ICIC Express Letters
Volume **11**, Number **2**, February **2017**
 ISSN 1881-803X
pp. 349–355

AN ENERGY SAVING DATA AGGREGATION SCHEME IN WIRELESS SENSOR NETWORKS

Jing Zhang[1,2], Pei-Wei Tsai[1,2], Jeng-Shyang Pan[1,2]
Longmei Tang[1,2] and Xiaorong Ji[1,2]

[1]School of Information Science and Engineering
[2]Fujian Provincial Key Laboratory of Big Data Mining and Applications
Fujian University of Technology
No. 3, Xueyuan Road, University Town, Minhou, Fuzhou 350118, P. R. China
jing165455@126.com

Received July 2016; accepted October 2016

Abstract. *Wireless Sensor Network (WSN) is widely used for the data gathering, especially for environmental monitoring. However, since the sensors are resources limited, energy saving and lifetime prolonging become more and more important. In this paper, we focus on the cluster and network coding based data aggregation for the area monitoring. At first, an efficient Crossed Cube based Connected Dominating Set (CDS) constructed algorithm is presented, which is based on the Induced Tree of Crossed Cube cluster. Furthermore, the adaptive random linear network coding method is used, and a Cluster-based Data Aggregation Scheme (CDAS) for Wireless Sensor Networks is presented. The corresponding analysis and experimental results indicate the feasibility and efficiency of the CDAS are greater than some existing algorithms, which does not use the cluster construction and the network coding.*
Keywords: Data aggregation, Wireless Sensor Networks, Connected Dominating Set, Network coding

1. **Introduction.** Wireless Sensor Network (WSN) is a typical self-organization network [1,2], which is composed by abundant sensors for environmental monitoring. Since the WSN can achieve data aggregating and transmission, it is a very attractive option for wide applications in different areas. However, there exists no fixed infrastructure such as switching centers or base stations in a WSN. Nodes that are within the communication range of each other can communicate directly whereas. The node that is far apart has to rely on an intermediary node to relay messages. Furthermore, most of the sensors are worked with the battery, which is resources limited. Therefore, it is essential to find an energy-saved, efficient data aggregation scheme for such network.

Network coding became widely significant after it was first proposed by Ahlswede et al. [3]. Some researches on data aggregation in WSN have been published [4-9]. Miao et al. [4] proposed an energy efficient algorithm using network coding for Gradient Based Routing (GBR). They proposed competing algorithms by considering two forward candidates. However, by the GBR algorithm, every node should be considered whether they have code opportunities, and it costs too high. Fasolo et al. [5] focused on the study of two components of a data aggregation protocol, which are routing scheme and aggregation functions. They have presented the routing scheme as the problem of forwarding packets in order to facilitate in-network data aggregation. Yousefi et al. [6] focused their study on aggregation scheduling problem and proposed an efficient distributed algorithm that generates a collision-free schedule with the least number of time slots. Shim and Park [7] proposed a practical secure data aggregation scheme, based on an additive homomorphic encryption scheme, an identity-based signature scheme, and a batch verification technique with an algorithm for filtering injected false data. Zhao et al. [8] proposed an energy-efficient scheme, which was called Treelet-based Clustered Compressive Data Aggregation (T-CCDA). Bagaa et al. [9] proposed a taxonomy and classification of existing data aggregation scheduling solutions. Chen et al. [13] proposed Connected Dominating Set based and Flow-oriented Coding-aware Routing (CFCR). However, by these algorithms, every node should be considered whether they have to code opportunities, which cost too high.

Since the self-organization of WSNs has no fixed infrastructure, topology control is one of the approaches to prolong network lifetime, and reduce interference and the packet retransmission. The purpose of using the topology control mainly is to form a virtual backbone by constructing a Connected Dominating Set (CDS). Network performance can be further optimized if the cluster based construction is used for the data aggregation. Furthermore, network coding can reduce energy consumption remarkably by forwarding only encoded packets instead of a large number of the original information, and then the cluster-based data aggregation scheme for energy saving in WSNs will be proposed in this paper.

The rest of this paper is organized as follows. Problem statement and preliminaries are given in Section 2. Section 3 presents Cluster based Date Aggregation Scheme. Section 4 gives an analysis of proposed, and Section 5 summarizes the results.

2. **Problem Statement and Preliminaries.** Sensors are placed in the interesting area randomly, and assumed that the transmission range of the sensors is the same. The link between any pair of nodes is bidirectional. One network is modeled as a connected bidirectional graph $G = (V, E)$, where V and E represent the node set and the link set in G, respectively. $\forall u, v \in V$, there exists an edge (u, v) in G if and only if u is in v's transmission range in the network, v is also in u's transmission range, and there is no obstacle preventing radio wave transmission between u and v. Consider a node u. The set of nodes covered by u is represented by $N(u)$, $N(u) = \{v \mid (v, u) \in E\}$ is called the open neighbor set of u. $N[u] = N(u) \cup \{u\}$ is called the closed neighbor set of u. A Dominating Set (DS) [10] of a graph $G = (V, E)$ is a set of nodes $V' \subseteq V(G)$ such that for every $(u, v) \in E(G)$, $u \in V'$ or $v \in V'$. A Connected Dominating Set of a graph $G = (V, E)$ is a DS of G such that the subgraph of G induced by the nodes in this set is connected.

The WSNs use a CDS to serve as a virtual backbone for energy efficiency. Many algorithms have been proposed for the purpose of fast approximations with small performance ratio for constructing Min Connected Dominating Set. Our previous research [10] has proposed the smallest one. The Induced Tree of Crossed Cube cluster algorithm, which will be used in this paper for the virtual backbone constructed. The algorithm will not be described in detail in this paper because of the page limited. Based on the cluster, the cluster head will be selected as the coding nodes. After the cluster-based routing constructed, the coding node will combine the received messages into an encoding packet and forward the packet in batch.

3. **Cluster-Based Data Aggregation Scheme.** Due to the limitation of computational power in WSNs, the adaptive random linear network coding [11] is adopted in the proposed network coding algorithm in this paper. Suppose that a sensor A generates an encoded packet by coefficients in the Galois field $GF(2^8)$, $X = \sum_{i=1}^{N} C_i D_i$, where D_i are the data which need to be encoded, and C_i are the random coefficients from $GF(2^8)$. These new encoded packets are then transmitted with the N random coefficients. Each of the receivers tries to solve this linear equation, for which at least $N_r \geq N$ packets must be received. There is one network coding scheme N which is not fixed for the entire network, and it is determined by the neighbours of the coding nodes, which is shown in Table 1 [11].

Table 1. The adaption coding scheme

Avg	11	10	9	8	7	6	5	4
scheme	$N = 8$	$N = 7$	$N = 5$	$N = 4$	$N = 3$	$N = 2$	$N = 2$	$N = 1$

The specific process of the Cluster-based Data Aggregation Scheme (CDAS) scheme can be divided into 2 stages.

The first stage is the cluster based routing. After the CDS is constructed, each node knows its own cluster head, and then sends the information to its own cluster head.

The second stage is encoding algorithm, which encodes the messages and the adaptive N random linear network coding. At first, depending on the neighbours queue, the adaptive value of N can be set. Then, the coding node selects N coefficients C_i, $i = 1, 2, \ldots, N$ from $GF(2^8)$ randomly. Also, it may receive the encoding packets from the upper layer cluster heads. If so, it needs to select 2 coefficients C_i, $i = 1, 2$ from $GF(2^8)$ randomly. Then it combines the information from its own cluster and the upper clusters: $X = \sum_{i=1}^{N} C_i D_i$. At last, the encoding packets will be sent to the next layer cluster heads.

The third stage is decoding, replying and resending the messages. Before describing this stage, the definition of the decoding rate should be given as follows.

Definition 3.1. [12] (The decoding rate). *Assuming that all the links share the same delivery ratio d_r. The probability that node can decode the packets is determined as*

$$P^N = P^{N-1}\left(r + d_r^2 - d_r^3\right)d_r^N + \left(P^{N-2} - P^{N-3}d_r^2\right)d_r^{2+N}\left(1 - d_r\right), \quad (1)$$

where $P^0 = 1$, $P^1 = d_r^2$ and $P^2 = d_r^4\left(2 - d_r^2\right)$.

First, it is easy to see that $P^2 = \sum_{i=2}^{4} C_4^i d_r^i \left(1 - d_r\right)^{4-i} d_r^2 = d_r^4\left(2 - d_r^2\right)$. When $n \geq 3$, $P^N = p[rank(C) = N]d_r^N$.

Each coding node receives the encoded packets and decodes them, if change decoded unsuccessfully. Let S refer to the number of encoded packets the considered node need to send out. $S = N - rank(C_n)$, each node sends out an S of encoded packets. We say that it is the negative ACK message. Nodes receiving less than N packets can send out $NACK$ message to some particular neighbours to retrieve missing data that they need. The nodes that have received $NACK$ message send out S encoded packets. This improves encoding algorithm with $NACK$ described as follows.

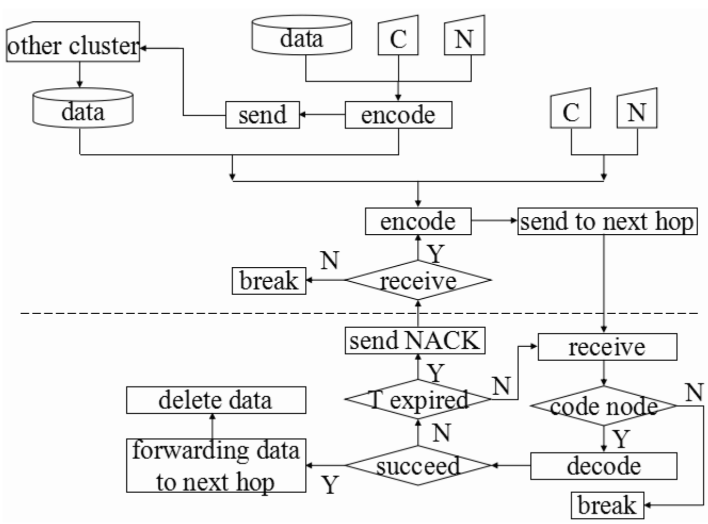

Figure 1. Cluster-based Data Aggregation Scheme (CDAS)

When the cluster head receives the encoding packets from upper cluster heads, begin the countdown. If it can be decoded within t time slot and the decoding is successful, the information is combined with their own information, and then forward to the next layer. If the decoding is unsuccessful, the coding node calculates $S \leftarrow N - rank(C_n)$, sending this S $NACKs$ to the upper cluster head. S more packages are needed to decode the information successfully. Restart the countdown. If it can be decoded within t times lot forward to the next layer. Otherwise goto this step till the encoding packet can be decoded successfully. If it receiving the $NACK$ from the next layer cluster head, execute the second stage.

The detail of the CDAS is described in Algorithm 1, and its framework is shown in Figure 1.

Algorithm 1. Cluster-based Data Aggregation Scheme (CDAS)

for each cluster head as the network coding, set a counting down time t randomly
if an encoded packet is received then
    $n\_messages$ = the messages number of this packet
    start counting down t
    construct $Matric\_C$ using coefficients from messages in $n\_messages$,
    $rank = gaussian(Matric\_C, Matrix\_inv)$
    if $rank >= N$ then
        solve the encoded packets by $Matrix\_inv$ and obtain N interest messages
        for $(j = 1; j <= S; j++)$
            select N coefficients C_i, $i = 1, 2, \ldots, N$ from $GF(2^8)$ randomly.
              generates new encoded packets by encoding the N data messages D_i given by $X = \sum_{i=1}^{N} C_i D_i$.
              broadcast the encoded packets X
        end for
    end if
end if
end for

---

4. **Numerical Analysis.**

Theorem 4.1. *The decoding rate by the Cluster-based Data Aggregation Scheme (CDAS) is* 1.

Proof: According to Definition 3.1, there is

$$P^2 = \sum_{i=2}^{4} C_4^i r^i (1-r)^{4-i} r^2 = r^4(2-r^2).$$

When $n \geq 3$, $P^N = p[rank(C) = N]r^N = P^{N-1}(r+r^2-r^3)r^N + (P^{N-2} - P^{N-3}r^2)r^{2+N}(1-r)$.

By the Cluster-based Data Aggregation Scheme, when the coding node cannot decode the encoding packets successfully, it sends the negative ACK message. The coding node calculates $S = N - rank(C_n)$, and sends the S value to some particular neighbours to retrieve missing data that they need. Later it can receive S encoded packets. Goto this step till the encoding packet can be decoded successfully. So the decoding rate by the CDAS scheme amounts to 1. □

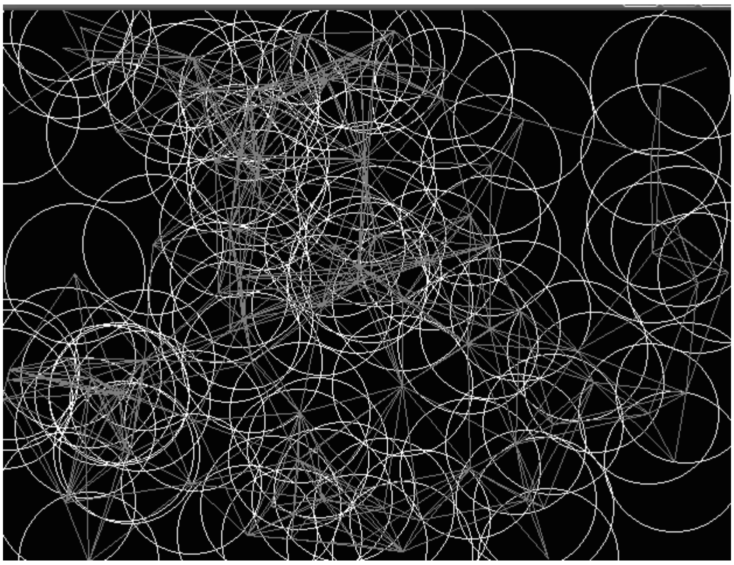

Figure 2. One example of the area containing sensors

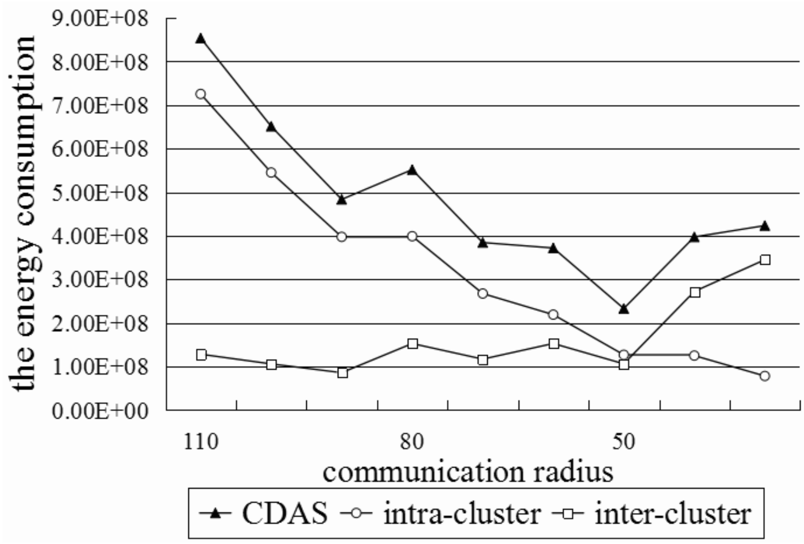

Figure 3. The intra-cluster, inter-cluster and the total energy consumption

In order to evaluate the influence of the energy consumption, three schemes are compared, which are the scheme without CDS and NC, such as GBR [4], the scheme with CDS and NC but each node has coding opportunities, such as CFCR [13], and our CDAS scheme. The density of nodes λ is a poisson distribution. Different densities have different influences on transmission energy. Figure 2 shows one example of the area containing 100 sensors with $\lambda = 2$. Figure 3 shows the intra-cluster, inter-cluster and the total energy consumption with different communication radiuses, which is decreased from 110 to 30. There is an interesting phenomenon that the total energy consumption does not always decrease. When the communication radius is large, the total energy consumption declines smoothly. When the length reaches some level, it can achieve the lowest value. Then when the communication radius continues to decrease, the total energy consumption grows swiftly, and the difference becomes larger and larger. The reason is perhaps that the intra-cluster energy consumption decreases with the communication radius reducing.

Figure 4 analyzes different cases under $\lambda = 2$, $\lambda = 3$ and $\lambda = 4$. It shows the energy consumption by these three schemes. More generally, we compare the average energy

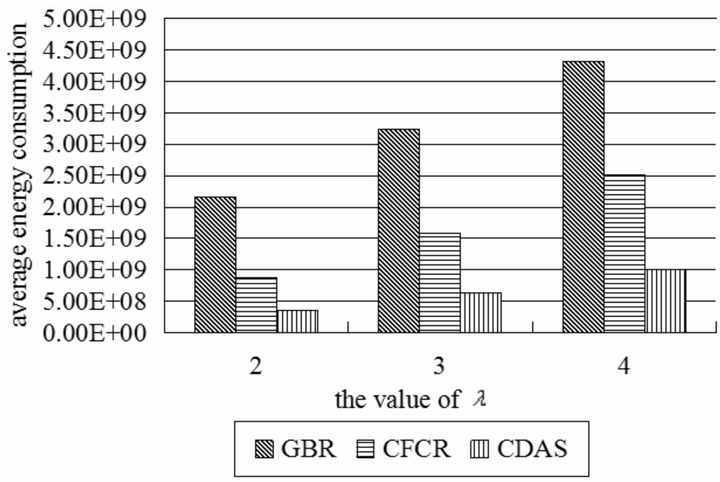

FIGURE 4. The energy consumption under different λ

consumption. Note that the energy consumption of the GBR scheme is very high, almost at $2.2E + 09$ when $\lambda = 2$. Then the CDAS scheme has the lowest energy consumption. The lowest energy consumption is only $3.46E + 08$, which is virtually one over twenty with the GBR scheme. In other wards, energy consumption by CDAS can be saved as 80% to 90% whatever the values λ are.

5. **Conclusions.** Since the self-organization of WSNs has no fixed infrastructure, topology control is one of the approaches to prolong network lifetime, and reduce interference and the packet retransmission. Network performance can be further optimized if the cluster based construction is used for the data aggregation. Furthermore, network coding can reduce energy consumption remarkably by forwarding only encoded packets instead of a large number of the original information. In this paper, we study the methods of the cluster and network coding, which are two most promising techniques to reduce the energy consumption. Based on them, the cluster-based data aggregation scheme for energy saving in WSNs is proposed in this paper. The Induced Tree of Crossed Cube cluster algorithm and the adaptive random linear network coding method are used in this scheme. The first stage is the cluster based routing. After the CDS is constructed, each node knows its own cluster head, and then sends the information to its own cluster head. The second stage is encoding algorithm, which encodes the messages and the adaptive N random linear network coding. The corresponding analysis and experimental results indicate the feasibility and efficiency of the CDAS are greater than GBR and CFCR.

For the future research direction, firstly, we will focus on security data aggregation management. Furthermore, maximize the encoding number, the number of messages that can be encoded by a coding node in each transmission, which is most researchers' purposes. However, delay will be produced when the coding nodes wait for more encoding number. In order to find a balance between delay and encoding number, we will propose a scheme based on a data collection rate prediction in the future.

Acknowledgment. The authors wish to thank the National Natural Science Foundation of China (Grant Nos: 61072080, 61572010), Key Project in Fujian Provincial Education Bureau (JZ160461, JAT160328, JA15329), the Scientific Research Project in Fujian University of Technology (GY-Z160066, GY-Z160130, GY-Z160138), Natural Science Foundation of Fujian Province of China (2014J01218), Fujian Provincial Key Project of Science and Technology (2014H0008, 2015H0009, 2014-G-83).

REFERENCES

[1] M. Dong, K. Ota, A. Liu and M. Guo, Joint optimization of lifetime and transport delay under reliability constraint wireless sensor networks, *IEEE Trans. Parallel and Distributed Systems*, vol.27, no.1, pp.225-236, 2016.
[2] J. Zhang, L. Xu, S. Zhou and X. Ye, A novel sleep scheduling scheme in green wireless sensor networks, *The Journal of Supercomputing*, vol.71, no.3, pp.1067-1094, 2015.
[3] R. Ahlswede, N. Cai, S. Li and R. Yeung, Network information flow, *IEEE Trans. Information Theory*, vol.46, no.4, pp.1204-1216, 2000.
[4] L. S. Miao, D. Karim, K. Anish and G. Noel, Network coding and competitive approach for gradient based routing in wireless sensor networks, *Ad Hoc Networks*, vol.10, no.6, pp.990-1008, 2012.
[5] E. Fasolo, M. Rossi, J. Widmer and M. Zorzi, In-network aggregation techniques for wireless sensor networks: A survey, *IEEE Wireless Commun.*, vol.14, no.2, pp.70-87, 2007.
[6] H. Yousefi, M. Malekimajd, M. Ashouri and A. Movaghar, Fast aggregation scheduling in wireless sensor networks, *IEEE Trans. Wireless Communications*, vol.14, no.6, pp.3402-3414, 2015.
[7] K. Shim and C. Park, A secure data aggregation scheme based on appropriate cryptographic primitives in heterogeneous wireless sensor networks, *IEEE Trans. Parallel and Distributed Systems*, vol.26, no.8, pp.2128-2139, 2015.
[8] C. Zhao, W. Zhang, Y. Yang and S. Yao, Treelet-based clustered compressive data aggregation for wireless sensor networks, *IEEE Trans. Vehicular Technology*, vol.64, no.9, pp.4257-4267, 2015.
[9] M. Bagaa, Y. Challal, A. Ksentini, A. Derhab and N. Badache, Data aggregation scheduling algorithms in wireless sensor networks: Solutions and challenges, *IEEE Communications Survey and Tutorials*, vol.16, no.3, pp.1339-1368, 2014.
[10] J. Zhang, L. Xu, S. M. Zhou, W. Wu and X. Ye, Induced tree of crossed cube based efficient connected dominating set algorithm in WSNs, *International Journal of Applied Mathematics and Computer Science*, vol.25, no.2, pp.295-309, 2015.
[11] I. H. Hou, Y. E. Tsai, T. F. Abdelzaher and I. Gupta, Adapcode: Adaptive network coding for code updates in wireless sensor networks, *Proc. of IEEE INFOCOM*, 2008.
[12] Z. Z. Tang, H. Y. Wang, Q. Hu and L. Hai, How network coding benefits converge-case in wireless sensor networks, *Proc. of IEEE Vehicular Technology Conference (VTC Fall)*, pp.1-5, 2012.
[13] J. Chen, K. He, R. Du, M. Zheng, Y. Xing and Q. Yuan, Dominating set and network coding-based routing in wireless mesh networks, *IEEE Trans. Parallel and Distributed Systems*, vol.26, no.2, pp.423-433, 2015.