1-1-2020

Cooperative relay selection for load balancing with mobility in hierarchical WSNs: A multi-armed bandit approach

Jian Zhang
Nanjing University of Information Science & Technology

Jian Tang
Beijing University of Technology

Feng Wang
University of Mississippi

Recommended Citation

Zhang, J., Tang, J., & Wang, F. (2020). Cooperative Relay Selection for Load Balancing With Mobility in Hierarchical WSNs: A Multi-Armed Bandit Approach. IEEE Access, 8, 18110–18122. https://doi.org/10.1109/ACCESS.2020.2968562

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 30, 2019, accepted January 15, 2020, date of publication January 22, 2020, date of current version January 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968562

Cooperative Relay Selection for Load Balancing With Mobility in Hierarchical WSNs: A Multi-Armed Bandit Approach

JIAN ZHANG[1,2], JIAN TANG[3], AND FENG WANG[4], (Senior Member, IEEE)

[1]College of computer and software, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing 210044, China
[3]Department of Information, Beijing University of Technology, Beijing 100124, China
[4]Department of Computer and Information Science, The University of Mississippi, Oxford, MS 38655, USA

Corresponding author: Jian Tang (freeflytang@bjut.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61573364 and Grant 618730099, in part by the Open Foundation of Education Ministry Demonstration Base of Internet Application Innovation Open Platform (Meteorological Cloud Platform and Application) under Grant 2201101401063, in part by the National & Beijing Key Laboratory of Process Automation in Mining and Metallurgy under Grant BGRIMM-KZSKL-2018-06, and in part by the a Project Funded by the Priority Academic Program Development of Jiangsu Higer Education Institutions (PAPD).

ABSTRACT Energy efficiency is the major concern in hierarchical wireless sensor networks(WSNs), where the major energy consumption originates from radios for communication. Due to notable energy expenditure of long-range transmission for cluster members and data aggregation for Cluster Head (CH), saving and balancing energy consumption is a tricky challenge in WSNs. In this paper, we design a CH selection mechanism with a mobile sink (MS) while proposing relay selection algorithms with multi-user multi-armed bandit (UM-MAB) to solve the problem of energy efficiency. According to the definition of node density and residual energy, we propose a conception referred to as a Virtual Head (VH) for MS to collect data in terms of energy efficiency. Moreover, we naturally change the relay selection problem into permutation problem through employing the two-hop transmission in cooperative power line communication, which deals with long-distance transmission. As far as the relay selection problem is concerned, we propose the machine learning algorithm, namely MU-MAB, to solve it through the reward associated with an increment for energy consumption. Furthermore, we employ the stable matching theory based on marginal utility for the allocation of the final one-to-one optimal combinations to achieve energy efficiency. In order to evaluate MU-MAB, the regret is taken advantage to demonstrate the performance by using upper confidence bound (UCB) index. In the end, simulation results illustrate the efficacy and effectiveness of our proposed solutions for saving and balancing energy consumption.

INDEX TERMS Wireless sensor networks, relay selection, mobility, marginal utility, matching theory, multi-armed bandit.

I. INTRODUCTION

Wireless sensor networks (WSNs) have recently gained increasing popularity in ubiquitous support of sensing system services which periodically sense and transmit collected data to the Sink [1], and has widely involved many applications including modern industrial processes and automation, environment monitoring, intelligent transportation, and military surveillance [2], [3]. Due to the limitation in battery resource of sensor nodes, energy conservation is always a pivotal challenge for prolonging the lifetime of WSNs. Actually, long geographical distance transmission often aggravates energy expenditure. For example, sensor nodes sending data directly to the Sink will consume more energy. Moreover, exchanging or recharging power supplies of sensor nodes is usually difficult, which generally incurs the phenomenon of disconnection in WSNs. So it is harmful or disastrous for applications to collect data in these situations.

The cluster-based hierarchical technology is one of the most effective and promising schemes to enhance energy conservation [4], [5]. However, cluster heads (CHs) forwarding packets to the Sink via either long-distance or multi-hop

The associate editor coordinating the review of this manuscript and approving it for publication was Chien-Fu Cheng.

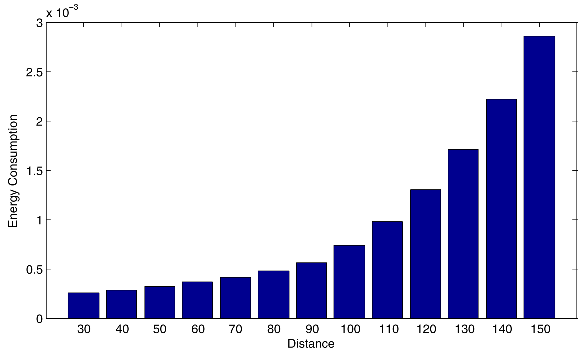


FIGURE 1. Energy consumption with transmission ranges.

transmissions usually drain their energy quickly, which leads to energy consumption unbalance and limitation of the network performance [6], [7]. Especially, CHs located nearby the Sink deplete their energy quicker than other ones, which incurs the problem of energy holes [8].In addition, the CH, which plays a role of the arrangement for the system operation in an associated manner to improve the performance of WSNs, is always overburdened during the process of cluster formation [9].

Combining the cluster-based hierarchical technology and the mobility for data collection seems to be a promising way to achieve a better performance of WSNs [10]. Accordingly, researchers have already exploited mobility of the mobile devices or robots as a novel innovation to assist conventional multi-hop transmission for data collection [11], [12]. However, the main concern of previous schemes mainly focuses on how to select CH to achieve energy consumption balance in terms of clustering [13], which exactly determines the lifetime of WSNs. On a broader level, it is not reasonable for the solution of energy efficiency to design a multi-hop path without paying attention to the communication between nodes in one cluster. Consequently, the energy consumption of a node distributed at the edge of the network will be consumed quickly owing to the long-distance communication between the node and CH [14]. In fact, the energy consumption of cluster members will exponentially increase with transmission distance largening according to their transmission power models. As shown in Fig. 1, the energy expenditure for diverse transmission distances differentiate significantly. For instance, the long-distance data communication will consume a large amount of energy when the large volume of data are exchanged or proceed in the cluster [15]. So one solution for this phenomenon is that a short-distance multi-hop path will be designed via introducing relay technologies for energy reservation to provide data services for upper applications [16].

Although multi-hop clustering algorithms can effectively expand the cluster coverage and reduce the number of CHs, Exchanging information frequently will incur more energy consumption in the process of forming a stable cluster [17]. Based on the analysis and research of multi-hop clustering algorithms published in recent years, we propose a multi-hop clustering mechanism with MS to collect data for energy efficiency. In the process of building the multi-hop cluster, we propose multi-user multi-armed bandit (MU-MAB) algorithms for relay selection to reduce and balance energy consumption without prior knowledge. The major contributions of this paper are summarized as follows.

1) we propose a CH selection mechanism based on definitions of node density and residual energy to relieve the burden of CH while reducing and balancing energy consumption of the cluster. In this case, nodes with long-distance transmission to CH are separated, whose data should be transferred by the relay node.

2) We create the conception of virtual head (VH) with a mobile sink (MS) to collect data for balancing energy consumption and shortening average distance of separated edge nodes.

3) In order to achieve energy optimization in terms of multi-hop transmission, we proposed MU-MAB for relay selection without prior knowledge to reduce and balance energy consumption. Via defining marginal utility function and introducing matching theory, we build the optimization framework for relay selection in terms of energy efficiency with a most upper bound of the expected regret.

4) The relations between clustering and routing in our algorithms are further exploited by theoretical and numerical analysis, and the results are respectively demonstrated from several aspects to verify the validation of our proposed algorithms.

This paper is organized as follows. In Section II, we introduce a brief survey of related works. Section III provides some preliminaries and formulations based on our proposed basic framework. In Section IV, we describe the principles for relay selection. Afterwards, we detailedly analyze the performance of system model and propose MU-MAB algorithms for data collection in Section V. Section VI gives the regret analysis for MU-MAB and simulations are clarified in Section VII. Finally, Section VIII concludes this paper.

II. RELATED WORKS

Clustering greatly contributes to energy efficiency and network lifetime, which enhances the power allocation and the benefit recapture of resource [18]. In [9], single hop communication is applied to minimize energy consumption at cluster level by creating an optimal data collecting chain based on the fact that node's energy capacity is rather limited and the communication overhead is proportional to the transmission distance. However, the energy hole problem incurs lower performance of WSNs at global level for data collection, which could be circumvented through an application-based optimization of multi-level clustering algorithm proposed in [19].

In practical applications, data collection always combines the mobility to achieve better performance of WSNs. In [20], a novel dynamic clustering called mobile-to-cluster scheme is proposed to optimize the service process, which employs mobile vehicles to balance the load of WSNs. The authors in [21] focus on low delay and high-throughput opportunistic data collection in WSNs with general network topologies and arbitrary numbers of MSs. The study in [22] utilizes mobile robots to create a connected path from the base station to the event for in-network forwarding. MS and CHs in [23] are collaboratively considered to minimize the total dissipated energy in communication. However, most of them only consider the built network pattern to better the energy efficiency of WSNs, not referring to the complex environment where the optimization performance of WSNs has to face uncertainty. For example in [24], the long range communication may incur the obstacle of power consumption, which is potential for reducing the power consumption by novel approaches combining WSNs with Unmanned Aerial Vehicles (UAV).

For multi-hop cluster data collection, one significant issue is the relay selection problem because of energy constrained WSNs. Relay selection is proposed in [1], where an optimized forwarding tree and a minimized UAV trajectory distance are jointly designed to gather data. Additionally, the relay selection problem can be solved by channel state information (CSI) at transmitter, which selects the most appropriate relay to achieve the best performance [25]. However, acquiring the channel state information at transmitter could increase the complexity of the communication system and introduce undesired overhead to the system. To deal with this problem, the multi-armed bandit (MAB) is employed to replace channel state information. In [26], the authors introduce UAV in wireless communication systems for dynamic relaying and large-area environmental sensing, which are referred to as the UAV-assisted networks. In order to handle large data volumes in the long term for stand alone UAVs with constrained energy and processing capacity, authors in [27] determine optimized selection of multi-hop path between a source and a target UAV with MAB. In [28], a DSMU-MAB algorithm based on stable matching and the designed back-off timer is presented to reduce the frequency of the information exchange for relay selection. It is better to use multi-hop intra-cluster communications, if there is a small number of CHs when member nodes are far from CHs or when there are transfer restrictions on sensors. Hence, [3] considers an evaluation criterion of parameters either one-hop or multi-hop WSNs.

III. PRELIMINARIES AND FORMULATIONS

In this section, we will formally introduce the energy efficiency problem discussed in this paper, which mainly focuses on a dynamic cluster changing as the density of CH to balance energy consumption of the cluster for data collection. After that, we consequentially analyze properties of Equation (1) as the criterion in the following sections.

A. ENERGY CONSUMPTION MODEL

The energy consumption formulas follow the popular models given in [18]. To transmit l-bit packets from node i to node j,

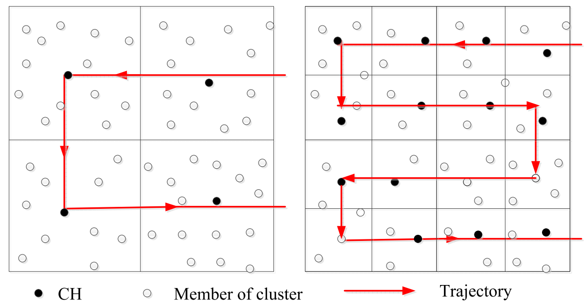


FIGURE 2. Trajectories for different size of clusters.

the energy expending formula is shown as follows:

$$E_i = E_{Tx}(l, d_{ij}) = \begin{cases} (E_{elec} + \varepsilon_{fs} \cdot d_{ij}^2) \cdot l & \text{if } d_{ij} < d_0 \\ (E_{elec} + \varepsilon_{amp} \cdot d_{ij}^4) \cdot l & \text{if } d_{ij} \geq d_0 \end{cases} \tag{1}$$

where d_{ij} stands for the distance between transmitter i and receiver j, E_{elec} indicates energy depletion of the electronic circuit. Parameters ε_{fs} and ε_{amp} denote the energy consumption of the amplifier in the free space and multi-path fading channel models, respectively. d_0 is equal to $\sqrt{\varepsilon_{fs}/\varepsilon_{amp}}$. To receive l-bit packets, the energy expending formula can be defined as:

$$E_{Rx}(l) = l \cdot E_{elec} \tag{2}$$

B. CLUSTERING

Supposed that sensor nodes are randomly deployed in the region of interest, we will introduce the mechanism of clustering for data collection in order to improve the energy-efficient performance of WSNs. In this paper, we explicitly consider two scenarios in the cluster for a MS to collect data, which exactly depends on the position where MS is going to stop. Actually, MS could move on to CH for efficient data collection or to the center of the cluster for energy optimization. However, the patten of clustering comes to affect the decision of MS's moving trajectory. We employ K-means clustering to partition WSNs into K clusters according to Euclidean distance and use the superiority of it to flexibly control the size of clusters. Actually, if K is large, the cluster size may be so smaller that MS's moving trajectory becomes longer. Basically, MS will take more time to collect data. On the contrary, if K is small, the cluster will include more sensor nodes and MS's moving trajectory could be shorten consequently. Seen from Fig.2, both of clustering and path of MS jointly play crucial roles to optimize and balance energy consumption. Therefore,the size of a cluster can be adjusted flexibly so that the performance of WSNs (e.g. the tradeoff between energy consumption and MS's moving trajectory assignment) are improved dramatically.

Different from [29], we adequately operate the generated cluster as many rounds as possible rather than frequently change it for keeping a single property of the cluster. In this way, we eventually partition WSNs into Voronoi uniform regions where MS goes through to collect data.

Note that 1) we set the size of cluster according to the energy model of data transmission Equation (1), that is, the data transmission ranges in the cluster include both of free space and multi-path fading channel models; 2) The value K definitely determines the MS's moving trajectory while changing the data transmission and energy consumption.

C. CLUSTER HEAD SELECTION

In this subsection, we select CH by combining the factors of node's residual energy and density. It is easy to understand the factor of node's residual energy. In terms of node density, we create the following definitions.

Definition 1 (Neighbor Node Set): For a given node i in the cluster and supposed that the size of cluster is L, the Neighbor Node Set of node i can be defined as:

$$\Gamma_i = \{Node_{ij} | d_{Node_{ij}} < d_0, j = 1, 2, \cdots, L\} \tag{3}$$

Definition 2 (Node's Density): Given the size of cluster L, Node i's density is calculated by

$$s^i_{density} = \frac{|\Gamma_i|}{L} \tag{4}$$

where the expression $|\Gamma_i|$ is the number of elements of set Γ_i.

In fact, the density property of node i in the cluster indicates a ratio of the number of Neighbor Node Set and the size of the cluster. In addition, the node with more residual energy also has an opportunity to be selected as CH. Combined residual energy and the definition of node density, the selection principle of candidate CHs is

$$Node_i = w\frac{E^i_{residual}}{E_{\max}} + (1-w)\frac{|\Gamma_i|}{L} \tag{5}$$

where $\omega \in [0, 1]$ is a weight coefficient, $E_{\max}$ indicates the maximal energy of node in the cluster, and $E^i_{residual}$ stands for the residual energy of node i.

After calculating all the candidate CHs' value, the CH is selected by maximizing the following objective function.

$$CH_i = \max\left(w\frac{E^i_{residual}}{E_{\max}} + (1-w)\frac{|\Gamma_i|}{L}\right) \tag{6}$$

The CH selection method aims to prevent edge nodes of cluster from being selected as CH, which will exacerbate the energy unbalance problem. CH can not only be used as backbone network to deliver data from source nodes for real-time quality of service, but also facilitate the design of MS' optimal trajectory to achieve energy-efficient data collection with latency. However, the topic of latency from introducing MS is beyond the scope of this paper and we will discuss it in the future. In this way,the energy-efficient data collection could be improved. On the other hand, we also consider energy distribution to minimize the dissipation energy of node as far as possible, thereby we have the definition of node density for CH to guarantee sufficient members, which play a role as the relay. Furthermore, the energy density could also contribute to energy balance designedly, which will be described in detail in the following part of this paper. The CH selection method is shown in Algorithm 1 by considering node's energy and density, which is built for a multi-hop cluster to enhance the energy-efficient performance of WSNs. where $\bar{\Gamma}_i$ indicates the complementary set of Γ_i in terms of cluster i in WSNs.

Algorithm 1 CH Election Algorithm

Initialization: Define K, a CH set $\Phi=\emptyset$ and density set $\Gamma=\emptyset$

for $i = 1$ toK **do**
  Difine $|i_K|$ is the size of i-th cluster
  for j=1 to $|i_K|$ **do**
    let $\Gamma_{ij} = \{Node_{ij} | R_{Node_{ij}} < d_0\}$
    Calculate the density of node j with radius $R = d_0$
    Compute $Node_{ij}$ using Equation (5)
    Broadcast a control packet including $Node_{ij}$
  end for
  Maximize $\{Node_i\}$ i.e. $CH_i = \max\{Node_i\}$
  Update CH set Φ by $\Phi \leftarrow CH_i$
  Update $\bar{\Gamma}_i$
end for
Broadcast a control packet including Φ

D. PREFERENCE RELATION IN CLUSTERS

According to CH selection based on Equation (6) and Equation (1), we assume that there must exist nodes whose range to CH is larger than d_0 in our cluster model. In order to transmit data to CH, they have to enhance power and inevitably consume more energy. Such phenomenon should be preventable in terms of energy balance, so we assign a node relay included in Neighbor Node Set of CH for the node exceeding range d_0. On the other hand, CH plays a role of leader to manege members of the cluster, which generally receives data from cluster members or other CHs and eventually transmits data to the Sink. So CHs usually undertake the responsible for WSNs as a backbone network, leading to consume more energy than cluster members. Consequently, we introduce MS to takes the responsibility for data collection so that workload of WSNs is mainly offloaded to MS.

Our proposed algorithms aim at collecting data in an energy-efficient manner with mobility. For this purpose, we design two different ways for MS to collect data. In the process of data collection, we actually adopt the traversal technology for MS to gather data from each cluster. As can be seen from Fig.3, sensory data is delivered to CH, and CH sends collected data to MS that travels and stops at CH's position. On the other hand, we primarily focus on mitigating long-distance energy consumption and unbalance. So we propose the definition of Virtual Head(VH) based on Neighbor Node Set to achieve equalization in terms of transmission distance. In this paper, we assume a two-hop transmission, that is, a source node transmits data to CH through a relay like the

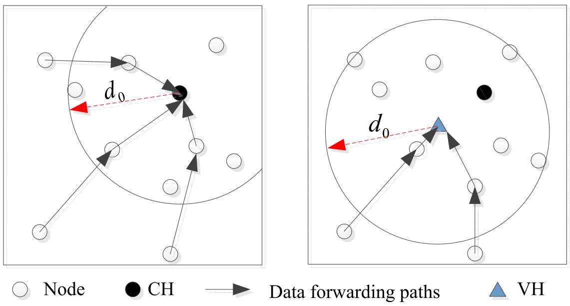


FIGURE 3. Two models of two-hop clustering data collection process by *MS*: CH-based data collection and VH-based data collection.

principle of power line communication transmission in [25]. A generalization to a multi-hop transmission is straightforward. Taking advantage of energy-efficient mobility of MS, it is noticeable that MS-based data collection strategies by CH or VH are performed to enhance the performance of WSNs, which are usually built for minimizing and balancing energy consumption.

IV. PRINCIPLES FOR RELAY SELECTION

Clustering has been employed as an effective approach for organizing the network to achieve energy efficiency. In this section, we first provide the properties of our models for cluster. Then, we analyze intra-cluster data relay in terms of energy efficiency.

A. PROPERTIES OF CLUSTER

According to the rationale that the intra-cluster routing load from cluster members to CH is minimized when CH tends to be located at the center of each cluster, we focus on the intra-cluster two-hop data transmission and build two strategies for data collection. As shown in Fig.3, one strategy is called CH-based data collection, which data are delivered to CH or transferred by a relay to CH. The other strategy called VH-based data collection will enable the rationale to enhance the performance of WSNs based on the CH-based data collection.

In order to reduce and balance energy consumption, MS is introduced to collect data by visiting CH or VH. Although CH and VH are proposed in the same framework, they play their respective roles and can be applied for different scenarios. For example, CH-based data collection strategy with or without MS can be modeled for random deployment WSNs in an inaccessible field to humans, while VH-based data collection strategy with or without CH can be applied for smart city where MS travels along street to gather data from sensor nodes deployed in blocks. Note that CH is not essential after initialization in VH-based strategy because MS can be responsibility for maintaining the basic information of cluster.

In this paper, we research two strategies for data collection in the perspective of energy efficiency. For both of them, we focus on the data relay communication in two-hop cluster.

B. RELAY NODES SELECTION

In fact, we should consider the situations about different transmission distances in terms of energy consumption. By investigating a wider transmission distance distribution, we specifically determine an optimization of energy consumption as far as the effect of transmission distance is concerned. For convenience, we can control the coverage area of the cluster with the distance $R \in [0, 2d_0]$.

In order to reduce energy consumption for two-hop cluster, we could exploit the relay node to transfer data. However, choosing appropriate quantities of relay nodes is difficult in terms of the optimization for energy efficiency. Based on the energy consumption model, we turn out the existence of the relay.

Theorem 1: For cluster i, there exists a point $Q \in \Gamma_i$ as a relay for one node in set $\bar{\Gamma}_i$ to reduce energy consumption.

Proof: Suppose that there exists one point Q such that the following inequalities $d_{node,Q} < d_0$, $d_{Q,CH} < d_0$ and $d_{node,CH} > d_0$ are hold. For convenience, E_{node} is the directly transmitted energy consumption of *node*. E_Q presents the transmitted and received energy consumption of Q, and $E_{node,CH}$ denotes the transmitted energy consumption between *node* and CH. They are calculated in detail as follows.

$$E_{node} = E_{elec} + \varepsilon_{fs} \cdot d^2_{node,Q} \tag{7}$$

$$E_Q = E_{Relec} + E_{Telec} + \varepsilon_{fs} \cdot d^2_{Q,CH} \tag{8}$$

$$E_{node,CH} = E_{Telec} + \varepsilon_{mp} \cdot d^4_{node,CH} \tag{9}$$

We will verify that the sum of E_{node} and E_Q is less than $E_{node,CH}$. Therefore, we combine E_{node} and E_Q to subtract $E_{node,CH}$:

$$\begin{aligned}\Delta E &= E_{node} + E_Q - E_{node,CH} \\ &= \varepsilon_{fs} \cdot (d^2_{node,Q} + d^2_{Q,CH}) + (E_{Relec} + E_{Telec}) \\ &\quad - \varepsilon_{mp} d^4_{node,CH}\end{aligned} \tag{10}$$

Due to $d_{node,Q}, d_{Q,CH} \in (0, d_0]$, $d_{node,CH} \in (d_0, +\infty)$, Equation (10) can be rewritten

$$\begin{aligned}\Delta E &= E_{node} + E_Q - E_{node,CH} \\ &= \varepsilon_{fs} \cdot (d^2_{node,Q} + d^2_{Q,CH}) + (E_{Relec} + E_{Telec}) \\ &\quad - \varepsilon_{amp} d^4_{node,CH} \\ &\leq 2\varepsilon_{fs} \cdot d_0^2 + (E_{Relec} + E_{Telec}) - \varepsilon_{mp} d^4_{node,CH}\end{aligned} \tag{11}$$

Owing to $\varepsilon_{fs} > 0, \varepsilon_{mp} > 0, d_0 > 0, E_{Relec} + E_{Telec} > 0$, which are all constants, there exists a $d_{node,CH} \in (d_0, +\infty)$ to make $\Delta E < 0$ held. So we prove the existence of the Q. □

For example, we assume every node transmits 4000-bit data. Let $d_0 = sqrt(\varepsilon_{fs}/\varepsilon_{mp}) = 87.7058$ and $d_{node,CH} = \sqrt{2}d_0$. According to Theorem 1, $E_{node} = 5.0769\text{e - }04$, $E_Q = 7.0769\text{e - }04$, and $E_{node,CH} = 0.0014$. So we have $\Delta E = E_{node} + E_Q - E_{node,CH} = -2.1538\text{e - }04 < 0$.

Considering all nodes in $\bar{\Gamma}$ to send data to CH or VH, we directly apply Theorem 1 to reduce the energy consumption of the cluster. So we have the following corollary and omit the proof.

TABLE 1. Notations.

Symbols	Statements
N	The number of nodes in set Γ
M	The number of nodes in set $\bar{\Gamma}$
T	The total number of time slot
t	$1 \leq t \leq n$ is any arbitrary time slot
m, n	User-relay pair
k_{opt}	A matching contains M user-relay pair
k^*_{opt}	A optimal expected matching contains M user-relay pair
$X_{m,n}(t)$	A reward that user m gets when m selects relay n at time t
$\mu_{m,n}$	The reward expectation of arm (m, n)
$\hat{\mu}_{m,n}$	Estimator of μ
$\mu^*_{m,n}$	The maximal value of $\mu_{m,n}$
δ	A confidence level
$b_{m,n}$	UCB index of user-relay pair at current time t.
$T_{m,n}(t-1)$	Observed samples from arm (m,n) at time t
$\Delta_{m,n}(t)$	$\mu^*_{m,n} - \mu_{m,n}$
Δ_{min}	$\min\limits_{k_{opt}} \Delta^{k_{opt}}_{m,n}$
Δ_{max}	$\max\limits_{k_{opt}} \Delta^{k_{opt}}_{m,n}$

Corollary 1: There exist a node subset of Γ to relay data from nodes in set $\bar{\Gamma}$ so that the energy consumption of the cluster is minimal.

Since choosing relay nodes to get a layover of transmitted data is challenging, so we have the following theorem.

Theorem 2: It is a NP-hard problem to select relay nodes in Γ for energy efficiency in the cluster.

Proof: Obviously, we can reduce the problem of selecting relay nodes into a knapsack problem, and here omits the process. □

V. MUTI-USER MAB FOR RELAY SELECTION

As shown in above sections, it is of importance to select energy-efficient relay nodes in the cluster to achieve optimization of energy consumption. In view of this, we adopt MU-MAB to perform the relay selection strategy while guaranteeing energy balance of the cluster.

A. NOTATIONS AND FORMULAS

In this section, we model the issue of relay selection as a MU-MAB problem. For convenience, we list usable notations in table 1.

In this paper, we take advantage of M nodes in set $\bar{\Gamma}$ to optimize relay nodes in set Γ including N nodes. According to the CH selection algorithm, we can control the scale N of set Γ, which meets the relationship $N \geq M$ shown in Fig.3. Suppose $m \in \{1, 2, \cdots, M\}$ and $n \in \{1, 2, \cdots, N\}$, we calculate the reward $X_{m,n}(t)$ by $|\Delta E|$ while assuming that $X_{m,n}(t)$ follows some unknown i.i.d. over time. Without loss of generality, we normalize $X_{m,n}(t) \in [0, 1]$. The mean of random variable $X_{m,n}(t)$ is $\theta_{m,n} = E_T[X_{m,n}(t)]$. We denote the set of all these means as $\Theta = \{\theta_{m,n}, 1 \leq m \leq M, 1 \leq n \leq K\}$. The performance of a relay selection is evaluated by the regret value, which is defined as the difference between the expected reward and that calculated by the selectable policy π [30]. Then we can obtain the mathematical formula after T time slots.

$$\Re^{\pi}(\Theta(t); T) = \sum_{t=1}^{T} \sum_{(m,n)\in k^*_{opt}(t)} EX_{m,n}(t) - E^{\pi}[\sum_{t=1}^{T} S_{\pi(t)}(t)] \quad (12)$$

where $S_{\pi(t)}(t)$ is the sum of rewards obtained by all users under policy $\pi(t)$, which is computed as:

$$S_{\pi(t)}(t) = \sum_{n=1}^{N} \sum_{m=1}^{M} X_{m,n}(t) \times I_{m,n}(t) \quad (13)$$

where $I_{m,n}(t) = 1$ when node m is the only one to select relay n, otherwise $I_{m,n}(t) = 0$.

B. FORMULATION OF MU-MAB

In this subsection, we consider the case where no prior reward distribution knowledge is provided throughout the relay selection process, however, we assume that the reward distribution remains constant during all games. The multi-user relay selection algorithm is proposed and the algorithm has a learning ability based on modified MAB in WSNs, which is built with the matching theory and an energy utility.

To evaluate the most promising relays, we employ a learning mechanism named upper confidence bound (UCB) to support an optimistic evaluation of the relay's quality, which associates an index referred to as UCB index for the user-relay pair. Then, we estimate the corresponding reward expectations via the computed index for each user-relay pair and choose the user-relay pair with the highest index.

In order to make this argument more precise, we need to define the UCB in our framework. To simplify the process, we assume $X_{m,n}(t)$ is a sequence of independent gaussian random variables with mean μ, variance $\sigma^2 = 1$. Given an observed sequence $\{X_1, X_2, \cdots, X_T\}$, we would like to estimate the mean μ. Actually, the most natural estimator can be calculated by the expression $\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} X_i$. Owing to $\mu - \hat{\mu} \sim N(0, \frac{\sigma^2}{\sqrt{T}})$, and then we have

$$P\left(\mu - \hat{\mu} \geq \sqrt{\frac{2\log(1/\delta)}{T}}\right) \leq \delta \quad (14)$$

where $\delta \in (0, 1)$. We assume that the learner has observed $T_{m,n}(t-1)$ samples from one arm at time t and received rewards from the arm (m, n) with an empirical mean of $\hat{\mu}_{m,n}(t-1)$. Then, the unknown mean of the arm can be expressed by

$$UCB_{m,n}(t-1) = \hat{\mu}_{m,n}(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_{m,n}(t-1)}} \quad (15)$$

Under the description in III-D, there are sufficient relay nodes in set Γ to meet energy-efficient policies. However, due to $P(N, M)$ arms, it is prohibitively expensive to solve this maximization via using exhaustive search, which actually refer to the problem of combinational optimization.

In order to solve the problem of combinational optimization, we employ the matching theory to design policies for this type of MAB problem with respect to regret. After that, we map the optimization of matching into an arm.

Compared with UCB index in Equation (15), an arm is determined through the optimization of the matching strategy and the corresponding UCB index is observed in each time as follow:

$$\sum_{(m,n)\in k_{opt}(t)} \hat{\mu}_{m,n}(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_{m,n}(t)}} \tag{16}$$

where $\hat{\mu}_{m,n}(t) = \frac{\hat{\mu}_{m,n}(t-1)T_{m,n}(t)+X_{m,n}(t)}{T_{m,n}(t-1)+1}$ and $T_{m,n}(t) = T_{m,n}(t-1)+1$. Our scheme selects M user-relay pairs with the maximum value $b(k_{opt})$ at each time slot after initialization period, and this optation is carried out once. For convenience, we denote the UCB index by $b_{m,n}$, and define the total UCB index of a matching strategy k_{opt} by

$$b(k_{opt}) := \sum_{m=1}^{M} b_{m,k_{opt}(t)} \tag{17}$$

C. MARGINAL-BASED MATCHING OPTIMIZATION

The main goal of matching is to optimally match relays and users, if their individual and learned information are observed. Then, each source node named user ranks relays by using a preference relation referred to as UCB index, which predicts the energy efficiency achieved by marginal utility function. Now we can define a matching stability when no user-relay pairs prefer each other in comparison to their current matching. Hence, the definition of stability is shown as follows:

Definition 3 (Stability [31]): A matching $S : [M] \to [K]$ is stable if for every $m \in [M]$ and $n \in [N]$ satisfying $S(m) \neq n$ if $b_{m,S(m)} < b_{m,n}$ then there exits some user $m' \in [M]$ such that $S(m') = n$ and $b_{m',n} < b_{m,n}$.

To obtain a stable matching strategy, we design a definition of utility function which have the property of marginal rule to select relay nodes.

Definition 4 (Marginal Utility Function): A utility function about energy is presented:

$$f_t(\Delta E_i) = \sum_{i=1}^{M} (E^i_{node}(t) + E^i_Q(t) - E^i_{node,CH}(t)) \tag{18}$$

where t indicates some time and M is the number of relay nodes in a cluster under the assumption of $n \geq 1$, $E^i_{node}(t)$, $E^i_Q(t)$ and $E^i_{node,CH}(t)$ are from expression(11).

Due to $E^i_{node}(t)+E^i_Q(t)-E^i_{node,CH}(t)$, we can create Equation (18) as a monotonically decreasing function through selecting relay nodes if the following inequality is held.

$$\Delta E_{i+1} - \Delta E_i < \Delta E_i - \Delta E_{i-1} \tag{19}$$

Ranking the sequence $\{\Delta E_{i+1} - \Delta E_i\}$ at time t, we could get a relay selection strategy $\min(f_t(\Delta E_{i+1}) - f_t(\Delta E_i))$.

Algorithm 2 Marginal-Baed Matching Optimization Algorithm

Input: M,N,
Output: k_{opt}

```
Calculate UCB index
for i = 1 toM do
    for j = 1 toN do
        if ΔE_{i+1}) + f_(t)(ΔE_{i-1}) < 2f_(t)(ΔE_i) then
            Update  Σ_{(m,n)∈k_opt(t)} μ̂_{m,n}(t − 1) + sqrt(2 log(1/δ) / T_{m,n}(t))
        end if
    end for
end for
```

Algorithm 3 MU-MAB Algorithm

```
for t = 1 → T do
    for m = 1 → M do
        Calculate X_{m,n}(t)
    end for
end for
while l do
    t = T + 1
    Run algorithm 2 to get UCB of an arm
    b(k_opt) := Σ_{m=1}^{M} b_{m,k_opt(t)}
end while
```

Through analyzing the property of the above utility function, we definitely calculates UCB index $b_{m,n}$ and maps it into a monotonically decreasing function with a user-relay pair.

Theorem 3: The marginal utility function satisfying inequality (19) is stable.

Proof: Obviously, the marginal utility function $f_t(\cdot) < 0$ guarantees the declining tendency with increment of $i \in \{1, 2, \ldots, M\}$. Next, we look for relay nodes in the cluster according to the following inequality,

$$f_t(\Delta E_{i+1}) + f_t(\Delta E_{i-1}) < 2f_t(\Delta E_i) \tag{20}$$

Owing to the property of declining tendency, inequality (18) accord with diminishing marginal utility with $i \in \{1, 2, \ldots, M\}$ increasing constantly. Therefor the following equality is satisfied mathematically.

$$\lim_{n\to+\infty} f_t(\Delta E_{i+1}) - f_t(\Delta E_i) = 0 \tag{21}$$

Generally, as long as we should guarantee the marginal increment between $f(\Delta E_{i+1})$ and $f(\Delta E_i)$ is minimum, that is, $\min(f_t(\Delta E_{i+1}) - f_t(\Delta E_i))$, there must be no $m \in [M]$ and $n \in [N]$ to match a smaller value of $b_{m,k}$. Therefore, we get the conclusion. □

According to the above analysis, we design an optimization algorithm based on MU-MAB to select relay nodes through a marginal-based matching strategy as shown in Algorithm 2 and 3.

VI. REGRET ANALYSIS

Traditionally, the regret of a policy for a multi-armed bandit problem is upper-bounded by analyzing the expected number of times that each non-optimal arm is played. In this paper, our work is influenced in [32] and references therein.

Theorem 4: For any horizon T, if $\delta = 1/T^2$, then the expected regret is at most

$$3MN\Delta_{\max} + \frac{16MN\log(T)}{\Delta_{\min}} \tag{22}$$

Proof: According to Equation (12), we change regret into the following expression.

$$R_T = \sum_{m=1}^{M}\sum_{n=1}^{N}\Delta_{ij}E[T_{m,n}(T)] \tag{23}$$

In order to bounding $E[T_{m,n}(T)]$ for each arm(m, n), we define a collection $G_{m,n}$ for convenience.

$$G_{m,n} = \{\mu^*_{m,n} < \min_{t\in[T]} UCB_1(t)\} \cap \{\bar{\mu}_{(m,n)u_{m,n}} + \sqrt{\frac{2}{u_{m,n}}\log(1/\delta)} < \mu^*_{m,n}\} \tag{24}$$

where $u_{m,n}$ indicates observations of arm (m, n) and can be solve as a constant. In this way, we can conclude $E[T_{m,n}(T)]$ according to the definition of expectation.

$$\begin{aligned} E[T_{m,n}(T)] &= E[I\{G_{m,n}\}T_{m,n}(T)] + E[I\{G^c_{m,n}\}T_{m,n}(T)] \\ &\le u_{m,n} + P(G^c_{m,n})n \end{aligned} \tag{25}$$

Naturally, we only verify that $T_{m,n}(T) \le u_{m,n}$ and $P(G_{m,n})^c$ is small enough with regard to T. First, we use contradiction method to proof $T_{m,n} \le u_{m,n}$. Suppose that $T_{m,n}(T) > u_{m,n}$ which means arm (m, n) was played more than $u_{m,n}$ times over T rounds. Then there exists a round $t \in [T]$ to get $T_{m,n}(t-1) = u_{m,n}$ and $A_t = (m, n)$ held.

$$\begin{aligned} UCB_{m,n}(t-1) &= \bar{\mu}_{m,n}(t-1) + \sqrt{\frac{2}{T_{m,n}(t-1)}\log(1/\delta)} \\ &= \bar{\mu}_{(m,n)u_{m,n}} + \sqrt{\frac{2}{u_{m,n}}\log(1/\delta)} \\ &< \mu^*_{ij} \\ &< UCB_1(t-1) \end{aligned} \tag{26}$$

Therefore, there indeed exist $A_t = \arg\max_{pq} UCB_{pq}(t-1) \neq (m, n)$. This is a contradiction and then we conclude $T_{m,n}(T) \le u_{m,n}$. Next, we concern the second part $P(G^c_{m,n})$ of inequality (25). According to the definition of $G_{m,n}$, it is easy to get the complement set $P(G^c_{m,n})$ defined as follows.

$$G^c_{m,n} = \{\mu^*_{m,n} \ge \min_{t\in[T]} UCB_1(t)\} \cup \{\bar{\mu}_{(m,n)u_{m,n}} + \sqrt{\frac{2}{u_{m,n}}\log(1/\delta)} \ge \mu^*_{m,n}\} \tag{27}$$

In Equation (27), we decompose the first part.

$$\begin{aligned} &\{\mu^*_{m,n} \ge \min_{t\in[T]} UCB_1(t)\} \\ &\subset \{\mu^*_{m,n} \ge \min_{s\in[T]} \bar{\mu}_{(m,n)s} + \sqrt{\frac{2}{s}\log(1/\delta)}\} \\ &= \bigcup_{s\in[T]}\{\mu^*_{m,n} \ge \bar{\mu}_{(m,n)s} + \sqrt{\frac{2}{s}\log(1/\delta)}\} \end{aligned} \tag{28}$$

Due to random variables are independent, we theoretically infer the following inequalities.

$$\begin{aligned} &P(\mu^*_{m,n} \ge \min_{t\in[T]} UCB_1(t)) \\ &\le P(\bigcup_{s\in[T]}\{\mu^*_{m,n} \ge \bar{\mu}_{(m,n)s} + \sqrt{\frac{2}{s}\log(1/\delta)}\}) \\ &\le \sum_{s=1}^{T} P(\mu^*_{m,n} \ge \bar{\mu}_{(m,n)s} + \sqrt{\frac{2}{s}\log(1/\delta)}) \\ &\le T\delta \end{aligned} \tag{29}$$

For the second part of Equation (27), we mainly confirm the value of $u_{m,n}$. Due to the definition of $\Delta_{m,n}$ and the relationship between Δ_{ij} and u_{ij}, we obtain an simplified format only about $\Delta_{m,n}$, which could be figured out here.

$$\Delta_{m,n} - \sqrt{\frac{2}{u_{m,n}}\log(1/\delta)} \ge c\Delta_{m,n} \tag{30}$$

where $c \in (0, 1)$. Using Hoeffding's bound, we have the following deduction.

$$\begin{aligned} &P(\bar{\mu}_{(m,n)u_{m,n}} + \sqrt{\frac{2}{u_{m,n}}\log(1/\delta)} \ge \mu^*_{ij}) \\ &= P(\bar{\mu}_{(m,n)u_{m,n}} - \mu_{m,n} \ge \Delta_{m,n} - \sqrt{\frac{2}{u_{m,n}}\log(1/\delta)}) \\ &\le P(\bar{\mu}_{(m,n)u_{m,n}} - \mu_{m,n} \ge c\Delta_{m,n}) \\ &\le \exp(-\frac{u_{m,n}c^2\Delta^2_{m,n}}{2}) \end{aligned} \tag{31}$$

Taking expression (31)together with Inequality (29), we obtain $P(G^c_{m,n})$ as follows.

$$P(G^c_{m,n}) \le n\delta + \exp(-\frac{u_{m,n}c^2\Delta^2_{m,n}}{2}) \tag{32}$$

We then substitute Inequality (32) into Inequality (25), and the result is

$$E[T_{m,n}(T)] \le u_{m,n} + T(T\delta + \exp(-\frac{u_{m,n}c^2\Delta^2_{m,n}}{2})) \tag{33}$$

What we need to solve in Inequality (33) is $u_{m,n}$. According to Inequality (30), we can get a smallest integer $u_{m,n} = \left\lceil \frac{2\log(1/\delta)}{(1-c)^2\Delta^2_{ij}} \right\rceil$. Then setting $\delta = 1/T^2$ and $c = 0.5$, we have

$$\begin{aligned} E[T_{m,n}(T)] &\le \left\lceil \frac{2\log(1/\delta)}{(1-c)^2\Delta^2_{m,n}} \right\rceil + T^{1-\frac{2c^2}{(1-c)^2}} + 1 \\ &\le 3 + \frac{16\log(T)}{\Delta^2_{m,n}} \end{aligned} \tag{34}$$

TABLE 2. Simulation parameters.

parameters	values	parameters	values
α	2/4	ε_{fs}	$10\text{J/bit}/m^2$
E_{elec}	50nJ/bit	ε_{amp}	$0.0013\text{PJ/bit } m^4$
E_0	0.5J	n_{num}	100
$Area$	300m × 300m		

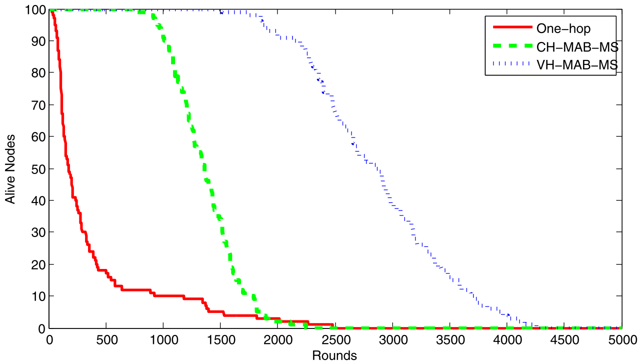


FIGURE 4. The lifetime of network when $k = 2$.

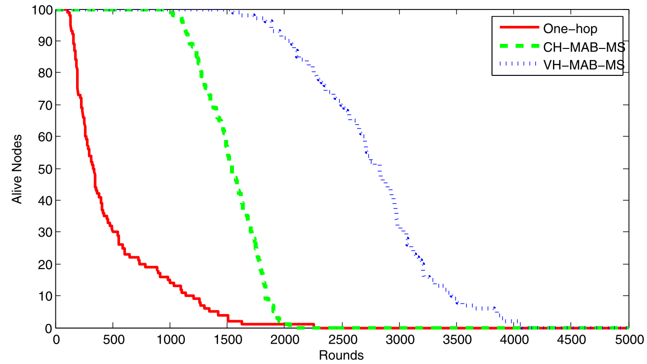


FIGURE 5. The lifetime of network when $k = 4$.

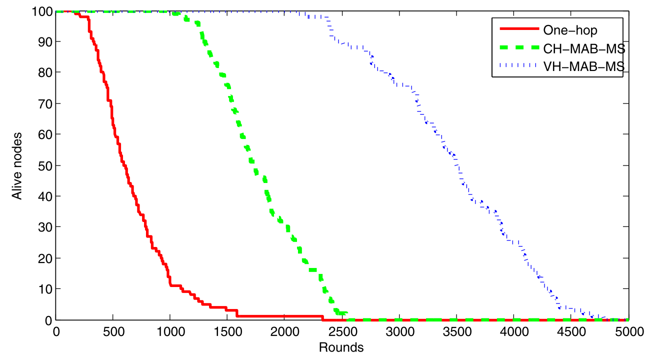


FIGURE 6. The lifetime of network when $k = 6$.

Therefore, the result of expected regret R_T follows

$$\begin{aligned} R_T &= \sum_{m=1}^{M}\sum_{n=1}^{N}\Delta_{m,n}E[T_{m,n}(T)] \\ &\leq \sum_{m=1}^{M}\sum_{n=1}^{N}\Delta_{m,n}(3+\frac{16\log(T)}{\Delta_{m,n}^2})] \\ &\leq 3MN\Delta_{\max}+\frac{16MN\log(T)}{\Delta_{\min}} \end{aligned} \tag{35}$$

□

VII. SIMULATIONS

In order to evaluate the energy efficiency of network performances, we carry out our proposed algorithms in the section.

A. PERFORMANCE COMPARISONS

We start from two WSNs frameworks named CH-MAB-MS and VH-MAB-MS respectively, and then simulate effects of traffic using the marginal-based matching theory for sensor relay to further verify proposed algorithms. We carry out simulations in different scenarios using MATLAB. For the sake of comparing to several critical indicators, we simultaneously refer to the traditional algorithm as One-hop which describes the data delivery with one-hop patten in the cluster. As expressed in the above sections, main parameters referred to energy efficiency in simulations are listed in table 2, which is used in a classical one-hop clustering algorithm(e.g. LEACH).

We consider the implementation of running time for lifetime based on two frameworks illustrated in Section III with K-mean cluster, where $K = 2$, $K = 4$ and $K = 6$ are designed. In the figures from Fig.4 to Fig.6, we compare our two metrics CH-MAB-MS and VH-MAB-MS with the benchmark named One-hop for the death of the first node in running WSNs.

Seen from Fig.4, the lifetime of benchmark One-hop is less than 100 rounds obviously. On one hand, long-distance transmission incurs much more energy consumption owing to the clustering scale with $K = 2$. On the other hand, sensory data from cluster members are delivered to CH, which increases the CH's traffic burden in spite of the rotating mechanism being adopted. In essence, our algorithms CH-MAB-MS and VH-MAB-MS manifest superior performances by introducing the CH selection mechanism, the mobility for energy efficiency and relay selection strategies. In terms of CH-MAB-MS, CH selection mechanism guarantees that no node at the edge of cluster is selected as CH, which avoids the increase of energy consumption in the cluster. MS traverses to CHs for data collection instead of directly or indirectly delivering data from CH to the Sink. In this way, CH's traffic burden is dramatically decreased to reduce the risk of energy depletion, especially for the application of raw data collection. More importantly, we employ MU-MAB with the marginal-based matching theory to reduce the energy consumption for prolonging the lifetime of WSNs. Due to optimality of MU-MAB for relay selection, the lifetime of WSNs is tremendously promoted. For VH-MAB-MS, we can see that the performance of the lifetime is optimal. This is because VH-MAB-MS is different from CH-MAB-MS in the process of data collection with MS, which stops at VH and gathers data from cluster members rather than CH.

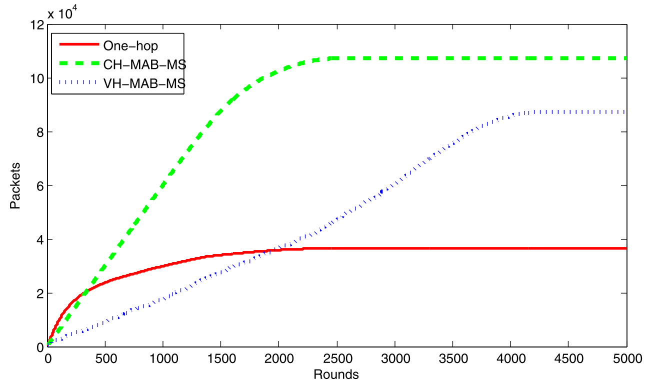


FIGURE 7. Packets of network when $k = 2$.

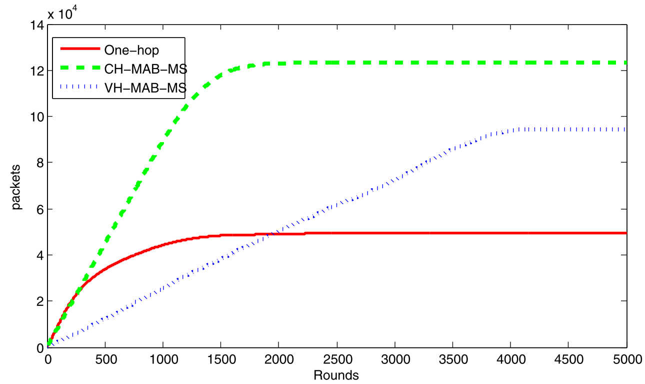


FIGURE 8. Packets of network when $k = 4$.

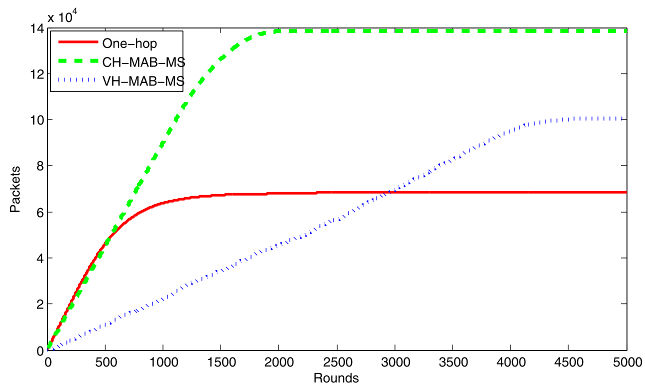


FIGURE 9. Packets of network when $k = 6$.

We can also get the similar results in Fig.5 and Fig.6 with $K = 4$ and $K = 6$, respectively. However, all the results of One-hop, CH-MAB-MS and VH-MAB-MS outperform what illustrated in Fig.4. The primary cause is that shorter transmission distance contributes to less energy consumption, which the cluster scale becomes smaller than the one of the case $K = 2$.

In order to study the energy efficiency from the perspective of network traffic, we consider packets generated and delivered by the three algorithms with different clustering scenarios. Seen from Fig.7 to Fig.9, the traffic tends to

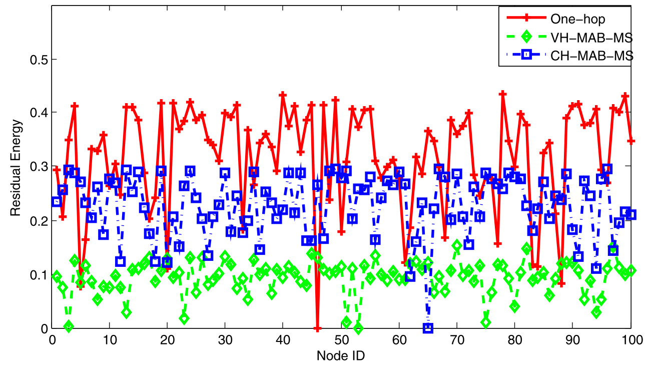


FIGURE 10. Residual energy when $k = 4$.

improvement with K increasing from 2 to 6. Intuitively, prolonging the lifetime of networks is an efficient method to achieve considerable packets, which necessitates optimization of network resources. Specifically, the death of first node play an important roles in the process of analyzing data traffic.

As shown in Fig.7, the total quantity of packets transmitted by One-hop algorithm is rather quiet small comparing to our two proposed algorithms. According to the above description of the lifetime, One-hop algorithm leads to a shorter lifetime so that data traffic is limited. However, the quantity of packets transmitted by One-hop algorithm is more than CH-MAB-MS's and VH-MAB-MS's before the first node comes to death. In terms of CH-MAB-MS, data are transmitted to MS rather than the Sink. For VH-MAB-MS, all data from cluster members are sent to MS dwelling at VH, which dramatically reduces the data traffic to spread in WSNs.

For further illustrating the statistical characters, we also simulate other scenarios with $K = 4$ and $K = 6$ to verify the rule of data traffic in Fig.8 and Fig.9. In fact, sustainable and sufficient data created by nodes can ensure a favorable quality of service. Therefore, our proposed algorithms CH-MAB-MS and VH-MAB-MS can provide a better data service than the One-hop algorithm in terms of advancing data supply.

In Fig.10, we demonstrate the residual energy of all nodes involved in the cluster when the death of the first node appears and $K = 4$. The scenarios of $K = 2$ and $K = 6$ have the similar property, so we omit the analysis of them. Seen from Fig. 10, due to the unbalanced energy consumption of the One-hop algorithm, most nodes remain major energy with 0.35J on average when the first node is dead. CH-MAB-MS algorithm's and VH-MAB-MS algorithm's are 0.25J and 0.1J, respectively. From the perspective of lifetime, the more most nodes remain energy, the poorer WSNs have performance. Actually, the traditional One-hop algorithm has tremendous residual energy with about more than 80% of the original energy. Contrasted to CH-MAB-MS and VH-MAB-MS, it only consumes a few energy and vulnerably makes networks disconnected, which leads to an enormous waste of resource. Although we introduce mobility for data collection to save energy in CH-MAB-MS algorithm, data from cluster members are still sent to CH and then MS.

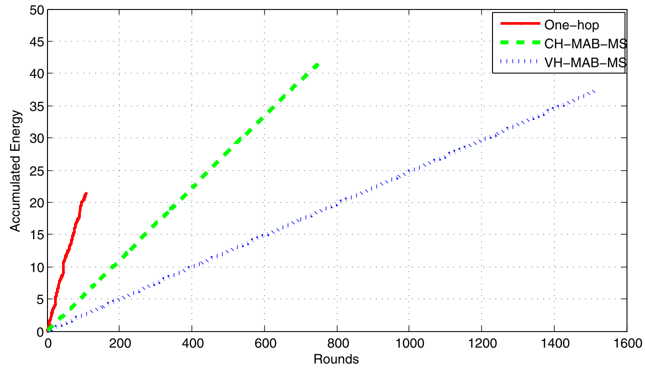


FIGURE 11. Accumulated energy consumption when $k = 2$.

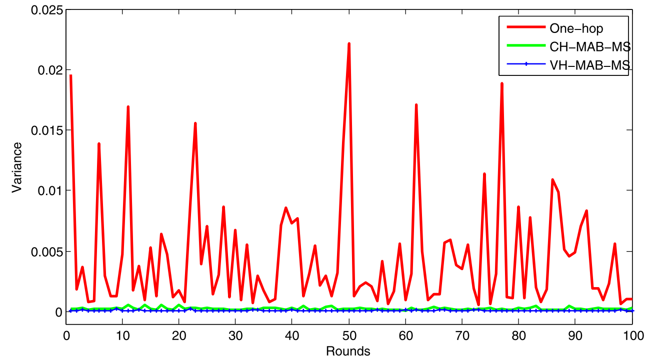


FIGURE 12. Energy variances when $k = 4$.

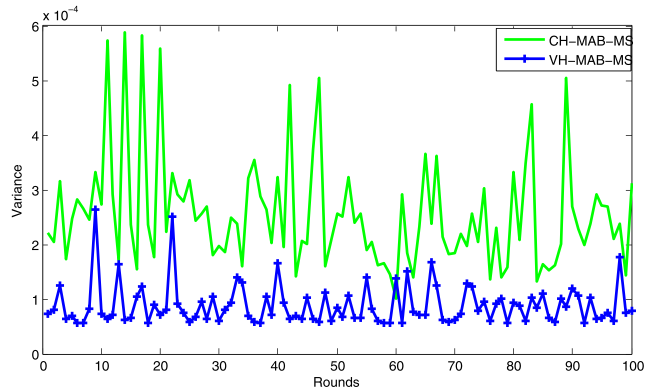


FIGURE 13. Energy variances when $k = 4$.

In this process, CH dominates the lifetime of network owing to higher energy consumption than cluster members. The VH-MAB-MS algorithm utilizes MS to collect data from cluster members instead of CH so that energy consumption of every cluster member is taken full advantage. In addition, the position of VH is always optimal to minimize the energy consumption, which makes the residual energy of VH-MAB-MS algorithm superior to CH-MAB-MS algorithm and the traditional One-hop algorithm.

For energy efficiency, we consider the accumulated energy consumption of WSNs when $K = 2$ occurs in Fig.11. Based on the death of the first node, the curve of accumulated energy consumption for the traditional One-hop algorithm is steep, because the energy of WSNs runs out in a shorter time. On the contrary, our proposed CH-MAB-MS algorithm and VH-MAB-MS algorithm take on flat performances. In other words, our proposed algorithms precede the traditional One-hop algorithm in total energy consumption, which means that they have good properties to prolong the lifetime of WSNs. In terms of our VH-MAB-MS algorithm, it is energy efficient for delay-tolerant events to collect data owing to constraints from the mobility of MS. It is also for this reason that the VH-MAB-MS algorithm consumes the least energy among the three algorithms in unit time.

Fig.12 depicts the energy variance with $K = 4$ and Fig.13 shows an elaborate specification between CH-MAB-MS algorithm and VH-MAB-MS algorithm. Due to variation of transmission distance distribution for the traditional One-hop algorithm, the energy variance fluctuates from 0.000572 to 0.002217, which reveals the fact that one-hop data collection strategy and CH selection mechanism determine the energy consumption. In Fig.13, energy variance of CH-MAB-MS algorithm is from 0.000101 to 0.000587. The reason is that we design data relay strategies to balance the energy consumption between all nodes in the cluster. From the perspective of CH, the selection strategy is given to reduce and balance energy consumption of CH according to weighted coefficient between the residual energy and the definition of density. For balancing the energy consumption among cluster members, the data relay strategy with MU-MAB achieves fairness by including the marginal rule to filtrate the potential optimal relay and the matching theory to determine the optimal solution. Similarly, energy variance of VH-MAB-MS algorithm varies from 0.000057 to 0.000265. In this way, VH-MAB-MS is more efficient in energy consumption than CH-MAB-MS. Therefore, we conclude that introducing mobility and optimizing cluster structure are energy efficient for data collection in WSNs.

B. DISCUSSION ON MAB REGRETS

In this section, we analyze the regret of MU-MAB applied for CH-MAB-MS algorithm and VH-MAB-MS algorithm in two scenarios. One is the predetermined relationship between source nodes and relay nodes in the cluster, the other is the stochastic application in a randomly deployed cluster.

Firstly, we discuss the former scenario. For convenience, we denote M as the number of source nodes(users) and N as the number of relay nodes(relays). We simulate 5 groups between users and relays, namely, $M = 1$ and $N = 4$, $M = 2$ and $N = 5$, $M = 3$ and $N = 6$, $M = 3$ and $N = 9$, $M = 4$ and $N = 8$. In Fig.14, the regret of VH-MAB-MS is less than that of CH-MAB-MS in the same cluster scale, which demonstrates VH-MAB-MS has uniformity in terms of Δ_{max} and Δ_{min} according to expression (24). Furthermore, CH-MAB-MS and VH-MAB-MS reflect the stability of regrets for different users and relays, that is, Δ_{max} and Δ_{min} vary little while M and N dominate the regret. The reason

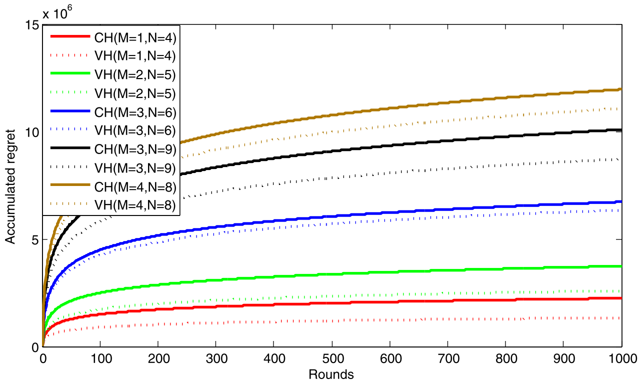


FIGURE 14. Accumulated regrets for the predetermined cluster.

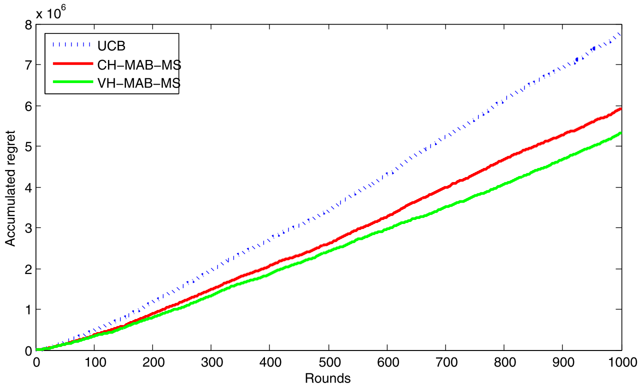


FIGURE 15. Accumulated regrets for the randomly deployed cluster.

is that the marginal-based matching theory contributes to the relay selection in the optimal solution.

Next, we investigate the latter one which illustrates changed relationship between users and relays. In Fig.15, owing to CH selection mechanism in Equation (6), M and N are constantly changed for CH-MAB-MS and VH-MAB-MS, so the accumulated regrets calculated with rounds show the figure of irregular stairs. In this process of selecting relay for energy efficiency, MU-MAB with the marginal-based matching theory makes the regret less than UCB, which means CH-MAB-MS and VH-MAB-MS have better performance for relay selection. Therefore, our proposed algorithms consume less energy and keep better performances in terms of energy efficiency.

VIII. CONCLUSION

To enhance energy efficiency of hierarchical WSNs, we exploit MU-MAB technology to solve the relay selection dilemma. Firstly, we build WSNs framework via K-mean method to create clusters including the long-distance transmission. In order to optimize and balance cluster's energy consumption, we design a CH selection mechanism with MS to collect data, which can control the proportion between long-distance transmission nodes and short-distance transmission nodes. Based on CH, we propose VH for MS to collect data in terms of energy efficiency. Secondly, we programme our hierarchical WSNs as a cooperative power line communication mechanism due to long-spanned transmission distance and nonuniform transmission distance distribution in the cluster. Thirdly, We propose the machine learning algorithm, namely MU-MAB, to solve the permutation problem for relay selection, which maybe incurs potential computation explosion as the increase in the quantity of nodes. Furthermore, we employ stable matching theory with marginal utility rule to allocate the final one-to-one optimal combinations for achieving energy efficiency of WSNs. At last, we simulate our proposed solutions in terms of saving and balancing energy consumption while evaluating the regret of MU-MAB.

REFERENCES

[1] D. Ebrahimi, S. Sharafeddine, P. H. Ho, and C. Assi, "UAV-aided projection-based compressive data gathering in wireless sensor networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1893–1905, Apr. 2019.

[2] L. M. Borges, F. J. Velez, and A. S. Lebres, "Survey on the characterization and classification of wireless sensor network applications," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1860–1890, 4th Quart., 2014.

[3] F. Fanian and M. K. Rafsanjani, "Cluster-based routing protocols in wireless sensor networks: A survey based on methodology," *J. Netw. Comput. Appl.*, vol. 142, pp. 111–142, Sep. 2019.

[4] Z. Xu, L. Chen, C. Chen, and X. Guan, "Joint clustering and routing design for reliable and efficient data collection in large-scale wireless sensor networks," *IEEE Internet Things J.*, vol. 3, no. 4, pp. 520–532, Aug. 2016.

[5] M. Zhao, Y. Yang, and C. Wang, "Mobile data gathering with load balanced clustering and dual data uploading in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 770–785, Apr. 2015.

[6] Y. Liu, K.-Y. Lam, S. Han, and Q. Chen, "Mobile data gathering and energy harvesting in rechargeable wireless sensor networks," *Inf. Sci.*, vol. 482, pp. 189–209, May 2019.

[7] R. Zhang, J. Pan, D. Xie, and F. Wang, "NDCMC: A hybrid data collection approach for large-scale WSNs using mobile element and hierarchical clustering," *IEEE Internet Things J.*, vol. 3, no. 4, pp. 533–543, Aug. 2016.

[8] K. Arthi and A. S. R. Lochana, "Zone-based dual sub sink for network lifetime maximization in wireless sensor network," *Cluster Comput.*, vol. 22, no. S6, pp. 15273–15283, Nov. 2019.

[9] P. Nayak and B. Vathasavai, "Energy efficient clustering algorithm for multi-hop wireless sensor network using type-2 fuzzy logic," *IEEE Sensors J.*, vol. 17, no. 14, pp. 4492–4499, Jul. 2017.

[10] N. Gharaei, K. A. Bakar, S. Z. M. Hashim, A. H. Pourasl, and S. A. Butt, "Collaborative Mobile Sink Sojourn Time Optimization Scheme for Cluster-Based Wireless Sensor Networks," *IEEE Sensors J.*, vol. 18, no. 16, pp. 6669–6676, Aug. 2018.

[11] C. Zhan, Y. Zeng, and R. Zhang, "Energy-efficient data collection in UAV enabled wireless sensor network," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 328–331, Jun. 2018.

[12] P. Sun and A. Boukerche, "Performance modeling and analysis of a UAV path planning and target detection in a UAV-based wireless sensor network," *Comput. Netw.*, vol. 146, pp. 217–231, Dec. 2018.

[13] N. Gharaei, K. Abu Bakar, S. Z. M. Hashim, and A. H. Pourasl, "Inter- and intra-cluster movement of mobile sink algorithms for cluster-based networks to enhance the network lifetime," *Ad Hoc Netw.*, vol. 85, pp. 60–70, Mar. 2019.

[14] A. Wang, J. Shen, P. Vijayakumar, Y. Zhu, and L. Tian, "Secure big data communication for energy efficient intra-cluster in WSNs," *Inf. Sci.*, vol. 505, pp. 586–599, Dec. 2019.

[15] A. Wang, J. Shen, C. Wang, H. Yang, and D. Liu, "Anonymous data collection scheme for cloud-aided mobile edge networks," *Digit. Commun. Netw.*, to be published, doi: 10.1016/j.dcan.2019.04.001.

[16] M. Wu, L. Tan, and N. Xiong, "Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications," *Inf. Sci.*, vol. 329, pp. 800–818, Feb. 2016.

[17] D. Zhang, H. Ge, T. Zhang, Y.-Y. Cui, X. Liu, and G. Mao, "New multi-hop clustering algorithm for vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1517–1530, Apr. 2019.

[18] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[19] D. Phanish and E. J. Coyle, "Application-based optimization of multi-level clustering in ad hoc and sensor networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4460–4475, Jul. 2017.

[20] K. Liu, J. Peng, L. He, J. Pan, S. Li, M. Ling, and Z. Huang, "An active mobile charging and data collection scheme for clustered sensor networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5100–5113, May 2019.

[21] S. Yang, U. Adeel, Y. Tahir, and J. A. McCann, "Practical opportunistic data collection in wireless sensor networks with mobile sinks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1420–1433, May 2017.

[22] A. Wichmann, T. Korkmaz, and A. S. Tosun, "Robot control strategies for task allocation with connectivity constraints in wireless sensor and robot networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 6, pp. 1429–1441, Jun. 2018.

[23] M. Abo-Zahhad, S. M. Ahmed, N. Sabor, and S. Sasaki, "Mobile sink-based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4576–4586, Aug. 2015.

[24] A. Gomez, M. Magno, M. F. Lagadec, and L. Benini, "Precise, energy-efficient data acquisition architecture for monitoring radioactivity using self-sustainable wireless sensor nodes," *IEEE Sensors J.*, vol. 18, no. 1, pp. 459–469, Jan. 2018.

[25] B. Nikfar and A. J. H. Vinck, "Relay selection in cooperative power line communication: A multi-armed bandit approach," *J. Commun. Netw.*, vol. 19, no. 1, pp. 1–9, Feb. 2017.

[26] Y. Lin, T. Wang, and S. Wang, "UAV-assisted emergency communications: An extended multi-armed bandit perspective," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 938–941, May 2019.

[27] A. Mukherjee, S. Misra, V. S. P. Chandra, and M. S. Obaidat, "Resource-optimized multiarmed bandit-based offload path selection in edge UAV swarms," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4889–4896, Jun. 2019.

[28] X. Li, J. Liu, L. Yan, S. Han, and X. Guan, "Relay selection for underwater acoustic sensor networks: A multi-user multi-armed bandit formulation," *IEEE Access*, vol. 6, pp. 7839–7853, 2018.

[29] L. Farzinvash, S. Najjar-Ghabel, and T. Javadzadeh, "A distributed and energy-efficient approach for collecting emergency data in wireless sensor networks with mobile sinks," *AEU-Int. J. Electron. Commun.*, vol. 108, pp. 79–86, Aug. 2019.

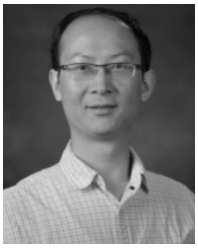
[30] P. Auer, N. C. Bianchi, and P. Fischer, "Finite-time analysis of the multi-armed bandit problem. Machine learning," *Mach. Learn.*, vol. 47, nos. 2–3, pp. 235–256, May 2002.

[31] A. Leshem, E. Zehavi, and Y. Yaffe, "Multichannel opportunistic carrier sensing for stable channel access control in cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 82–95, Jan. 2012.

[32] T. Lattimore and C. Szepesvari. (2019). *Bandit Algorithms*. [Online]. Available: https://tor-lattimore.com/downloads/book/book.pdf

JIAN ZHANG received the M.S. degree in applied mathematics from Liaoning University, in 2008, and the Ph.D. degree in pattern recognition and intelligent systems from Northeastern University, China, in 2012. Since 2012, he has been working at the Nanjing University of Information Science and Technology, China. His research interests include wireless sensor networks, edge computing, and target tracking.

JIAN TANG received the bachelor's degree from the Naval College of Engineering, Wuhan, China, in 1998, and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University, in 2006 and 2012, respectively. Since 2017, he has been working at the Beijing University of Technology, China. His current research interest is intelligent control and modeling for complex industrial processes.



FENG WANG (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in computer science from Tsinghua University, Beijing, China, in 2002 and 2005, respectively, and the Ph.D. degree in computer science from Simon Fraser University, Burnaby, BC, Canada, in 2012. His research interests are mainly in computer networking area, in particular, wireless sensor/mesh networks, peer-to-peer networks, socialized content sharing, cyber-physical systems, the Internet of Things, cloud computing, big data, and distributed computing. He serves as a TPC member for various international conferences, such as the IEEE/ACM IWQoS, ACM Multimedia, the IEEE ICC, the IEEE GLOBECOM, the IEEE CloudCom, and the IEEE ICME. He was a recipient of the Chinese Government Scholarship for Outstanding Self-financed Students Studying Aboard, in 2009, and the IEEE ICME Quality Reviewer Award, in 2011. He serves as the Program Vice-Chair in International Conference on Internet of Vehicles (IOV), in 2014, and also serves as the TPC Co-Chair in the IEEE International Conference on Communications in China (ICCC), in 2014, for Symposium on Wireless Networking and Multimedia.