

ENERGY-BALANCING, LOCAL DATA CORRELATION-AWARE CLUSTERING ALGORITHM FOR WIRELESS SENSOR NETWORKS

Zefeng Lv,* Fan Wang,* Xiaopeng Hu,* and Yan Yang**

Abstract

Geographically proximate sensor nodes usually temporally and spatially correlated in wireless sensor networks (WSNs). Clustering is considered to eliminate data redundancy and improve in-network data aggregation efficiency. In this paper, an energy-balancing, local data correlation-aware (LDCA) clustering algorithm is proposed for WSNs. Comprehensively, considering the data correlation, energy consumption, communication distance, and other factors, we designed an average entropy and a data correlation coefficient (DCC) to make clustering and aggregation performance more effective. It not only measures data correlation properly but also reduces data volume. We also use the sensor's residual energy as one of the key elements in the cluster-head-selection phase to achieve energy balance. Simulation results indicate that the LDCA clustering algorithm achieves a higher aggregation ratio and performs better with respect to energy consumption and load balance compared to other algorithms.

Key Words

Clustering, data correlation, energy balance, wireless sensor network

1. Introduction

WSNs include a large number of small, cheap, wireless sensor nodes with low power in a monitored environment [1]–[3]. The sensor nodes collect lots of similar data, which results in redundancy [4], [5]. In general, sensors have limited resources, especially energy, but data transmission consumes about 80% of the power [6]. Thus, transmitting redundant data wastes energy. To solve this problem, a clustering routing algorithm is used to eliminate this

kind of data redundancy and improve the in-network data aggregation efficiency. It constructs a reasonable network topology that is fit for distributed algorithms [7].

The Low Energy Adaptive Clustering Hierarchy (LEACH) algorithm is considered as the most popular and classic clustering routing protocol [8]. It operates periodically by “rounds”; each round is separated into two phases: a cluster formation phase and a data transmission phase. In the cluster formation phase, a node is selected as the head of each cluster with a probability. In the data transmission phase, each cluster member node directly communicates with a cluster head (CH), instead of transmitting data to the sink node. Many other algorithms based on LEACH have been proposed, such as HEED [9], EECS [10], and PEGSIS [11]. However, the major objective of such algorithms is to design routing schemes that can minimize communication costs. Therefore, CHs are selected stochastically, and geographically proximate nodes join one cluster regardless of the data correlations between them [12].

In recent years, some data correlation-based clustering algorithms have been proposed [13]–[16]. In these algorithms, some of the data correlation models were designed to determine the nodes that could be selected as CH nodes. Based on the assumption that the set of all observations in the network satisfy a Gaussian distribution, Dabirmoghaddam *et al.* used the Euclidean distance between sampled data to describe the correlation and propose a centralized greedy clustering algorithm [13]. However, the centralized algorithm increases energy costs and network latency due to the global information. Bijan *et al.* proposed a cluster head selection (CHS) methodology based on an energy–cost ratio, which compares the total energy for all CHs to the total network costs [14]. However, to select the CH, the total amount of data in the previous round of data communication and the distance between the CH and each sensor node should be considered. Yang *et al.* propose the low-energy minimum average distance algorithm for cluster distribution; in this algorithm, the CHs are selected to prolong network lifetime [15]. If the sensing area is stable, some nodes may be selected as CHs multiple

* School of Computer Science and Technology, Dalian University of Technology, Dalian, China; e-mail: kkzf_2001@163.com, {wangfan, xphu}@dlut.edu.cn

** School of Computer Science and Technology, Liaoning Normal University, Dalian, China; e-mail: xuemeng62038216@163.com
Corresponding author: Fan Wang

Recommended by Dr. Xiaonv Hu

(DOI: 10.2316/Journal.206.2018.5.206-0063)

times. This can result in energy imbalance due to high-energy consumption by the representative nodes. A data density correlation degree (DDCD) clustering method is provided in [16]. The objective of DDCD is to increase the accuracy of the representative data. So, a correlation model with many custom parameters is proposed to measure the spatial correlation. The parameters include the number of correlated neighbouring nodes, the distance between the node and the data centre, and the average distance between the node and its neighbouring nodes. The selected CH node works as a representative of its cluster for sending gathered data to the sink node. The main problem with these methods is the lack of an adaptive way to confirm the parameters of the correlation model. In addition, the energy balance problem still exists.

However, the above clustering algorithms can still be improved by adopting proper data correlation models. Researchers usually assume that the degree of correlation increases with a reduction in the distance between sensors. Some special nodes are spatially proximate but may not be correlated, such as those near the boundary of two separated fields. The existing algorithms may improperly describe the data correlation.

In this paper, we present an energy-balancing LDCA clustering algorithm to overcome these shortcomings. Based on the Slepian–Wolf theorem [17], we design an average entropy and a DCC to properly measure the correlation between nearby sensors. During the CHS, we comprehensively consider the distance, energy consumption, temporal and spatial correlations, and other factors to optimize clustering and data aggregation performance. Finally, we consider residual node energy as one of the key elements in the CHS phase, so the network energy distribution is balanced.

The remainder of this paper is organized as follows: Section 2 presents the problem statement and network models. Section 3 defines the average entropy, the DCC, and the details of the LDCA algorithm. Section 4 shows the simulation results of evaluating the performance of our clustering algorithm by comparing it to the LEACH and DDCD algorithms. In Section 5, we provide our conclusion and suggest our planned future work.

2. Problem Statement and Network Model

2.1 Problem Statement

In WSNs, a large number of sensor nodes are densely deployed in a monitored environment. The sensors collect data consecutively and work together to observe the sensing field [5]. Therefore, the sample data have the following features: (1) data gathered by different nodes deployed in the same sensing area have strong correlations, which are called *spatial correlations*. (2) In a stable environment, data gathered by one node during a certain time have strong correlations, which are called *temporal correlations*. (3) The scene information can be described through the comprehensive expression of all original data.

Facing the data correlation problem, researchers usually consider that spatially proximate nodes are highly

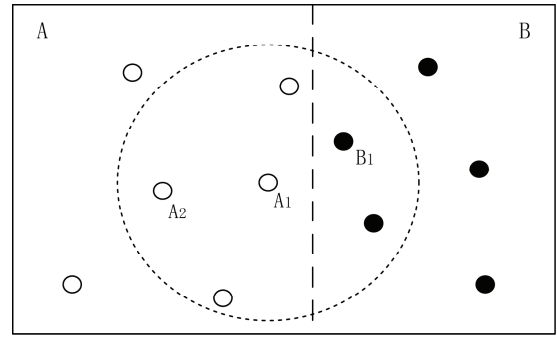


Figure 1. Data correlation between the nodes near the boundary.

correlated. However, some nodes that have adjacent locations maybe uncorrelated in practice, such as the nodes shown in Fig. 1. Assuming that fields A and B are sensing different data the dashed line in Fig. 1 indicates the boundary. The white nodes collect data from A and the black nodes collect data from B. Let us take node A1 as an example. The node lies near the boundary and its communication range is shown as the circle. Although the distance between B1 and A1 is shorter than that between A2 and A1, they are apparently not correlated. Therefore, using spatial distance as the only measurement cannot solve the data correlation problem.

Another problem is the network energy balance. In some current clustering algorithms, CH nodes are responsible for transmitting data while the cluster member nodes are sleeping most of the time. Some nodes may work as CH nodes for a long time, which leads to network energy imbalance. Then, the representative nodes consume energy so fast that the network dies very quickly.

Thus, we comprehensively consider the distance, energy consumption, local data correlation, and other factors to measure the local data correlation, so the clustering performance can be optimized and the network energy consumption minimized. We also consider the node residual energy as one of the key elements in the clustering phase to achieve energy balance. The details are presented in Section 3.

2.2 Network Model

The network model proposed in this paper is based on the following assumptions.

- (1) All N sensor nodes are distributed in the $M \times M$ rectangular area.
- (2) There is only one sink node with a fixed geographical location and unlimited resources.
- (3) Each sensor node has a unique ID and communication range r , which can be adjusted by altering the transmitting power.
- (4) The distances between sensor nodes are calculated by the received signal power level.

Computation and communication costs are the two main resources consumed by a sensor node. In most WSN applications, sample data such as temperature, humidity, and pressure are linear. In this case, we mainly consider the

communication cost and ignore the computation process. The transmitting energy cost is described by the following equation:

$$E_t = \sum_{i=1}^N e_i \times b_i \quad (1)$$

where e_i is the energy coefficient, which is the energy required to transmit one bit of data. According to the first-order radio model [18], e_i is directly determined by communication distance. b_i is the data coefficient, which is the total data volume. To minimize the energy consumption in our algorithm, we must make sure that (1) results in a minimum value. In our work, we use average entropy to achieve low communication consumption and minimize the number of transmitted bits. Average entropy can measure the data correlation at the same time.

3. Algorithm Description

In this section, we first introduce the Slepian–Wolf theorem. Then, we specifically define an average entropy and a DCC. Finally, details of the LDCA algorithm are presented.

3.1 Problem Statement

A theoretical method of coding two correlated sources in lossless compression was discovered by David Slepian and Jack K. Wolf in 1973 [17]. The Slepian–Wolf theorem proves that two correlated sources can achieve a certain compression efficiency without communicating with each other in lossless compression. The Slepian–Wolf theorem can determine the minimum number of bits per source character required for the two encoded message streams.

According to the Slepian–Wolf theorem, two correlated sources can achieve a coding rate equal to the joint entropy as long as their individual rates are at least equal to the condition entropy. The relationship can be described as follows:

$$\begin{aligned} R_1 &\geq H(Y|X) \\ R_2 &\geq H(X|Y) \\ R_1 + R_2 &\geq H(X, Y) \end{aligned} \quad (2)$$

Sensors X and Y are correlated sources. When Y sends data to X , X can reconstruct it based on local information. So, the output of Y can be compressed at a rate of $R_1 \geq H(Y|X)$. Then, the total rate is minimized as follows:

$$R = R_1 + R_2 \geq H(Y|X) + H(X) = H(X, Y) \quad (3)$$

Thomas Cover generalized this to a situation with multiple correlated sources to prove that the Slepian–Wolf theorem can also be applied for the N -dimensional case [19].

Given a network $S = \{ni | i = 1, 2, 3, \dots, N\}$ with N sensor nodes, where n_i represents one sensor node, we use a node as the centre and its communication range r as the

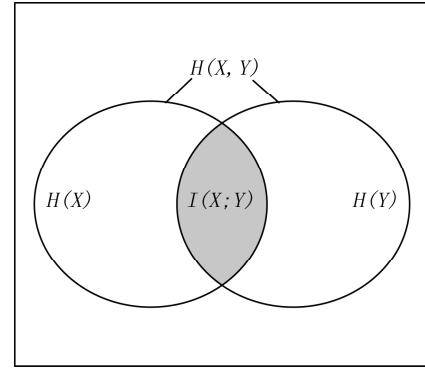


Figure 2. Relationship between entropy and mutual information.

radius to construct a circle area. Let NS_i be the neighbour set containing all nodes within the area. According to our analysis of the Slepian–Wolf theorem, the minimal local data rate transmitted from Y to X can be expressed as

$$H(Y|X) = H(X, Y) - H(X). \quad (4)$$

In information theory [20], the joint entropy can also be defined as

$$H(X, Y) = H(Y) + H(X) - I(X; Y), \quad (5)$$

where $I(X; Y)$ is the mutual information between centre node X and neighbour node Y . Mutual information is a measurement of the information correlation, which is used to represent the relationship between two random variables. Figure 2 shows the relationship among all variables in (5), and the mutual information is represented in the shaded parts. The mutual information probability between X and Y is given as

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (6)$$

where $p(x)$ and $p(y)$ are the probability distributions of source X and source Y , and $p(x, y)$ is the joint probability distribution.

According to Fig. 2, we can see that the entropy and mutual information are highly related. Thus, we can further discuss the correlation of the data ECC as follows:

$$ECC = \frac{2I(X; Y)}{H(X) + H(Y)}. \quad (7)$$

Thus, the joint entropy can be described as

$$H(X, Y) = \left(1 - \frac{1}{2}ECC\right) (H(X) + H(Y)). \quad (8)$$

By knowing the mutual information between X and Y , we can obtain the minimal number of total transmitted bits. However, the mutual information calculation requires a lot of computing and exchange of a large volume of data. This process wastes much energy and creates extra network costs. In our work, ECC is the normalized form of the mutual information and ranges from 0 to 1.

3.2 DCC

According to (1), the total energy consumption can be minimized when the network achieves a minimum data volume. We combine the Slepian–Wolf theorem with the clustering strategy to reduce the local data rate among the highly correlated nodes. Thus, local data correlation can also be measured by the condition entropy. We use correlation coefficient ρ to estimate the joint entropy, and design an average entropy and a DCC to obtain optimal clustering results.

Definition 1. *Average Entropy.* For any node $n_i \in S, n_j \in NS_i$, the average entropy Eh_i is defined as

$$Eh_i = \sum_j \frac{H(Y_j|X_i)}{|NS_i|} = \sum_j \frac{H(Y_j) - \frac{1}{2}\rho(H(X_i) + H(Y_j))}{|NS_i|}. \quad (9)$$

The average entropy describes the average coding rate from each node in NS_i to its centre node n_i . The average entropy measures the data correlation between nodes and ensures the minimum amount of local data transfer between nodes. As indicated in (9), Eh_i is zero if X_i is correlated with Y_i (in this case, Y_i is completely determined by X_i) and it reaches its maximum when X_i and Y_i are independent. As a result, the lower the average entropy, the higher the increase in the correlation between node n_i and its neighbours. If node n_i serves as a CH, the local energy consumption within the cluster can be minimized. Thus, the average entropy can determine the node's ability to be a CH.

Definition 2. *DCC.* To determine the local data correlation, the DCC between node n_j and n_i is defined as

$$\delta_{ij} = |H(Y_j|X_i) - Eh_i|. \quad (10)$$

Therefore, all the nodes in NS_i can be divided into two types – related nodes and unrelated nodes – according to the decision variable t_{ij} , which mainly measures the relationship between the mutual entropy and the average entropy for nodes and their neighbours:

$$t_{ij} = \begin{cases} 1, & \delta_{ij} \leq \varepsilon \\ 0, & \delta_{ij} \geq \varepsilon \end{cases} \quad (11)$$

where ε is a predefined decision threshold. If $t_{ij} = 1$, the nodes are related to the centre node and classified as the related node type. If $t_{ij} = 0$, the nodes are unrelated to the centre node and belong to the unrelated node type. For each node n_i , we construct a related nodes set $RS_i = \{n_j | n_j \in NS_i, t_{ij} = 1\}$ and an unrelated nodes set $US_i = \{n_j | n_j \in NS_i, t_{ij} = 0\}$.

Definition 3. *Connection Degree:*

$$D_i = |RS_i| \quad (12)$$

The connection degree describes the density of correlated nodes. As the value of D_i increases, node n_i is more likely to be in the centre of the densely deployed correlated area. Thus, we consider it as one of the criteria in the CHS phase. The data correlation can be measured precisely. The connection degree can also indirectly reflect the characteristics of the geographical area.

We use the above fundamental definitions to propose the distributed, energy-balancing, LDCA clustering algorithm. The algorithm can effectively measure the local data correlation and minimize total energy consumption. During the clustering process, we also consider the nodes' residual energy to achieve energy balance, so that network lifetime is extended.

3.3 Energy-balancing LDCA Clustering Algorithm

The LDCA clustering algorithm consists of three steps: the initialization, the CHS, and the cluster formation (CF). In the initialization phase, each node calculates its own average entropy, DCC, and other parameters in parallel. After this procedure, the related nodes in NS_i are added to RS_i and the unrelated nodes are added to US_i for each node n_i . Then, all nodes in the network initialize their status to “ON” in preparation for the following phases.

In the CHS phase, each node exchanges an initialization message with all its neighbours to select the local optimal nodes as cluster heads. This process includes two steps. In Step 1, each node n_i sends the ID, average entropy, and connection degree to its neighbours in NS_i . In Step 2, the goal is to select a suitable cluster head according to the local data correlation and selection rules, which include minimal average entropy, maximal connection degree, and residual energy. All the selected cluster heads are then added to a global cluster head set and labelled as a cluster head.

In the CF phase, cluster head nodes broadcast notification messages in the sensor network. Other nodes must join a cluster according to some predefined rules. Therefore, strongly related sensor nodes are then divided into the same cluster. First, all the cluster head nodes in C broadcast notification messages. Then, other nodes will receive n messages. When $n = 1$, the node is directly added to the corresponding cluster and it becomes a cluster member. If $n \geq 2$, there is at least one message from its RS nodes. The node will join a cluster with the shortest communication distance between the CH nodes. The status is labelled as a cluster member. If all n messages are from US nodes, the node itself will become a cluster head and set its status to cluster head. Finally, all nodes with a cluster member status send a message to join their cluster heads, respectively. At this point, the clustering network topology has been constructed.

4. Performance Analysis

In this section, we present the simulation experiments used to evaluate the performance of the LDCA clustering algorithm. To prove the superiority of the LDCA algorithm, we

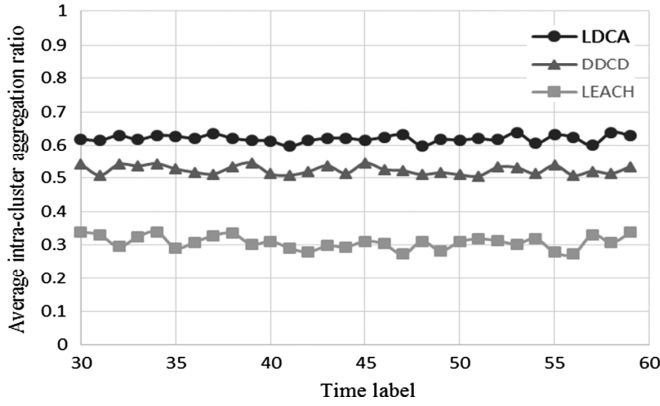


Figure 3. Comparison of average intra-cluster aggregation ratios among the LDCA, LEACH, and DDCD algorithms.

compare it with the current typical algorithms including LEACH and DDCD.

In our experiments, the dataset collected in the Intel Berkeley Research Lab in 2004 [21] is considered. There are 54 nodes deployed in the network. We chose the communication radius $r = 25m$ and set the sampling time interval to 2 min. Then, we analysed the clustering performance, total data volume, energy consumption, and energy balance performance.

4.1 Clustering Performance Analysis

One of our major objectives was to measure the correlations between sensor nodes. The nodes that belong to the same cluster have high correlation, so the data aggregation is more effective in the network. We use the average intra-cluster aggregation ratio R_{avg} to measure the clustering performance

$$R_{avg} = 1 - \frac{\sum K B_s / B_r}{K} \quad (13)$$

where B_r represents the total data volume received from all cluster members, B_s represents the total data volume forwarded by a CH node to the sink node, and K is the number of clusters in the network. Obviously, if the value of R_{avg} is large, the clustering performance is more effective. In the experiment, we recorded the amount of data transmission in 30 consecutive time periods during the stable data transmission phase. The results are shown in Fig. 3. According to the figure, the LEACH algorithm has the lowest aggregation ratio because it does not consider the data correlation between nodes. The DDCD and LDCA algorithms both perform better, but the LDCA achieves the best aggregation ratio. The results illustrate that our algorithm can effectively measure the local data correlation and have better clustering performance.

We changed the communication radius from 15 m to 40 m, simulated the LDCA algorithm repeatedly, and found that the communication radius can influence the average intra-cluster aggregation ratio. The relationship between R_{avg} and the communication radius is shown in Fig. 4.

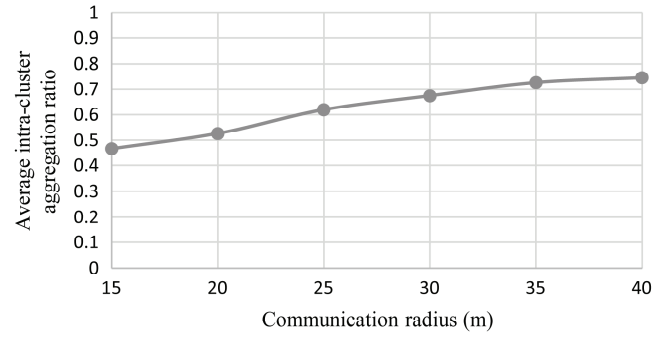


Figure 4. Relationship between the average intra-cluster aggregation ratio and the communication radius.

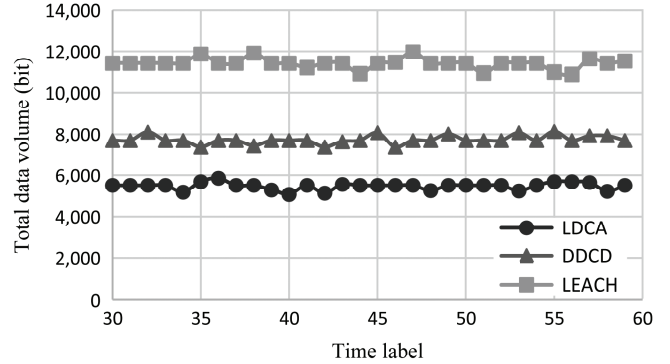


Figure 5. Comparison of network total data volume among the LDCA, LEACH, and DDCD algorithms.

As we know, after the clustering phase, an isolated node will become a cluster head. In this situation, the intra-cluster aggregation ratio equals 0. With a reduction in the communication range, the probability of being an isolated node increases, so the number of isolated nodes can affect the clustering performance. Thus, we can conclude that the communication radius plays an important role in the improvement of clustering performance.

4.2 Total Data Volume Analysis

According to the energy analysis in Section 3, most of the energy is consumed in data transmission, so we track the total data volume from the CH nodes to the sink node within a certain time period. The comparison of data transmission among the LDCA, LEACH, and DDCD algorithms is shown in Fig. 5. In the stable data transmission phase, the LDCA algorithm has a minimum number of bits, which means the total data rate of each cluster has been minimized. Therefore, our algorithm effectively reduces the amount of data transmission and performs better than other algorithms.

4.3 Energy Consumption Analysis

We recorded the residual network energy in 60 time periods and drew a residual energy-time curve for each algorithm, as shown in Fig. 6. As the figure shows, in the first several time periods, the total energy consumption of the three algorithms is low because they are all in the

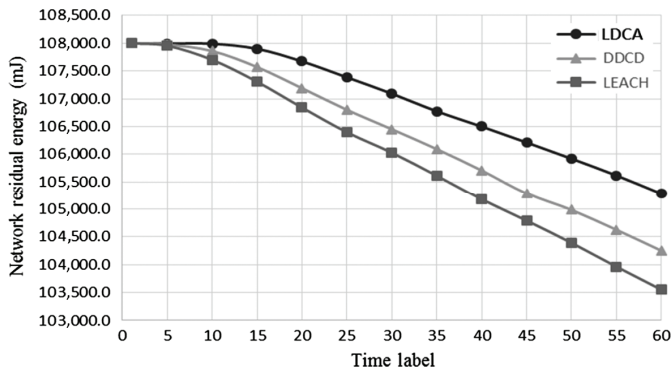


Figure 6. Comparison of network energy consumption among the LDCA, LEACH, and DDCD algorithms.

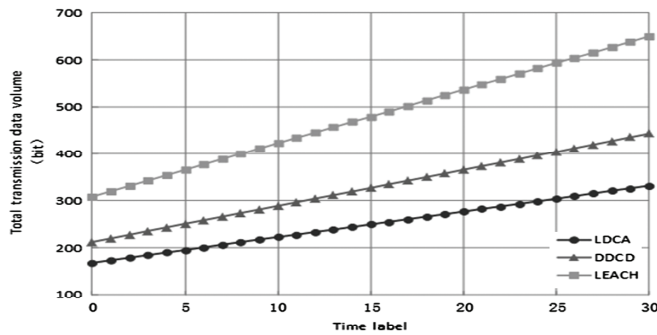


Figure 7. Comparison of total data transmission volume.

clustering phase. When data transmission begins, both the LEACH and DDCD algorithms consume more energy than LDCA. Because we comprehensively consider the distance, energy consumption, temporal and spatial correlation, and other factors, the LDCA conserves more energy than the other algorithms. According to the simulation results and analysis, our algorithm more effectively reduces energy consumption and extends network life.

4.4 Energy Balance Analysis

Energy consumption is mainly a result of data transmission, so we tracked the total data volume (bits) that was transmitted to the CH and compared the results of the different clustering algorithms, as shown in Fig. 7. When comparing the different curves, we found that the total data transmission rate of the LDCA algorithm had the smallest number of bits in the stable transmission stage, thus proving that the local data transmission volume in the cluster is reduced, and performance is better than the LEACH and DDCD clustering algorithms.

5. Conclusion

In this paper, we proposed an energy-balancing, LDCA clustering algorithm for WSNs. An average entropy and a DCC were designed in our algorithm. We comprehensively considered the communication distance, energy consumption, and temporal and spatial correlations to optimize

clustering and aggregation performance and reduce the total data volume. In our simulation experiments, we compared the LDCA clustering algorithm with the DDCD and LEACH algorithms. The results indicate that the LDCA algorithm has a higher inter-cluster average aggregation ratio and performs better with respect to energy consumption and energy balance compared to other algorithms. In future work on the LDCA clustering algorithm, we will focus on reducing the message overhead and designing a data aggregation method for event detection.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (No. 61272523, No. 61471084), the National Key Project of Science and Technology of China (No. 2011ZX05039-003-4), and the Fundamental Research Funds for the Central Universities.

References

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, Wireless sensor networks: A survey, *Computer Networks*, 300–301(4), 2013, 490–493.
- [2] J. Yick, B. Mukherjee, and D. Ghosal, Wireless sensor network survey, *Computer Networks*, 52(12), 2008, 2292–2330.
- [3] W. Cui, X. Meng, B. Yang, *et al.* An efficient lossy link localization approach for wireless sensor networks, *Frontiers of Information Technology & Electronic Engineering*, 18(5), 2017, 689–707. <http://dx.doi.org/10.1631/FITEE.1601247>
- [4] S. Pattem, B. Krishnamachari, and R. Govindan, The impact of spatial correlation on routing with compression in wireless sensor networks, *ACM Transactions on Sensor Networks (TOSN)*, 4(4), 2008, 1–33.
- [5] M.C. Vuran, Ö.B. Akan, and I.F. Akyildiz, Spatio-temporal correlation: Theory and applications for wireless sensor networks, *Computer Networks*, 45(3), 2004, 245–259.
- [6] H. Çam, S. Özdemir, P. Nair, D. Muthuavinashiappan, and H.O. Sanli, Energy-efficient secure pattern based data aggregation for wireless sensor networks, *Computer Communications*, 29(4), 2006, 446–455.
- [7] Y. Gao, K. Wu, and F. Li, Analysis on the redundancy of wireless sensor networks, *Proc. 2nd ACM International Conf. on Wireless Sensor Networks and Applications*, San Diego, CA, USA, 2003, 108–114.
- [8] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, *Proc. the 33rd Annual Hawaii International Conference on System Sciences*, 2, 2000, 10–19.
- [9] O. Younis and S. Fahmy, HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, *IEEE Transactions on Mobile Computing*, 3(4), 2004, 366–379.
- [10] G. Chen, L.I. Chengfa, and Y.E. Mao, EECS: An energy-efficient clustering scheme in wireless sensor networks, *Ad Hoc & Sensor Wireless Networks*, 3(2–3), 2007, 99–119.
- [11] A.G. Karegowda, B.G. Premasudha, and G. Devika, A pragmatic study of evolutionary techniques based energy efficient hierarchical routing protocols – LEACH and PEGASIS, *International Journal of Applied Engineering Research*, 10(17), 2015, 38274–38285.
- [12] H.W. Ferng, R. Tendean, and A. Kurniawan, Energy-efficient routing protocol for wireless sensor networks with static clustering and dynamic structure, *Wireless Personal Communications*, 65(2), 2012, 347–367.
- [13] A. Dabirmoghaddam, M. Ghaderi, and C. Williamson, Cluster-based correlated data gathering in wireless sensor networks, *International Symposium on IEEE Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MAS-COTS)*, Miami Beach, FL, USA, 2010, 163–171.

- [14] B.K. Debroy, M.S. Sadi, and M.A. Imran, An efficient approach to select cluster head in wireless sensor networks, *Journal of Communications*, 6(7), 2011, 529–539.
- [15] L.L. Yang and L. Tao, Research on the low-energy minimum average distance algorithm for the distribution of cluster-head nodes in wireless sensor networks, *Journal of Convergence Information Technology*, 7(16), 2012, 208–213.
- [16] F. Yuan, Y. Zhan, and Y. Wang, Data density correlation degree clustering method for data aggregation in WSN, *IEEE Sensors Journal*, 14(4), 2014, 1089–1098.
- [17] D. Slepian and J.K. Wolf, Noiseless coding of correlated information sources, *IEEE Transactions on Information Theory*, 19(4), 1973, 471–480.
- [18] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications*, 1(4), 2002, 660–670.
- [19] T.M. Cover, A proof of the data compression theorem of Slepian and Wolf for ergodic sources, *IEEE Transactions on Information Theory*, 21(2), 1975, 226–228.
- [20] T.M. Cover and J.A. Thomas, *Elements of information theory*, (Hoboken, NJ, USA: John Wiley & Sons, 2006), 21–22.
- [21] Institute for Nuclear Theory, Seattle, WA, USA, 2004, Intel Lab Data [Online]. Available: <http://db.csail.mit.edu/labdata/labdata.html>.



Xiaopeng Hu is currently a professor in computer science at the School of Computer Science and Technology, Dalian University of Technology. He received a PhD. degree in computer science from Imperial College London, UK, in 2005. His research interests include computer vision, machine learning, and sensor fusion.



Yan Yang is currently a lecturer at the School of Computer Science and Information Technology, Liaoning Normal University. Her research interests include sensor network, machine learning, and pattern recognition.

Biographies



Zefeng Lv is a Ph.D. candidate in computer science at Dalian University of Technology and a senior engineer at China Liaohe Petroleum Engineering Co., Ltd. His research interests include sensor networks, machine learning, and instrument control systems.



Fan Wang is currently an associate professor at the School of Computer Science and Technology, Dalian University of Technology. Her research interests are wireless sensor networks, computer vision, and machine learning.