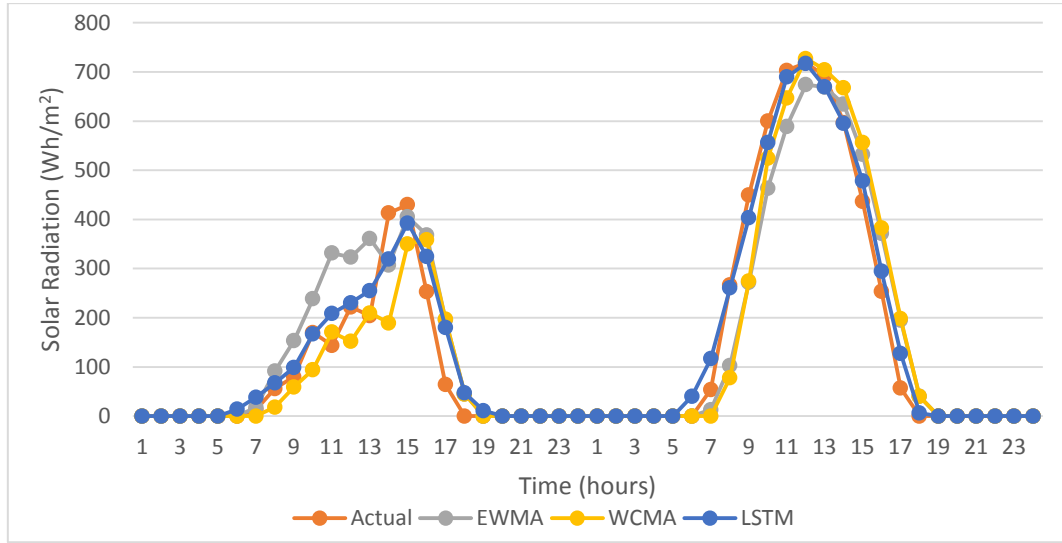


Fig. 4. Prediction results by WCMA, EWMA, LSTM compared with actual data

3.3 Energy Supplement Model and Energy Consumption Model

Energy is a limited resource for sensors in IoT, hence the supplement model and consumption model needs to be accurately modeled from real situations. As to the energy supplement model, we survey various products of solar energy harvesting sensor nodes, such as Prometheus [33], HydroWatch [34], Heliomote [35], Everlast [36], Sunflower [37], recently proposed DuraCap [38] and OpenWise [39], all of which use *Harvest-Store-Use* architecture. Specifically, we assume that the harvested energy of the sensors is stored in a supercapacitor or battery. Furthermore, when one day is divided into t time slots, the harvested energy supplement to sensors several times during a day. If st is to represent the supplementing time, we have $st = k * t$ where k is the supplementing sequence of the charging. $E_H^n(st)$ represents the amount of energy that can be used by a sensor n after harvesting energy in k time slots. According to the reported research [40], the average solar panel size is 8.15 inch². The LSTM-based energy predictor proposed in Section 3.2 accurately estimates the solar radiation for the next several time slots and then $E_H^n(st)$ could be derived accordingly. In every time slot, the energy management framework help efficiently assign energy through the whole network and make sure that the energy consumption in each node should be below the harvested energy plus the residual energy in this node.

For any node, energy is consumed during transmission, reception of packets and data dissemination. Adopted from the model of LEACH [13], the energy consumed by a sensor to transmit one bit of data is defined as E_t in Eq. (1).

$$E_t = \begin{cases} e_{et} + e_{sp} * d^2 & d \leq d_0 \\ e_{et} + e_{mp} * d^4 & d > d_0 \end{cases} \quad (1)$$

where e_{et} is the energy consumed by the transmitter for modulation, filtering and spreading data. d is the distance between the sending node and the receiving node. When d is less than a specified distance d_0 , the free space (d^2) model is used, and when it is more than d_0 , the multi-path fading (d^4) channel model is applied. The amplifier energy e_{sp} and e_{mp} depend on the transmission distance and the acceptable bit-error rate. The energy consumption model for receiving one bit of data is defined as E_r .

$$E_r = e_{er} \quad (2)$$

where e_{er} is the amount of energy that a receiver needs when receiving one bit of data.

For each CM in a cluster, the node needs to sense and transmit data to its CH. Comparing to transmitting data, the energy needed for sensing data could be neglected, so the energy used by CM is $E_{CM} = E_t$. It is assumed that the data received by a CH could be aggregated into one single packet due to the data correlation. So each CH consumes energy on receiving data from their CMs in the cluster, aggregating the data and transmitting the aggregated data to the BS as $E_{CH} = E_r + \alpha E_{DA} + E_t/\alpha$ where α is the aggregation rate and E_{DA} is the energy consumed by the CH for data aggregation.

4. Uneven Clustering Protocol for Maximizing Information Transmission

In a large-scale WSN, the nodes grouped into K clusters ($K > 1$) can help balance the energy through the whole network, where the CH nodes are able to aggregate information and send the information to the BS. To maximize utilizing the harvested energy, an uneven clustering protocol which has different sizes of clusters for maximizing information transmission is designed and the schemes of CHs reselection and multihop routing strategies are employed in our protocol.

4.1 Uneven Clustering Protocol Designed for EH-WSNs

Assume that, a unique sensor node ID is assigned to each sensor node and the location of each node is known. A node in each cluster is selected to be the head of the cluster, and each CH forwards the data to the BS through other intermediate CHs in a multihop path. When the distance from the CH to the BS is far, multihop routing is an excellent option to save energy. For each round, CHs are reselected according to the residual energy and predicted harvested energy of cluster nodes from the energy predictor. The nodes could be periodically rotated as the CHs, but the hot spot problem cannot be mitigated when nodes near the BS could die quicker. Therefore, unequal sizes of clustering are adopted to achieve better energy balance. Clusters close to the BS have smaller cluster sizes which comparatively consume less energy for processing data in the cluster.

Cluster Formation Algorithm

Initialization

For each node S_i , the location (x_i, y_i) is known. The field area is $(0,0)-(L,L)$. The coordinates of the BS are $(L/2, L/2)$ and the distance $d(S_i, BS)$ from each node to the BS is calculated.

Step 1. Tentative Cluster Head Selection

At the beginning of each round, nodes are randomly selected as tentative CHs, if $\text{Rand}() \leq p/(1 - p \times \text{mod}(r, \text{round}(1/p)))$, where p indicates the probability of being CH and r indicates the number of rounds.

Step 2. Cluster Head Competition

For any tentative CH_j , it determines the competition area R_c as

$$R_c = R_0 \times \left(1 - c \times \frac{d_{max} - d(S_i, BS)}{d_{max}}\right) \times I_{EH}$$

R_c guarantees that clusters closer to the BS have less competition ranges. In this competition range, only one CH with the highest residential energy and predicted harvesting energy in the next time slot becomes the final CH.

Step 3. Cluster Formation

All the nodes join the closest CH.

The process repeated when the next round starts from Step 1.

Fig. 5. Cluster formation algorithm

As illustrated in **Fig. 5**, in the initialization period of the cluster formation algorithm, the network is to be grouped into K clusters. The number of cluster K is controlled by the user which is explained later thoroughly in the simulation part. In this period, some sensors are randomly selected as the tentative CHs. First, each sensor node generates a random number from 0 to 1, and it then becomes the tentative CH if the number is smaller than the predefined probability P_k defined as $p/(1 - p \times \text{mod}(r, \text{round}(1/p)))$ where p indicates the probability of being CH, r indicates the number of rounds. In the cluster formation period of our uneven clustering algorithm, different competition ranges are produced according to the distance from the node to the BS and the harvesting energy prediction for the next time slot. The equation of the CH's competition range is adopted from [19] for an unequal clustering protocol and added a harvesting energy adjusting index with the closer the node to the BS, the smaller the size of the cluster. The tentative CH's competition range R_i is expressed in Eq. (3) as:

$$R_i = R_0 \times \left(1 - c \times \frac{d_{max} - d(S_i, BS)}{d_{max}}\right) \times I_{EH} \quad (3)$$

where R_0 denotes the maximum competition radius in the network which can be estimated from the analysis in Section 5.2, $d(S_i, BS)$ denotes the distance between S_i and the BS, d_{max} is the maximum distance of the node to the BS. The minimum competition range is close to $R_0 \times (1 - c)$. c is a constant coefficient between 0 and 1. When c is close to 1, the maximum competition range is much greater than the minimum competition range. I_{EH} is the predicted harvesting energy adjusting index which is between 0.5 and 1 where $I_{EH} = 0.5 + 0.5 * E_h/E_{max}$. If the expected solar radiation energy for next time slot (E_h) will be the maximum amount of energy we can harvest (E_{max}) according to the history data, e.g., at the noon during the summer, I_{EH} is set to be 1. In the night, the index is set to be close to 0.5. Therefore the maximum competition radius changes according to the harvested energy during a day. Each tentative CH also has knowledge of its adjacent tentative CHs. The node with maximum residual energy and estimated harvesting energy in the next time duration before the reselection of the CH is chosen as the final CH. If there is a tie, the node with the smaller ID is chosen. After the formation of clusters is completed, data transmission starts. In any round, the reselection of any CHs happens for each predefined time period.

4.2 Multihop Routing

In each cluster, all the CMs collect the data and directly send them to their CHs. After the CH receives all the data from its CMs, the CH first aggregates the data and then sends the packets to the BS by a multihop path through intermediate CHs.

Researches have been proved that a multihop routing protocol is more efficient in utilizing the energy more efficiently [9, 13]. To select the next hop node, if the distance of a node S_i to the BS is less than a certain distance d_0 , no intermediate node is needed, and it transmits the data to the BS directly. Otherwise, choosing an adjacent node S_j where S_j belongs to the CHs and $d^2(S_i, S_j) + d^2(S_j, BS) < d^2(S_i, BS)$ with the maximum residential energy. When the intermediate node is selected by considering factors other than distance, the harvested rate needs to be considered. Among all the eligible neighbors, nodes with the highest harvested energy are selected as the intermediate node.

5. Simulations

5.1 Parameter Setting

Communication-related parameters in our approach are shown in **Table 1**. Under these

settings, simulations are carried out on the Matlab platform [41] under a simplified situation. In these experiments, assume that it is in an ideal MAC protocol neglecting the packet loss error. The network fields are 200×200 square meters and the BS is located in the middle of the field. The data packet size of a CM sent to CH is 4000 bits/packet and the control packet size is 500 bits/packet. One time slot is set to be 1 hour (3600 s). The solar radiation data is retrieved from the US Solar Radiation Database [33] which is to be used as an estimation of the harvested energy in each time slot. For example, [0, 0, 0, 0, 0.00072, 0.02882, 0.11503, 0.20562, 0.24646, 0.27146, 0.28380, 0.29236, 0.27904, 0.27127, 0.26922, 0.24033, 0.16847, 0.07057, 0.00776, 0, 0, 0, 0, 0] is the average solar radiation (J) for each hour in one location of a day in the year of 2010. The harvested energy of days and nights have an obvious difference. Considering the harvested energy usage and hardware restriction, the charging sequence is set as once in every four time slots (hours).

Table 1. Parameter settings

Parameters	Value
Network field	(0, 0)-(200, 200)
Base station location	(100,100)
Data packet size	4000 bits/packet
Control packet size	500 bits/packet
Initial energy	0.5 J
Time slot	1 hour (3600 s)
n	100-400
e_{et}	5×10^{-8} J/bit
e_{er}	5×10^{-8} J/bit
E_{sp}	1×10^{-10} J/(bit \cdot m 2)
E_{mp}	1.3×10^{-15} J/(bit \cdot m 4)
E_{DA}	5×10^{-9} J/bit

According to Eq. (3) in Section 4, the competition radius of a CH varies according to the specified c where R_0 is set as a constant 90 meters in our experiments. The results of packages transmitted to the CHs under different c with two distinct number of nodes in the field is plotted in Fig. 6. From Fig. 6, we can see that when c is between 0.2-0.3, the CHs in the whole network receive the most number of packets. Therefore, we set c as 0.2 in later simulations.

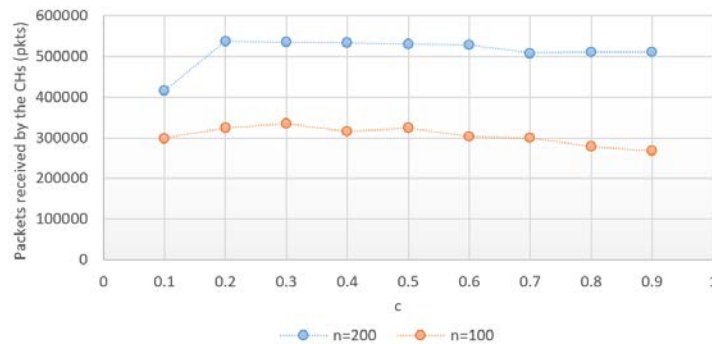


Fig. 6. Packets received by the CHs with different c

5.2 Discussion on Estimating Optimum Number of Clusters

In our approach, the network with K clusters could be estimated by formulating the

optimization problem to maximize the amount of information sent by all the CMs. As the number of clusters increases, the average number of CMs in each cluster will be decreased which lower the overall transmitted information and the burden for a CH. So K will be chosen as the smallest K for fulfilling the harvested data.

Here we simplify the situation by using the average value of cluster members. Assume that K clusters are formed and the average number of sensors in each cluster is N/K . For each cluster we have $N/K - 1$ average number of CMs, so we have:

$$\max (K * \left(\frac{N}{K} - 1\right) * L_t) \quad (4)$$

Subject to:

$$\left(\frac{N}{K} - 1\right) * L_t * \left(E_r + \left(\frac{N}{K} - 1\right) * E_{DA} + \frac{E_t}{\frac{N}{K} - 1}\right) \leq E_H^n(t) \quad (5)$$

Eq. (4) is the objective function which represents the data that could be transmitted by these CMs for one time slot t . L_t is the amount of information that a CM transmits in a time slot t . For both CMs and CHs, there are the energy constraints for energy neutrality condition. Since for each CH, the energy required for transmitting the collected information in the cluster is higher than the energy required for each CM. So we only formulate the energy neutrality constraint for the CHs in Eq. (5) which is a well-known convex optimization problem with optimal points (L_t, K) . To solve this problem, the following Lagrangian function Eq. (6) is formulated.

$$L = K * \left(\frac{N}{K} - 1\right) * L_t + v * \left(\left(\frac{N}{K} - 1\right) * L_t * \left(E_r + \left(\frac{N}{K} - 1\right) * E_{DA} + \frac{E_t}{\frac{N}{K} - 1}\right) - E_H^n(t)\right) \quad (6)$$

where $v \geq 0$ is the Lagrange Multiplier.

Additional complement slackness condition is shown as Eq. (7).

$$v * \left(\left(\frac{N}{K} - 1\right) * L_t * \left(E_r + \left(\frac{N}{K} - 1\right) * E_{DA} + \frac{E_t}{\frac{N}{K} - 1}\right) - E_H^n(t)\right) = 0 \quad (7)$$

By applying the Karush-Kuhn-Tucker condition, we have Eq. (8) and Eq. (9).

$$\frac{\Delta L}{\Delta L_t} = 0 \quad (8)$$

$$\frac{\Delta L}{\Delta K} = 0 \quad (9)$$

Through Eq. (7)- Eq. (9), we can solve L_t , K , and v to get the optimal point. The L_t and K will be the closest integer to the optimal point. For example, when the number of sensors N in the field equals 100 and the range of area is $100*100$ square meters, the result is shown in [Table 2](#).

Table 2. Optimum number of clusters

E(J)	Number of Clusters (K)
E = 0.000005	K = 14
E = 0.00005	K = 6
E = 0.0005	K = 3
E = 0.001	K = 2

During the process, to calculate the energy E_t by a node to transmit one bit of data, the distance d from the CH to the BS and the distance from the CM to the CH need to be estimated. All the other parameters are constant in one specific setting. From [42], we have Eq. (10) and Eq. (11) which could be put into Eq. (1).

$$E[d_{toCH}^2] = \frac{L^2}{2\pi K} \quad (10)$$

$$E[d_{toBS}] = 0.765 * L/2 \quad (11)$$

By calculating the equations, we can roughly estimate K first under a simplified situation.

The performance of uneven clustering depends on the value of c and R_{avg} . For each circular-shaped cluster having a radius of R_{avg} , we have Eq. (12).

$$K\pi R_{avg}^2 = L^2 \quad (12)$$

Suppose L equals to 200 m and n equals to 200, K is set to be 15 and R_{avg} is 29.1. If n equals to 100, K is set to be 10 and R_{avg} is 35.7.

5.3 Experimental Results and Analysis

To show that the proposed protocol can achieve better performance on maximizing the network information transmission, we first simulate and compare the number of packets sent by these different protocols respectively, i.e., the UCEH with one hop, the UCEH with multihop, LEACH and EEUC. The results in [Table 3](#) and [Fig. 7](#) show the transmission of the maximum packets received by the CHs in UCEH with multihop is about 6.1%-10.5% more than LEACH when the number of sensors ranges from 100 to 300. When the number of nodes in the field increases to 400, more noticeable performance differences between these two protocols regarding to the packages received by CHs, which UCEH outperforms LEACH by 33.6% in this situation. EEUC which is also an uneven clustering protocol has the similar performance as EEUC and the CMs in both protocols can achieve high energy utilization and efficiency. While referring to the number of packages received by the BS, the results in [Table 4](#) and [Fig. 8](#) show the more obvious performance difference between UCEH and EEUC which UCEH overperforms EEUC by 9.1%-12.1%. EEUC considers hot spot problems to maximize the lifetime of the network over LEACH and other network protocols while our algorithm UCEH is focusing on maximizing data transition and tuning the parameters for this aim. It implies in turn that UCEH can achieve better energy balance under EH-WSNs.

Table 3. Packets received by CHs in different protocols

Number of sensors	UCEH-one hop	UCEH-multihop	LEACH	EEUC
100	482390	499630	470743	482410
200	1703200	1773689	1604585	1702000
300	1805200	1893235	1773220	1850000
400	2110023	2444627	1829254	2360770

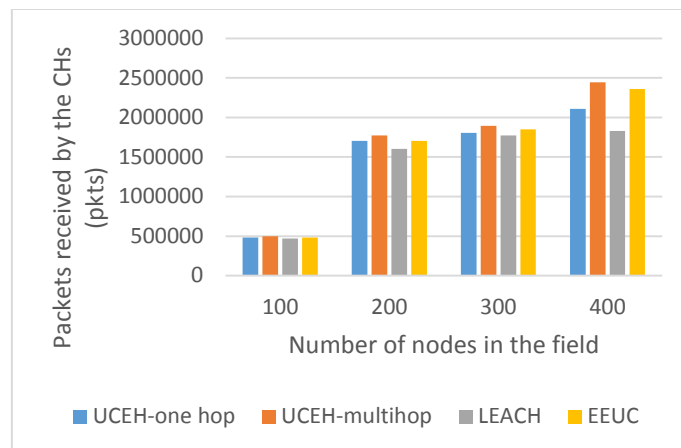


Fig. 7. Packets received by CHs in different protocols

Table 4. Packets received by the BS in different protocols

Number of sensors	UCEH-one hop	UCEH-multihop	LEACH	EEUC
100	39102	41700	37248	38200
200	131933	147807	133715	131833
300	130433	157769	147768	144166
400	175835	183718	152437	166730

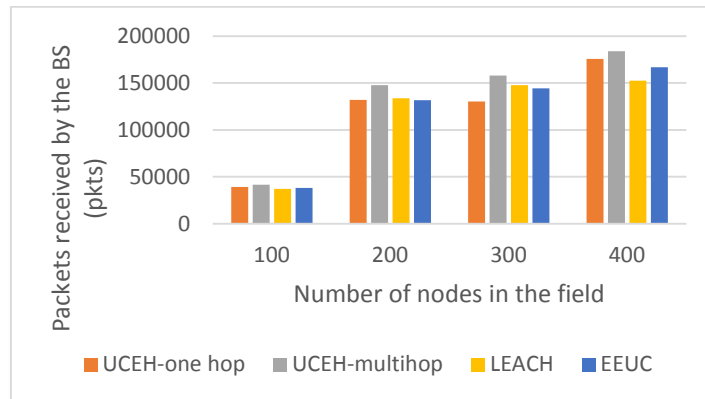


Fig. 8. Packets received by the BS in different protocols

Next, the simulation on the performance metric referring to the percentage of failure nodes in the network with the number of rounds is conducted under the same network configurations and energy budget as previously stated. If a node runs out of all the energy it contains, it will be shut down which is counted as a failure node. After its energy is refilled by the harvested source, the node could be alive and active again. Fig. 9 shows the trends of the percentage of failure nodes with the number of rounds under different protocols. In this one-day long (24 hours) simulation, the reselection rate is set as 10 times per time slot and the number of overall rounds is 240. The energy later supplemented by solar radiation activates all the failure nodes. The number of failure nodes under all the protocols is recorded during the simulation. Under LEACH, the number of failure nodes increases which reaches 22.3% of the whole network running out of the energy during the first 8 time slots (80 rounds). The percentages under EEUC and UCEH are 14.5% and 11.0% respectively. UCEH achieves comparably less percentage of failure nodes than other protocols which means that it achieves better network energy balance and also indicates that more packets are sent to the BS.

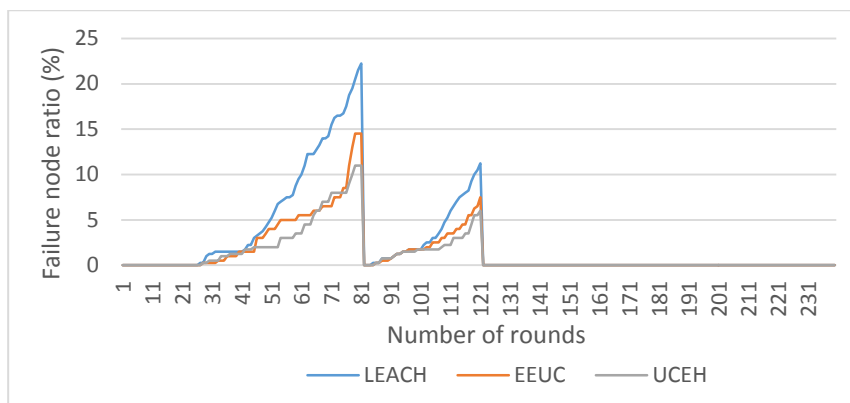


Fig. 9. Failure node ratio (%) under different protocols

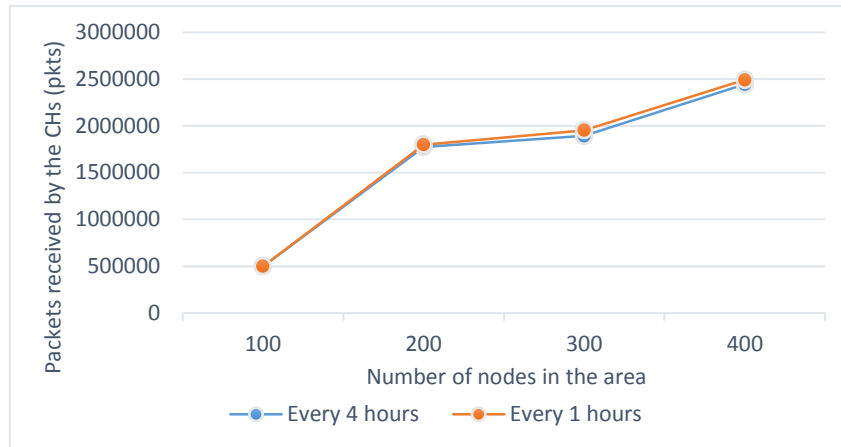


Fig. 10. Packets received by CHs (energy supplementing per hour vs per 4 hours)

We also compare different performance under different Energy Supplement Model. Due to the harvested energy usage and hardware restriction, the charging sequence is set as once in every 4 time slots (hours). **Fig. 10** shows the result when the energy harvesting is supplemented every hour. The numbers of packages sent to the CH and received by the BS are similar with a little bit increase. The results do not change a lot, especially when solar radiation is in an ordinary situation because our model has already aimed to maximize network throughput.

Other than the uneven clustering, we compare one CH in each cluster with multi-CHs as in [20]. We set the number of CHs in each cluster as 3. One CH is selected during the formation of clusters in UCEH. The other two heads are the two CMs that are closest to this previously selected CH. These three heads take turns as the CH and no selection needs for three rounds. The results in **Table 5** do not show much difference between UCEH with and without multiple CHs, which means no reselection could not help much in improving the information gathering. In our simulations, the reselection burden is comparatively small compared to transmitting data.

Table 5. Comparison of packets received by CHs with one and three CHs in a cluster

Number of sensors	UCEH	UCEH-3 CHs
100	499630	498356
200	1773689	1792342
300	1893235	1876234
400	2444627	2452230

6. Conclusions

In this paper, we have proposed an uneven clustering protocol for EH-WSN, where the CH selection and cluster size control are designed for maximizing the information transmission in the whole network. The outcomes of this paper can provide some insights for clustering routing protocol design in EH-WSN, by exploiting uneven clustering, multiple CHs, and multihop routing.

Under our proposed protocol, we analyze the optimum parameters derived simplified network environment settings. Empirical studies verify that the proposed scheme can outperform other classic schemes, i.e., LEACH and EECH, by having more balanced energy and more packets received by the CHs and the BS with a lower number of failed nodes. We also compare the environment under different solar supplementing models and compare different performance with different number of CHs in a cluster. Results show the performance does not show much difference in maximizing data transmission under the energy harvesting situation.

Besides the fixed data transmission rate, the changeable data transmission rate can apparently improve the information throughput especially when the harvested energy is not always stable during a day. For our future work, we plan to extend the framework on automatically adjusting the network parameters by the power management to maximizing the information transmission, such as dynamic data transmission rate and more rigorous simulations will be conducted.

References

- [1] W. Wu, B. Wang, Y. Zeng, H. Zhang, Z. Yang and Z. Deng, "Robust Secure Beamforming for Wireless Powered Full-Duplex Systems With Self-Energy Recycling," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10055-10069, Nov. 2017. [Article \(CrossRef Link\)](#).
- [2] W. Wu, B. Wang, Z. Deng and H. Zhang, "Secure Beamforming for Full-Duplex Wireless Powered Communication Systems With Self-Energy Recycling," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 146-149, April 2017. [Article \(CrossRef Link\)](#).
- [3] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443-461, 2011. [Article \(CrossRef Link\)](#).
- [4] A. Babayo, H. Anisi, A. Hossein & I. Ali, "A Review on energy management schemes in energy harvesting wireless sensor networks," *Renewable and Sustainable Energy Reviews*, vol. 76, pp.1176-1184, 2017. [Article \(CrossRef Link\)](#).
- [5] F.K. Shaikh, S. Zeadally, "Energy harvesting in wireless sensor networks: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 55, 2016, pp. 1041-1054, 2016. [Article \(CrossRef Link\)](#).
- [6] Adam, C. Tapparello, H. Ayatollahi, W. Heinzelman, "Energy-Harvesting Wireless Sensor Networks (EH-WSNs): A Review," *ACM Transactions on Sensor Networks*, vol. 14, no. 2, Article No. 10, July 2018. [Article \(CrossRef Link\)](#).
- [7] C. Li, S. Zhang, P. Liu, F. Sun, J. M. Cioffi and L. Yang, "Overhearing Protocol Design Exploiting Intercell Interference in Cooperative Green Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 441-446, Jan. 2016. [Article \(CrossRef Link\)](#).
- [8] C. Li, H. J. Yang, F. Sun, J. M. Cioffi and L. Yang, "Multiuser Overhearing for Cooperative Two-Way Multiantenna Relays," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3796-3802, May 2016. [Article \(CrossRef Link\)](#).
- [9] N. Pantazis, S. Nikolidakis, D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, pp. 551-591, 2013. [Article \(CrossRef Link\)](#).
- [10] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy-harvesting sensor nodes," *IEEE Trans. Wireless Commun.* vol. 9, no. 4, pp. 1326-1336, April 2010. [Article \(CrossRef Link\)](#).
- [11] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy-harvesting sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 6, no. 4, article no. 32, 2007. [Article \(CrossRef Link\)](#).

- [12] S. Baghaee, S. Chamanian, H. Ullusan, O. Zorlu, E. Uysal-Biyikoglu, and H. Kulah, "Demonstration of energy-neutral operation on a WSN testbed using vibration energy harvesting," in *Proc. of the 20th European Wireless Conference, Barcelona, Spain*, pp.1-6, 2014.
- [13] W. B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. of the 33rd Annual Hawaii International Conference on System Sciences, Wailea Maui, Hawaii*, pp.10-19, 2000. [Article \(CrossRef Link\)](#).
- [14] A. Manjeshwar and D. P. Agrawal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," in *Proc. of 15th International Parallel and Distributed Processing Symposium. IPDPS 2001, San Francisco, CA, USA*, pp. 2009-2015, 2001. [Article \(CrossRef Link\)](#).
- [15] Seema Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. of Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)(INFOCOM 2003), San Francisco, CA*, vol.3, pp.1713-1723, 2003. [Article \(CrossRef Link\)](#).
- [16] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366-379, Oct.-Dec. 2004. [Article \(CrossRef Link\)](#).
- [17] A.M. Wang, D.L. Yang, D.Y. Sun, "A Clustering Algorithm Based on Energy Information and Cluster Heads Expectation for Wireless Sensor Networks," *Computers and Electrical Engineering*, vol. 38, no. 3, pp.662-671, 2012. [Article \(CrossRef Link\)](#).
- [18] D. Wei, Y. Jin, S. Vural, K. Moessner and R. Tafazolli, "An Energy-Efficient Clustering Solution for Wireless Sensor Networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3973-3983, November 2011. [Article \(CrossRef Link\)](#).
- [19] G. Chen, C. Li, M. Ye, J. Wu, "An unequal cluster-based routing protocol in wireless sensor networks," *Wireless Networks*, vol. 15, no. 2, pp.193-207, 2009. [Article \(CrossRef Link\)](#)
- [20] Shuai Peng and Chor Ping Low, "Energy Neutral Routing for energy harvesting Wireless Sensor Networks," in *Proc. of 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai*, pp. 2063-2067, 2013. [Article \(CrossRef Link\)](#).
- [21] T. Voigt, A. Dunkels, J. Alonso, H. Ritter and J. Schiller, "Solar-aware clustering in wireless sensor networks," in *Proc. of IEEE Ninth International Symposium on Computers And Communications (ISCC 2004), Alexandria, Egypt*, Vol. 1, pp. 238-243, 2004. [Article \(CrossRef Link\)](#).
- [22] Muhammad, Hassaan Khaliq Qureshi, Umber Saleem, Muhammad Saleem, Andreas Pitsillides, and Marios Lestas, "Harvested Energy Prediction Schemes for Wireless Sensor Networks: Performance Evaluation and Enhancements," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 6928325, 14 pages, 2017. [Article \(CrossRef Link\)](#).
- [23] J.R. Piorno, C. Bergonzini, D. Atienza, T.S. Rosing, "Prediction and management in energy harvested wireless sensor nodes," in *Proc. of 2009 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Aalborg, Denmark*, pp. 6-10, May 17-20 2009. [Article \(CrossRef Link\)](#).
- [24] A. Cammarano, C. Petrioli and D. Spenza, "Pro-Energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks," in *Proc. of 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012), Las Vegas, NV*, pp. 75-83, 2012. [Article \(CrossRef Link\)](#).
- [25] W. Lee, K. Kim, J. Park, J. Kim and Y. Kim, "Forecasting Solar Power Using Long-Short Term Memory and Convolutional Neural Networks," *IEEE Access*, vol. 6, pp. 73068-73080, 2018. [Article \(CrossRef Link\)](#).
- [26] S. Kosunalp, "A New Energy Prediction Algorithm for Energy-Harvesting Wireless Sensor Networks With Q-Learning," *IEEE Access*, vol. 4, pp. 5755-5763, 2016. [Article \(CrossRef Link\)](#)
- [27] T. Wang, W. Heinzelman and A. Seyed, "Maximization of Data Gathering in Clustered Wireless Sensor Networks," in *Proc. of 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL*, pp. 1-5, 2010. [Article \(CrossRef Link\)](#).
- [28] A. Mehrabi and K. Kim, "General Framework for Network Throughput Maximization in Sink-Based Energy Harvesting Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 7, pp. 1881-1896, 1 July 2017. [Article \(CrossRef Link\)](#).

- [29] O. Koyi, H. Yang, and Y. Kwon, "Impact of Base Station Location on Wireless Sensor Networks," in *Proc. of the 4th Computer Science On-line Conference 2015 (CSOC2015), vol 2: Intelligent Systems in Cybernetics and Automation Theory, Zlin, Czech Republic*, pp.151-162, April 27-30, 2015. [Article \(CrossRef Link\)](#).
- [30] X. Jiang, J. Polastre, and D. Culler, "Perpetual Environmentally Powered Sensor Networks," in *Proc. of Fourth International Symposium on Information Processing in Sensor Networks, Piscataway, NJ, USA*, pp. 463-468, April 2005.
- [31] B. Huang, A. Kunoth, "An optimization based empirical mode decomposition scheme," *J. Comput. Appl. Math.*, vol. 240, pp. 174-183, 2013. [Article \(CrossRef Link\)](#).
- [32] National Solar Radiation Data Base. http://redc.nrel.gov/solar/old_data/nsrdb/. Retrieved December 10, 2019.
- [33] J. Taneja, J. Jeong and D. Culler, "Design, Modeling, and Capacity Planning for Micro-solar Power Sensor Networks," in *Proc. of 2008 International Conference on Information Processing in Sensor Networks (ipsn 2008), St. Louis, MO*, pp. 407-418, 2008. [Article \(CrossRef Link\)](#).
- [34] Vijay Raghunathan, A. Kansal, J. Hsu, J. Friedman and Mani Srivastava, "Design considerations for solar energy harvesting wireless embedded systems," in *Proc. of IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, Boise, ID, USA*, pp. 457-462, 2005. [Article \(CrossRef Link\)](#)
- [35] F. Simjee and P. H. Chou, "Everlast: Long-life, Supercapacitor-operated Wireless Sensor Node," in *Proc. of 2006 International Symposium on Low Power Electronics and Design (ISLPED'06), Tegernsee*, pp. 197-202, 2006. [Article \(CrossRef Link\)](#).
- [36] P. Stanley-Marbell and D. Marculescu, "An 0.9 x 1.2, Low Power, Energy-Harvesting System with Custom Multi-Channel Communication Interface," in *Proc. of 2007 Design, Automation & Test in Europe Conference & Exhibition, Nice*, pp. 1-6, 2007. [Article \(CrossRef Link\)](#).
- [37] S. Sudevalayam and P. Kulkarni, "Energy Harvesting Sensor Nodes: Survey and Implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443-461, Third Quarter 2011. [Article \(CrossRef Link\)](#).
- [38] C. Chen and P. H. Chou, "DuraCap: A supercapacitor-based, power-bootstrapping, maximum power point tracking energy-harvesting system," in *Proc. of the 16th ACM/IEEE International Symposium on Low Power Electronics and Design*, pp. 313-318, 2010. [Article \(CrossRef Link\)](#).
- [39] A. Gonzalez, R. Aquino, W. Mata, A. Ochoa, P. Saldana, and A. Edwards, "Open-wise: A solar powered wireless sensor network platform," *Sensors*, vol.12, no. 6, pp. 8204-8217, 2012. [Article \(CrossRef Link\)](#).
- [40] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. [Article \(CrossRef Link\)](#).
- [41] MATLAB, version 7.11.0 (R2010b). The Mathworks Inc., Natick, Massachusetts, 2010.
- [42] G. Smaragdakis, I. Matta and A. Bestavros, "SEP: A Stable Election Protocol for Clustered Heterogeneous Wireless Sensor Networks," in *Proc. of Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA), Boston, US*, pp. 251-261, 2004.



Yujia Ge received her M. Eng in Computer Applications from Zhejiang University, China in 2001 and her M. S. in Computer Science from Iowa State University, USA in 2004. She currently teaches in School of Computer and Information Engineering at Zhejiang Gongshang University, China. She is pursuing her doctoral degree under the supervision of Dr. Yuhong Nan in College of Information Engineering, Zhejiang University of Technology, China. Her research interests are IoT applications and Search-based Software Engineering.



Yurong Nan is a full professor in College of Information Engineering, Zhejiang University of Technology, China. His research interests are embedded systems, motion control, and computer control technology. He is the academic leader of the research direction “Information Processing and Automation Technology” in Zhejiang Province.



Yi Chen received her Ph. D in the Hong Kong Polytechnic University in 2006 and currently is a full professor in Zhijiang College of Zhejiang University of Technology, China. Her research interests are power electronics and energy harvesting technology.