

PAPER • OPEN ACCESS

HDCDS-CDG: A Hierarchically Diffused Connected Dominating Sets based Compressed Data Gathering Scheme

To cite this article: Xiaohan Yu *et al* 2020 *J. Phys.: Conf. Ser.* **1570** 012056

View the [article online](#) for updates and enhancements.

You may also like

- [The Imprint of Cosmic Reionization on the Luminosity Function of Galaxies](#)
Sownak Bose, Alis J. Deason and Carlos S. Frenk
- [Inlaying Silicon in SiC-Derived Graphite with Unique Cavity Structure as a High-Capacity Anode for Li-Ion Batteries](#)
Mengfei Hu, Yujie Ma, Houzheng Wu *et al.*
- [Very Large Array Radio Study of a Sample of Nearby X-Ray and Optically Bright Early-type Galaxies](#)
Romana Grossová, Norbert Werner, Francesco Massaro *et al.*



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

HDCDS-CDG: A Hierarchically Diffused Connected Dominating Sets based Compressed Data Gathering Scheme

Xiaohan Yu*, Wenlei Xu and Chao Chen

School of Information and Electronic Engineering, Zhejiang Gongshang University, HangZhou, ZheJiang, 310018, China

*Corresponding author's e-mail: yuxiaohan188@126.com

Abstract. How to reduce the energy consumption and prolong the network lifetime is one of the most important issues in Wireless Sensor Networks (WSNs). This paper proposed a Compressed Data Gathering (CDG) scheme based on Hierarchically Diffused Connected Dominating Sets (HDCDS). In the proposed scheme, a connected dominating set was constructed as the backbone of the network. All of the other nodes, i.e., non-backbone nodes, are connected to the backbone. Nodes in the backbone collect data from non-backbone nodes using the traditional data collection scheme. Data within the backbone are transmitted to the base station in CDG manner. Theoretical analysis and simulations showed that, when compared to the LEACH methods and traditional compressive sensing-based data collection methods, HDCDS-CDG not only reduces the consumption of network resources but also increases the lifetime of the network.

1. Introduction

A Wireless Sensor Network (WSN) is composed of a large number of stationary or moving sensors in a self-organizing and multi-hop manner [1]. In WSNs, how to reduce the energy consumption and balance network loads is a very important problem. The Low Energy Adaptive Clustering Hierarchy (LEACH) scheme is widely used in WSNs [2-3]. Compared to common routing algorithms, LEACH has many advantages: 1. Cluster heads are randomly selected, which distributes the network energy consumption and hence balances network loads; 2. A clustering mechanism is adopted to manage cluster heads. As nodes in a higher level of the network, cluster heads are responsible for managing cluster members, which results in free selection of transmission paths and storage of routings; 3. The clustering mechanism has a good expansion effect. The ability of self-organizing and the method of sequential election of cluster heads extends the lifetime of the network. However, the LEACH scheme still has disadvantages: cluster heads are not uniformly distributed in the network, which fails to consider the location and residual energy of nodes and excessive energy consumption of cluster heads in the stable phase. Considering these disadvantages, reference [4] proposed the LEACH-EDH method, which fully considers the energy and geographical location of nodes during the clustering process, and adopts the probabilistic hybrid routing algorithm to reduce energy consumption in the transmission phase. Reference [5] proposed the LEACH-improved method, which improves the calculation of threshold by adding spacing factor, residual energy factor and node density factor, and thus reduces the energy consumption of nodes.

This paper considered the energy consumption and network lifetime of data acquisition and transmission in WSNs. Traditional methods generally compress data in the network by exploiting the correlation of data, and then send it to the base station [6-7]. Recently, the Compressed sensing (CS)



technology, which is a new direction in mathematics and information science, is very popular[8]. The idea is to preserve the useful information and discard the useless part in data. Reference [9] applied the CS technology in the data collection process in WSNs, and proposed the Compressive Data Gathering (CDG) method. Different from the traditional data collection methods, CDG collects linear combinations of the original data instead of the original data itself. The base station is able to recover the original data as long as enough number of linear combinations are collected. CDG not only reduces the energy consumption but also balances the network loads.

In this paper, we proposed a Hierarchically Diffused Connected Dominating Set-based Compressed Data Gathering (HDCDS-CDG) method by combining CDG and the Connected Dominating Set (CDS). It has the following advantages:

1. In HDCDS-CDG, the hierarchical diffusion is used to find cluster heads, i.e., nodes in the connected dominating set. The distribution of each cluster head is reasonable. It considers the residual energy of each node in the process of selecting cluster head. This avoids the problem that a node is selected as cluster heads for several times due to the randomness, and thus extends the network lifetime. Since HDCDS-CDG uses the idea of CDG to transmit linear combinations of the original data, the number of transmission for link within the CDS is fixed to m . The data can be finally transmitted to the aggregation node with relatively small transmissions.

2. Theoretical analysis and numerical experiments showed that, when compared to the LEACH algorithm, the selection of cluster heads and their distribution are more reasonable. In the process of data transmission, the energy consumption of cluster heads is reduced and the efficiency of data transmission is improved. Generally, the HDCDS-CDG algorithm can achieve less energy consumption, more efficient data transmission and longer lifetime of the network.

2. Hierarchically Diffused Connected Dominating Set-based Compressed Data Gathering

The HDCDS-CDG scheme first constructs a backbone which is consisted of nodes with large transmission gains. Nodes in the backbone can transmit data to the sink within a relatively small hops. We divide the data collection into two phases: Phase 1: non-backbone nodes transmit original data to the backbone; Phase 2: nodes in the backbone transmit linearly combined data to the sink using the idea of pipelining.

2.1 The network model

Suppose the network can be represented as an undirected graph $G=(V, E)$, where V is the set of nodes with $|V|=n$, and E is the set of edge. Suppose nodes are randomly distributed in a unit square, and the cluster heads and cluster members are fixed after the deployment. Moreover, we assume that each node has the same initial energy E_0 and communication range R_c . Each node is able to fuse data and sense its own residual energy. A unique identifier, i.e., the IDs $(1, 2, \dots, n)$, is used to identify nodes in the network. Before introducing the proposed scheme, we first introduce some definitions:

- **Definition 1:** A subset $U \subseteq V$ is a dominating set of G if each node in $V-U$ is adjacent to some node in U . U is a CDS if the subgraph induced by U is connected.
- **Definition 2:** A subset $U \subseteq V$ is an Independent Set (IS) of G if no two nodes in U are adjacent.
- **Definition 3:** An IS U is a Maximal Independent Set (MIS) of G if $U \cup \{v\}$ is not an IS for any $v \in V-U$.

An important step is to construct the backbone of the network. In this paper, the backbone is a connected dominating set. A dominating set is called as the Connected Dominating Set (CDS) if each pair of the nodes are connected. The less number of nodes in the CDS, the higher efficiency of the data transmissions. A key point is to find a CDS with small size.

2.2 Selection of cluster heads

The construction of the CDS is based on a MIS. Initially, all nodes are white, and finally these nodes will be converted to black or gray. Black nodes represent cluster heads. During the process of

constructing the CDS, nodes are colored green, blue and gray when they are active, chosen as the connected nodes and chosen as the cluster members respectively. Figure 1 is a schematic diagram of color transitions. The selection of cluster heads is divided into three stages: Stage 1: constructing a information table of neighbors; Stage 2: constructing an MIS; Stage 3: selecting the connected nodes and generating the CDS.

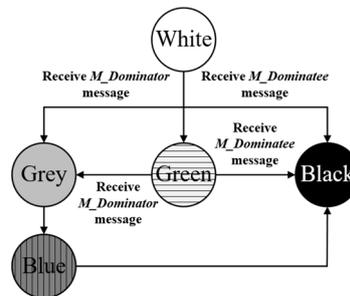


Figure 1. Schematic diagram of node state transition

Stage 1 : Constructing an information table of neighbors. Each node broadcasts an M_Hello message with its ID. Upon receiving M_Hello messages from neighbors, each node can construct an information table. The information table is a $C \times 4$ two-dimensional array where C is the number of neighbors. The first column is the number of the neighbors. The second column is the corresponding color of neighbors (initially, each node is white). The third column is the number of black neighbors (the initial value of this column is null). The fourth column is the corresponding residual energy of neighbors.

Stage 2 : Constructing an MIS. Suppose each node knows its level in the network. The sink starts the process of constructing an MIS by broadcasting an $M_Dominator$ message.

1. After receiving the $M_Dominator$ message, white nodes set the sink as their parent node and mark the sink black in the information table. Meanwhile, white nodes mark themselves as gray and transmit the $M_Dominatee$ to neighbors.

2. After receiving the the $M_Dominatee$ message, white nodes mark themselves as green. Meanwhile, they enter the election state and transmit the M_Active message to neighbors. This message carries information about the number of white neighbors and the residual energy. A white node marks itself as gray if the both the $M_Dominator$ message and the $M_Dominatee$ message are received. Then it will broadcasts the $M_Dominatee$ message. If a green node receives the M_Active message, all of its green neighbors will join in the election. A green neighbor with the maximum residual energy wins and marks itself as black. Then the winner broadcasts the $M_Dominator$ message. After receiving the $M_Dominator$ message, a green node marks itself as gray and then broadcasts the $M_Dominatee$ message.

3. If a gray node receives the $M_Dominator$ (respectively $M_Dominatee$) message, the color of the transmitting node in the information table will be marked as black (respectively gray). This process is repeated until no white nodes appear in the network.

Stage 3 : Constructing the CDS.

1. Each gray node marks itself as blue if there exist at least 2 black neighbors which are in different levels. If all of the black neighbors connect to at least one blue node, the gray node remains unchanged.

2. After all of the gray nodes are checked, each blue node marks itself as black. So far, all of the black nodes consists a CDS.

2.3 The clustering

After the selection of cluster heads, each cluster member needs to choose its corresponding cluster head. Each node in the CDS sends an M_Head message with its own residual energy to neighbors indicating that it is a cluster head. After receiving M_Head messages from several cluster heads, the cluster member searches its information table and transmits an M_Join message to the cluster head

with largest residual energy. Finally, each cluster head deletes nodes from which M_Join message is node received, and updates corresponding information.

2.4 Data transmissions in the stable state

Data transmissions in the stable state is divided into two phases: In the phase 1, each cluster members transmits the original data to its corresponding cluster head. In the phase 2, data are transmitted to the sink within the backbone by adopting the CDG scheme. In the CDG scheme, each node sends a linear combination of the original data instead of the original data itself. Figure 2(a) shows the CDG scheme, where each node linearly combines its own data with the received data from the previous hop and transmits the result to the next hop. Figure 2(b) shows the idea of the proposed scheme in a linear network. It can be seen that, when compared with the pure CDG scheme, the proposed scheme reduces the number of transmissions.

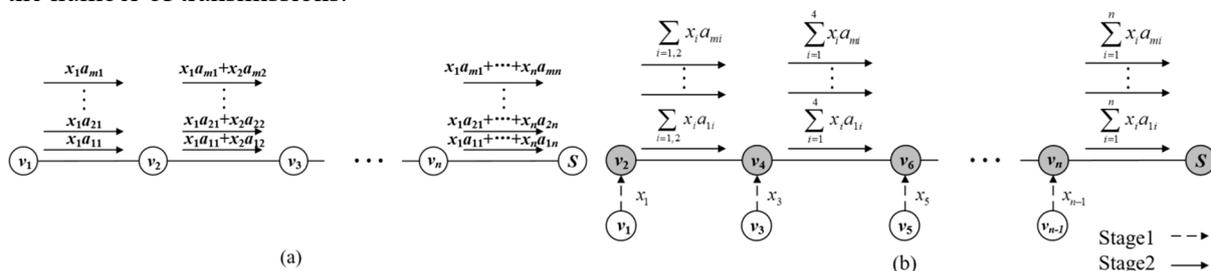


Figure 2. The schematic diagram of (a) CDG transmission (b) Linear network scheme with CDS

3. Theoretical analysis

In this paper, we only consider the energy that is consumed by data transmissions. This is because the energy consumed by computation is trivial. In the phase 1, each cluster member transmits its original data to the corresponding cluster head. Therefore, the number of transmissions in the phase 1 is given by $|D|$ where D is the number of cluster members. In the phase 2, each cluster head transmits the fused data m times. Therefore, the number of transmissions in the phase 2 is given by $|I \cup W|$ where I is the number of cluster heads and W is the number of node in the CDS excluding cluster heads. Next, we use Lemma 1 proposed by Oler [10] to determine $|I \cup W|$.

Lemma 1: Suppose a compact convex area C contains non-overlapping disks with a diameter of 1. Then, the total number of discs does not exceed:

$$\frac{2}{\sqrt{3}} A(C) + \frac{1}{2} P(C) + 1. \quad (1)$$

In equation (1), $A(C)$ is the area of C , and $P(C)$ is the perimeter of C . According to Lemma 1, the maximum number of non-overlapping circles with radius r is given $(2n/\sqrt{3}r^2) + (2\sqrt{n}/r) + 1$. Since I is an MIS, we have that

$$|I| \leq \frac{2n}{\sqrt{3}r^2} + \frac{2\sqrt{n}}{r} + 1. \quad (2)$$

By using results in the reference [11], we can get equation (3):

$$\Delta_i^* \leq \begin{cases} 4 & \text{if } i \text{ is even;} \\ 11 & \text{if } i \text{ is odd and } i > 1; \\ 12 & \text{if } i = 1. \end{cases} \quad (3)$$

where Δ_i^* represents the number of neighbors for any cluster head node.

Lemma 2: Let us consider the CDS. For any node that belongs to I_j ($1 \leq j \leq R-1$), there are at most $11\beta_{\rho+1}$ nodes in W_{j+1} connecting to it, i.e., $|W_{j+1}| \leq 11\beta_{\rho+1}|I_j|$ and $|W_1| \leq 12\beta_{\rho+1}$. For any node that belongs to W_i ($1 \leq i \leq R$), there are at most $4\beta_{\rho+1}$ nodes connecting to I_i , i.e., $|I_i| \leq 4\beta_{\rho+1}|W_i|$.

According to Lemma 2, we have that:

$$\begin{aligned}
 |IUW| &= \sum_{i=1}^R |W_i| + |I| \leq 12\beta_{\rho+1} + 11\beta_{\rho+1} \sum_{i=1}^{R-1} |I_i| + |I| \\
 &= 12\beta_{\rho+1} + 11\beta_{\rho} + (|I| - |I_R| - |I_0|) + |I| \leq (11\beta_{\rho} + 2)|I| + 12\beta_{\rho+1} = O(nr^{-2}).
 \end{aligned} \quad (4)$$

where $\beta_{\rho+1} = (4\sqrt{3} + 3)\pi/3 + 3$. Finally, the total number of transmissions in the HDCDS-CDG scheme is given by

$$T = |D| + m|IUW| = n - |IUW| + m|IUW| = O(n + nmr^{-2} - 1) = O(n + kr^{-2}n \log n - 1). \quad (5)$$

Next, we introduce the energy consumption model that is widely used in WSNs [12]. The energy consumption of transmitting s bits is given by

$$E_t = \begin{cases} (\beta + \varepsilon_{fs} \cdot d^2)s & d \leq r; \\ (\beta + \varepsilon_{mp} \cdot d^4)s & d > r. \end{cases} \quad (6)$$

where ε_{fs} is the coefficient in the free space attenuation model, ε_{mp} is the coefficient of the multipath attenuation channel model, $r = \sqrt{\varepsilon_{fs} / \varepsilon_{mp}}$ is the distance threshold, β is the energy consumption of processing one bit data. Obviously, $E_t = (\beta + \varepsilon_{fs} \cdot d^2)s$ in this paper. The energy consumption of receiving s bits is $E_r = \beta \cdot s$.

4. Numerical experiments

In this paper, *Matlab* software is used for simulation experiments. We assume that n nodes are randomly deployed in the $100\text{m} \times 100\text{m}$ rule area. The initial energy of each node is $E_0 = 5\text{J}$, the free space attenuation coefficient is $\varepsilon_{fs} = 10\text{pJ/bit/m}^2$, the multipath channel attenuation coefficient is $\varepsilon_{mp} = 0.011\text{pJ/bit/m}^4$, the data packet size is set to $DM = 4000\text{bit}$, the distance threshold $r = 30\text{m}$, the matrix sparsity in the compressed data collection is set as $k = 100$, and the probability of a node becoming a cluster head in the LEACH is set to $p = 10\%$. To smooth the randomness, each result is averaged over 50 experiments.

E_{min} and E_{ave} represent the minimum residual energy and the average proportion of the residual energy of each node respectively. Figure 3 shows the the ratio the normalized E_{min} between HDCDS-CDG scheme and the LEACH scheme. It can be observed that the node with the least residual energy in the LEACH scheme consumes more energy than that in the HDCDS-CDG scheme. Moreover, as the network size increases, the node with the least residual energy in the LEACH scheme decreases faster than that in the HDCDS-CDG scheme. This showed that the HDCDS-CDG scheme performed better than the LEACH scheme from the perspective of network lifetime. Figure 4 compares the performance of HDCDS-CDG scheme and LEACH scheme in terms total energy consumption. It can be observe that, as the size of the network increases, the total energy consumption of LEACH scheme is between HDCDS-CDG scheme and the LEACH scheme increases. This implies that the proposed scheme is more suitable for large-scale networks.

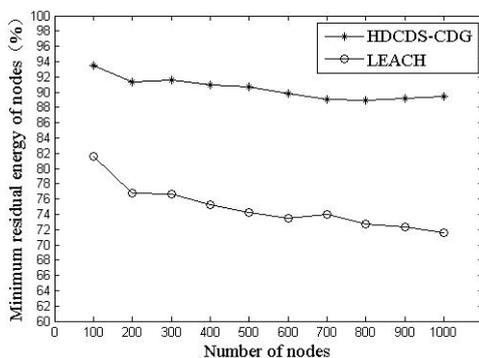


Figure 3. Minimum residual energy of nodes

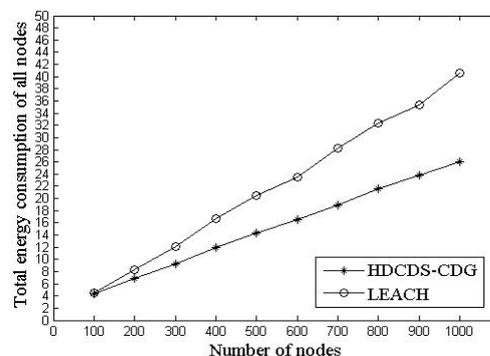


Figure 4. Total energy consumption of nodes

Figure 5 compares the lifetime between the HDCDS-CDG scheme and the LEACH scheme. The lifetime of a network is defined as the time until the first node die. It can be observed that, as the network size increases, the network lifetime decreases. However, the network lifetime in the HDCDS-CDG scheme decreases slower than that in the LEACH scheme. This implies that the HDCDS-CDG scheme is able to prolong the network lifetime.

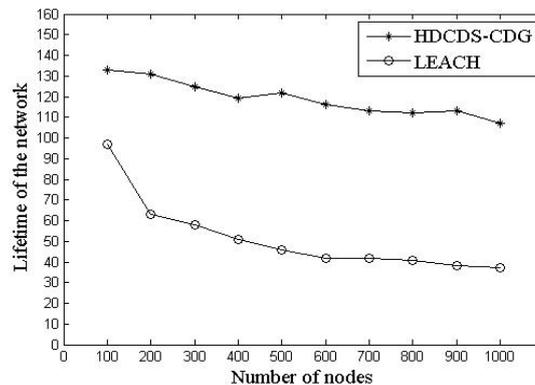


Figure 5. Network life comparison graph

5. Summary

This paper introduces a Hierarchically Diffused Connected Dominating Set-based Compressed Data Gathering method. Compared with the traditional data collection method of wireless sensor networks, the nodes near the base station need to forward the data from the nodes far away from the base station, which often causes the load imbalance of the whole network, resulting in the energy consumption of the nodes near the base station faster than the nodes far away from the base station. The HDCDS-CDG algorithm adopts the hierarchical diffusion connected dominating set election method, which makes the distribution of nodes more balanced, the selection more reasonable, and further saves energy consumption and improves efficiency.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China under Grants 61701441, 61801427 and the Zhejiang Provincial Natural Science Foundation under Grant LQ17F010001.

References

- [1] Fan, B. (2018) Key technologies of wireless sensor network security [J]. Building engineering technology and design, 30: 3397.
- [2] Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H. (2000) Energy-efficient communication protocol for wireless microsensor networks[C]//Proceedings of the 33rd annual Hawaii international conference on system sciences. IEEE, 2:3005-3014.
- [3] Alvarez, J.A., Villagra, P.E., Villalba, R., et al. (2011) Wood productivity of *Prosopis flexuosa* DC woodlands in the central Monte: influence of population structure and tree-growth habit[J]. Journal of Arid Environments, 75(1): 7-13.
- [4] Chen, X.J., Li, X.Y. (2009) Research and improvement of LEACH protocol in WSN [J]. Computer applications, 29 (12): 3241-3243.
- [5] Huang, L.X., Wang, H., Yuan, L.Y., et al. (2017) LEACH Protocol Improvement Algorithm Based on energy balanced and efficient WSN [J]. Journal of communications, 38 (Z2): 164-169.
- [6] Xie, K., Wang, X., Liu, X., et al. (2015) Interference aware cooperative communication in multi-radio multi-channel wireless networks[J]. IEEE Transactions on Computers, 65(5): 1528-1542.

- [7] Wu, D., Arkhipov, D.I., Kim, M., et al. (2016) ADDSEN: Adaptive data processing and dissemination for drone swarms in urban sensing[J]. IEEE transactions on computers, 66(2): 183-198.
- [8] Luo, C., Wu, F., Sun, J., et al. (2009) Compressive data gathering for large-scale wireless sensor networks[C]//Proceedings of the 15th annual international conference on Mobile computing and networking. 145-156.
- [9] Candès, E.J., Wakin, M.B. (2008) An introduction to compressive sampling[J]. IEEE signal processing magazine, 25(2): 21-30.
- [10] Oler, N. (1961) An inequality in the geometry of numbers[J]. Acta Math, 105(1): 19-48.
- [11] Wan, P.J., Huang, S.C.H., Wang, L., et al. (2009) Minimum latency aggregation scheduling in multihop wireless networks[C]// Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing. 185-194.
- [12] Fang, W.S., Chen, Z.H., Hu Z.D. (2019) LEACH Routing improvement algorithm for zoning and energy balance [J]. Computer engineering and design, 40(10).