Proceedings of the Seventh International Conference on Research in Intelligent and Computing in Engineering pp. 219–225 DOI: 10.15439/2022R16 ISSN 2300-5963 ACSIS, Vol. 33

CRSDF: Improved Network Lifespan through Chain-routing Scheme and Data Fusion in Wireless Sensor Network

Nguyen Duy Tan
Faculty of Information Technology,
Hung Yen University of Technology and Education,
Hungyen, Vietnam
tanndhyvn@email.com

Hong-Nhat Hoang
School of Computer Science and Engineering
Pusan National University
Busan, Republic of Korea
nhat.hqh@gmail.com

***Abstract*—How to use efficient energy in wireless sensor networks (WSN) is one of the major challenges due to limited energy batteries and computation capacity. Therefore, in this paper, we propose combining a chain-base routing scheme and data fusion sensor information (CRSDF for short). CRSDF contains two major works: Firstly, the chain-based routing method is applied to connect sensor nodes into a chain in which each node transmits only with the nearest neighbor using the remaining energy and distance of nodes as standard parameters to determine which node will be selected the chain leader, secondly, we fuse and compress one or more data packets to generate a result packet with small size base on the Slepian-Wolf and Dempster-Shafer theory. The simulation results exhibit that the energy efficiency of our proposed protocol can be improved by 40%, 20%, and 15% compared to low-energy adaptive clustering hierarchy (LEACH), power-efficient gathering in sensor information system (PEGASIS), and an improved energy-efficient PEGASIS-Based protocol, respectively.**

***Index Terms*—Energy-efficient, routing protocol, chain-based clustering, wireless sensor networks, data fusion.**

I. Introduction

WSN consists of many sensor nodes that are deployed in diverse fields with different applications such as healthcare, environment monitoring, smart cities, intelligent transportation system, and so on [1]. Sensor nodes are small in size, limited bandwidth, processor abilities, memory, and resources, particularly, the battery energy is very difficult to recharge and substitute. Therefore, it is very valuable for designing routing protocols to efficiently employ energy resources and lengthen the lifespan of the network. Chain-based clustering routing protocol and data compression are satisfactory methods to decrease energy consumed and extend the lifespan of sensor networks [2, 3]. If the chain-based clustering routing method enormously reduces the distance communication among nodes in the network, then data aggregation will remove the data redundancy gathered from surrounds in various sensors to obtain more accurate and concise information. Furthermore, data pressure also is a suitable solution for maintaining battery energy by abating the number of bits to be transferred, however, the normal compression method as Lempel-Ziv; Huffman is difficult acceptance for sensor networks because they need a strong processor and large storage capability [4, 5] for activities. Distributed Source Coding (DSC) technique [6, 7], which is proposed by Slepian-Wolf, perform compresses the lossless data of two correlated sources employing side information. The data correlated can be the observed stream or the data in the past of its neighbors. DSC is one of the most relevant solutions for WSNs to conserve energy due to the storage and processing constraint of sensor nodes. Furthermore, in WSN, sensor nodes usually are deployed densely in a sensor region to monitor the environment, so, this correlated condition can be easily satisfied [8, 9]. The hierarchical clustering routing technique is known as a good solution to limit energy consumption, and prolong the lifespan of the network. For instance LEACH [10], in LEACH the nodes are grouped into a few clusters. Each cluster votes for a leader node called the cluster head node (CHN) to take the responsibility for transmitting data to the sink device, other nodes (called cluster members) will only send measured data to its CHN. However, the communication single-hop intra-cluster consumes a lot of energy because the coverage area of clusters is wide. To overcome this limitation, the chain-based routing technique is proposed to maximize the network lifespan such as PEGASIS [2]. According to this technique, sensor nodes are connected and communicated with the closest neighbor in a long chain. It has been verified that the performance of the PEGASIS protocol is better the LEACH protocol in terms of network lifespan and energy efficiency. However, the selecting chain leader node (CLN) is the same as CHN without considering the remaining energy, and the distance from candidate nodes to the sink is also a limitation. Therefore, there have been many researchers focused on improving the chain routing scheme base on PEGASIS, typically, an improved energy-efficient PEGASIS-based protocol is introduced by Sen et al. [11], an energy-efficient chain-based routing protocol for orchard WSN, [12], or enhanced energy-efficient routing for WSN using extended power-efficient gathering in sensor information systems [13], and so on. However, none of the above improvements consider the collaboration between chain routing and data fusion to enhance energy efficiency.

In this paper, we propound a data fusion and chain-based clustering routing scheme (called CRSDF), which can achieve energy and bandwidth efficiency by eliminating redundant data by fusing and avoiding "long links" in the chain. In CRSDF, the Greedy algorithm is employed to build the chain like IEEPB [11], but CRSDF chooses CHN in rounds by deliberating the remaining energy of candidate nodes and the distance between theirs and the sink to determine which node will be voted CHN. CHNs will be responsible for forwarding data to the sink. In addition, the Slepian-Wolf and Dempster-Shafer (DS) evidence theory is also used to aggregate data going along the chain, and then the DSC scheme is used to reduce data packet size in correlated sources. The simulation results exhibit that the network lifetime of CRSDF can be lengthened by 40%, 30%, and 10% compared to LEACH, PEGASIS, and IEEPB respectively.

The remainder of this paper is organized as follows. Section II presents the related works and Section III presents the system model, Section IV analyzes the data fusion architecture, Section V describes the details of CRSDF. In Section VI, we evaluate and analyze the experiment results. Finally, we conclude the study in Section VII.

II. RELATED WORKS

There are many researchers focused on improving energy-efficient chain-based routing protocols. For example, in [11], Sen et al. proposed an improved energy-efficient PEGASIS-based (IEEPB for short), in which the criteria for selecting CLN is considered both the remaining energy and distance among nodes and the sink. Moreover, in order to reduce the "long link" between nodes in the building chain, the IEEPB selects a node that is not only compared to the end node of the chain but also compared to the other nodes in the chain to find the closest node for connecting into the chain. The simulation results show that IEEPB overcomes PEGASIS in terms of network lifespan and energy efficiency. However, the chain still contains long links since the algorithm for building the chain is still not optimal.

Recently, Zi et al. [3] proposed a novel chain-based routing protocol (BranChain) to avoid long links in a chain by connecting nodes in an independent branched chain (IBC) in the chain constructing phase. Each IBC contains several nodes that are connected by finding the optimal paths between two small chains. Although this method can improve network lifespan, the complexity algorithm is increased. Sadhana et al. [13] proposed an extended PEGASIS protocol called E-PEGASIS to enhance energy-efficient by combining the remaining energy and distance between the node and the sink in the chain leader (CL) selection phase. In addition, Wu et al. [12] propose an improved chain-based clustering hierarchical routing based on the LEACH algorithm called ICCHR, where the remaining energy and distance among nodes and the sink is attached to the probability function T(n) as a criterion for electing CHN, and then, CHNs are connected together into chain beginning the furthest CH from the sink based on the greedy algorithm. However, most of the current routing protocols do not consider data aggregation/fusion to reduce the number of packets that need to be transmitted to the sink, thereby significantly reducing energy consumption. Huy and Viet [14] proposed sliding windows for multi-sensor data fusion (DF-SWin) where the concepts of data mining, sliding window, and Rough set are integrated to get several specimens from sensor nodes in the cluster based on the energy level, distance from the node to CHN, and the residual packet properties to select value parameters for data fusion in CH node. In [9], Ullah et al. proposed multi-sensor data fusion based on modified belief entropy in Dempster–Shafer theory for heterogeneous sensors (DFUDS). DFUDS applied the rules of measuring uncertainty in DS theory for incorporating the data from sensors to achieve more accurate and concise quantification of data. Besides data aggregation, data pressure is also a technique to diminish the size of data packets transmitted to the sink, thus also saving energy. Hau et al. [20] introduced the improvement of the AODV routing protocol by using intelligent agents for efficient energy. Tan et al. [21] proposed a sector tree based on a clustering routing protocol for efficient energy, in which the nodes in a cluster communicate with each other based on the minimum spanning tree connecting. Sadler et al. [15] proposed the sensor Lempel Ziv Welch (S-LZW) algorithm based on the LZW algorithm for sensor systems by balancing three parameters: the dictionary size, and the data size for compression and processes with a full dictionary. In [4], Malleswari et al. introduced the implementation of modified Huffman coding for WSN, where the Huffman algorithm is employed for the compression and decompression of data in order to minimize the size of the data packet and save energy.

III. SYSTEM MODEL

A. Network Model

In our network model, we hypothesize that a network system consists of a gateway or sink device and a lot of sensor nodes that can support applications [3, 10, 11].

- The gateway device is fixed and not limited to energy and computation capacity, which can aggregate and forward the data packets.
- N sensor nodes are randomly deployed in the network zone of two-dimensional.
- All sensor nodes are static state after deployment and know the location of each other based on the global positioning system (GPS)
- All sensing nodes use the battery energy and cannot recharge or replace
- The radio channels are symmetry

B. Energy Consuming Model

In the radio energy dissipation model used in our proposal to transmit *q-bit* data, the energy consumed may be computed as Equation (1) [10, 12]:

$$E_{TX}(q,d) = \begin{cases} q \times E_{elec} + b \times E_{friis} \times d^2 & ,if\ d < d_0 \\ q \times E_{elec} + q \times E_{tworay} \times d^4 & ,if\ d \geq d_0 \end{cases} \quad (1)$$

where E_{elec} is the energy consumption part for the electric circuit, E_{towray} and E_{friis} are the part of energy consumption by the amplifier in two ray ground mode and free space mode with the distance d, corresponding, and d_0 is the threshold distance between two nodes. The geographical distance between nodes a and b is calculated as Equation (2) below:

$$d_{ab} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (2)$$

The energy dissipated in receiving *b-bit* data is calculated by the number of bits and energy dissipated by the electric circuit as Equation (3) below:

$$E_{RX} = qE_{elec} \quad (3)$$

IV. DATA FUSION AND FRAMEWORK

A. Distributed Source Coding

DSC based on the Slepian-Wolf theorem is one of the most efficient techniques, to compress correlated data sources [4, 16]. In DSC, the correlated signals from a few sensor nodes are compressed with a totality rate greater than or equal to the joint entropy thus they decrease data packet size. For instance, we have two distributed derivations X and Y which create u bits of binary data as demonstrated in Fig.

1. We assume that the data of sources X and Y are highly correlated and can be different by no more than 1 bit. In other words, the Hamming distance between X and Y is $d_H(X, Y) \leq 1$ or $P_r(X_i = 0) = P_r(Y_i = 0) = P_r(X_i = 1) = P_r(Y_i = 1) = 0.5$, $i = 1, ..., u$. So, $H(X) = H(Y) = u$ bits, where $P_r(X)$ and $H(X)$ denote the probability distribution density function of random source X and the entropy function, respectively. In this case, if Y is sent to the decoder as side information that contains u bits, then we can send X with $H(X|Y) = b$ bits per specimen, where b is the index number of the subset, without any loss data at the joint decoder.

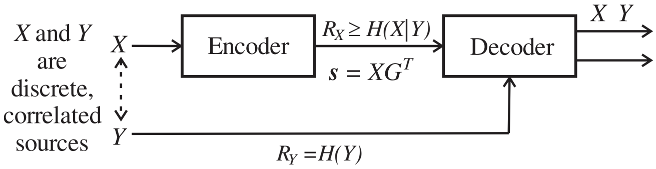

Fig. 1. Distinct encoding and common decoding of two correlated data sources Y and X

We suppose that Slepian-Wolf coding of X and Y origins are equiprobable 2^u specimens ($u = 4$ bits binary per specimen).

Let $k = u - b$, where k is a integer and $2^k - 1 \geq d_H(X, Y)$, then $H(X) = H(Y) = 4$ bits per specimen, $H(X|Y) = b = 3$ bits, and $H(X, Y) = 4+3=7$ bits per pair of specimen for common decoding, where $H(X|Y)$ and $H(X,Y)$ denote the conditional entropy and the joint entropy, respectively. Consequently, we can group 2^4 specimens as $2^3 = 8$ subsets, in each subset the Hamming distance between any two elements is greater than or equal to 4, and assign 8 different binary index numbers by Z_{xxx}, respectively, as follow:{Z_{000} = (0000, 1111): 000, Z_{001} = (0111, 1000): 001, Z_{010} = (0100, 1011): 010, Z_{011} = (0011, 1100): 011, Z_{100} = (0010, 1101): 100, Z_{101} = (0101, 1010): 101, Z_{110} = (0110, 1001): 110, Z_{111} = (0001, 1110): 111}. We study how to send data of source Y, whose data is used as side information, as {0100} and the index of X as {110}. When the decoder received the index as {110}, the probable value of X is set {0110, 1001} in Z_{110}. The decoder will obtain the precise value of X as {0110} since the Hamming distance between X and Y is less than or equal to one ($d_H(X, Y) = 1$).

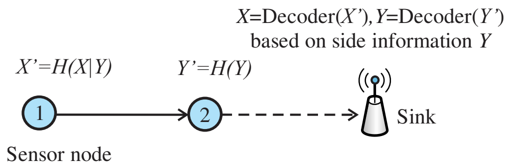

Fig. 2. Execution of distributed source coding in WSN

In fact, the side information Y is perfectly available at the decoder (e.g. sink device) since, in a sensor network, where the sensor node is deployed densely in a sensor field, these correlated data conditions can be satisfied easily. However, the collaboration of encoders is not easy to perform in the sensors node, as exhibited in Fig. 1 and 2, the Y is only ready for the decoder but not the encoder, therefore, how can we press two origins X and Y into a total $(u + b)$ bits as the mentioned example with lossless decoding? To solve this puzzle, we can employ a Linear Block Code (LBC) model in Galois Field two ($GF(2)$) representing in the form of a (u, b) LBC, which is given by $\boldsymbol{G = [I_m: P]}$ and $\boldsymbol{H = [P^T: I_{u-b}]_{(u-b)\times u}}$ with the property that $GH^T = 0$, where G is a b by u binary generator matrix and $\boldsymbol{P^T}$ is the exchange of the b by $(u - m)$ matrix $\boldsymbol{P}$. $\boldsymbol{H}$ is a parity check matrix used at the decoder using syndrome decoding, while G is used to encode the message. With the above example, we have an H and G matrix as shown follows:

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad H = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$

Therefore, the encoder can simply compute the syndrome of the source $\boldsymbol{s = XG^T}$, (associated with all $X \in Z_s$), which is the index of X using $H(X|Y) = b$ bits of syndrome instead of u bits original data, achieving the Slepian-Wolf coding in a pressure ratio of $u:b$ [16]. The decoder will associate the side information Y and all subsets to find out the original of X.

B. Dempster-Shafer theory

In WSN applications, sensor nodes measure the environment and periodically send correlated high data packets to the sink that is connected to the user's PC or through the internet. In order to preserve the battery energy of nodes, we can apply DSC techniques as above. However, it will be much extravagant energy in sending the unnecessary collected data because when nodes are very close, the observed values are the same. Fig. 3(a) illustrates the number of packets transmitted in a chain without data aggregation, in which there are seven nodes in a chain transmitting the overall seven packets along the chain to the sink. The total demolish energy may be computed as Equation (4):

$$E(a) = 14E_{Rx}(q) + 21E_{Tx}(q, d(i, j)) + 7E_S(q) \tag{4}$$

$$E(b) = 6E_{Rx}(q) + 7E_{Tx}(q, d(i, j)) + 4E_{DF}(q) + 7E_S(q) \tag{5}$$

where $E(b)$ and $E(a)$ are energy consumption totality with data fusion and without data fusion, respectively. Let the distance between every two nodes is d and the packet size is $q = 1$ bit, then we have:

$$E(a) = 35E_{elec} + 21E_{friss}d^2 + 7E_G + 7E_S \tag{6}$$

$$E(b) = 13E_{elec} + 7E_{friss}d^2 + 4E_{DF} + 7E_G + 7E_S \tag{7}$$

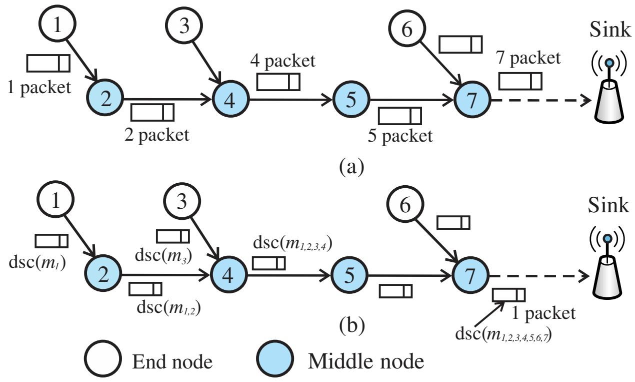

Fig. 3. The model of data sending in the chain (a) without data fusion and (b) with data fusion

Accordingly, it is clear that the energy exhaustion without data fusion is higher than about three times in comparison to data fusion. However, how to aggregate two or more correlated data packets from many sensor nodes (multi-sensor for short) into an individual packet as shown in Fig. 3(b)? In our proposal, we apply the DS evidence theory to reduce redundant information and improve the reliability of inference based on collected data from the external environment [17, 18, 22]. According to DS reasoning, all mutually exclusive events have the same style are enumerated in "a finite set Θ". Assuming that $\Theta = \{\theta_1, \theta_2, ..., \theta_v\}$, where θ_i ($i=1,v$) is the observation mutually exclusive or hypothetical possibilities that are obtained from the source sensor. The overall possible subsets of Θ are 2^v subsets.

$$2^v=\{\phi, \theta_1, \theta_2, ..., \theta_v, \{\theta_1, \theta_2\}, ..., \{\theta_1, \theta_2,..., \theta_i\},..., \Theta\} \quad (8)$$

Let m_i be "probability mass function" of the sensor *i-th* (i= identification), which the detected value in subset "*A*", and denoted by a "confidence interval" as [*Belief(A)*, *Plausibilityl(A)*] [9].

$$\begin{cases} \sum_{A\subseteq\Theta} m(A) = 1 \\ m(\phi) = 0 \end{cases} \quad (9)$$

where ϕ and *m(A)* is the null subset of Θ and probability mass function that shows how the accurate evidence for supporting hypothesis A is possible, respectively.

The lowest zone of the confidence interval is called *Belief(A)*, which is the total of the probability mass *m(A)* and supports for the reasons to believe in the values "θ_i" in subset *A*, including itself.

$$Belief_i(A) = \sum_{P_A\subseteq A} m_i(P_A), \ \forall A \subseteq 2^v \quad (10)$$

The upper zone of the confidence interval is the sum of all the probability mass of the subsets that intersect with subset *A*.

$$Plausibility_i(A) = 1 - \sum_{P_A\cap A=\phi} m_i(P_A), \ \forall A \subseteq 2^v \quad (11)$$

DS has also given a rule of combining for fusing (⊕) two sensor sources called $m_{12}(A)$ with the subset $A \neq \phi$ and $m_{12}(\phi)$=0 [18].

$$m_{12}(A) = m_1 \oplus m_2(A) = \frac{\sum_{B\cap C=A\neq\phi} m_1(B)m_2(C)}{1-\sum_{B\cap C=\phi} m_1(B)m_2(C)} \quad (12)$$

For example, as shown in Fig. 3(a), consider two sensors 1 and 2 observe together the outside environment temperature, obtain values Θ={"θ_1=30^{0}C", "θ_2=27^{0}C", "θ_3=unknow"}, where the θ_3 cannot determine the measurement value.

$$2^3=\{\phi,\theta_1,\theta_2, \theta_3, \{\theta_1, \theta_2\},\{\theta_1, \theta_3\},\{\theta_2, \theta_3\},\{\theta_1, \theta_2, \theta_3\}\}$$

TABLE I. THE RESULTS OF DATA FUSION OF THE DS'S RULE OF COMBINATION

			Sensor 1			Fused data
	Mass function		$m_1(\theta_1)$	$m_1(\theta_2)$	$m_1(\theta_3)$	
			0.2	0.7	0.1	1.0
Sensor 2	$m_2(\theta_1)$	0.6	{θ_1}=0.12	{ϕ}=0.42	{θ_1}=0.06	0.385
	$m_2(\theta_2)$	0.3	{ϕ}=0.06	{θ_2}=**0.21**	{θ_2}=**0.03**	**0.596**
	$m_2(\theta_3)$	0.1	{θ_1}=0.02	{θ_2}=**0.07**	{θ_3}=0.01	0.019
Fused data		1.0	0.385	**0.596**	0.019	1.0

The probability mass function achievements as follows: with sensor 1, $m_1(\theta_1)$=0.2, $m_1(\theta_2)$=0.7, $m_1(\theta_3)$=0.1; with sensor 2, $m_2(\theta_1)$=0.6, $m_2(\theta_2)$=0.3, $m_2(\theta_3)$=0.1, the other value are zero.

Table 1 illustrates the results of the calculation after applying the combination rule of DS, in which:

$$m_{12}(\theta_1) = \frac{0.12+0.02+0.06}{1-(0.06+0.42)} = \frac{0.2}{0.52} = 0.384$$

$$m_{12}(\theta_2) = \frac{0.21+0.07+0.03}{1-(0.06+0.42)} = \frac{0.31}{0.52} = 0.596$$

$$m_{12}(\theta_3) = \frac{0.01}{1-(0.06+0.42)} = \frac{0.01}{0.52} = 0.019$$

Accordingly, to fuse the data of *N* sensor nodes, each node generates a sequence of *v* values, we can apply Equation (13) as below:

$$m_{12,...,v}(A) = m_1 \oplus m_2,...\oplus m_v(A) = \frac{\sum_{\cap_i C_i = A\neq\phi}\left(\prod_{0\le i\le v} m_i(C)\right)}{1-\sum_{\cap_i C_i=\phi}\left(\prod_{0\le i\le v} m_i(C)\right)} \quad (13)$$

V. PROPOSED PROTOCOL

Here, we depict the detail of CRSDF protocol, which is inspired by the PEGASIS protocol [2, 3, 19]. The operation of the CRSDF protocol is divided into rounds, each round composed of four phases: (1) chain leader selection, (2) chain formation, (3) data fusion, and (4) data transmission phase.

A. Phase 1: Chain Leader Selection

In this phase, all nodes alive will send a hello message containing ID, location information, and residual energy to the sink. The sink will select the CLN having the residual energy higher than E_{ave} and maximum cost function as Equation (14) and (15):

$$E_{ave} = \frac{1}{n}\sum_{i=1}^{n} E_{res}(i) \quad (14)$$

Where *n* and $E_{res}(i)$ are the overall living nodes and the residual energy of contestant node i-*th* at the present time, respectively.

$$\cos t(i) = Max\left(\frac{c_1}{c_2} \times \frac{E_{res}(i)}{d_{i,Sink}}\right) \quad (15)$$

Where $d_{i,Sink}$ is the distance from the node i-*th* to the sink device, which is computed as Equation (2). Besides, the coefficient of cost factors values c_1 and c_2 are constant, and ($c_1 + c_2 = 1$).

Algorithm 1 Chain Leader Selection Phase

Input: N sensor nodes
Output: chain leader node (CLN)
1: **for** i = 1 **to** N **do**
2: Send a HELLO message containing the remaining energy and the location to the Sink
3:**end for**
4: The Sink calculates average energy as in Equation (15)
5: Chooses the chain-leader node, of which the cost function value as in Equation (16) is the highest
6: Broadcast CLN-ADV to all nodes alive in the network
7: Nodes receive the CLN-ADV message and go to **Algorithm 2**

return: CLN

B. Phase 2: Chain Formation

Fig. 4 shows an example of the network topology acquired by running different schemes with 100 sensor nodes in one round. Here, our proposed chain is shorter distance communication than the PEGASIS pattern of the network. After selecting CLN, the sink distributes this information to all living nodes to construct a chain as in Algorithm 2 below based on the Greedy algorithm to reduce the "long link" in the undirected weight graph problem.

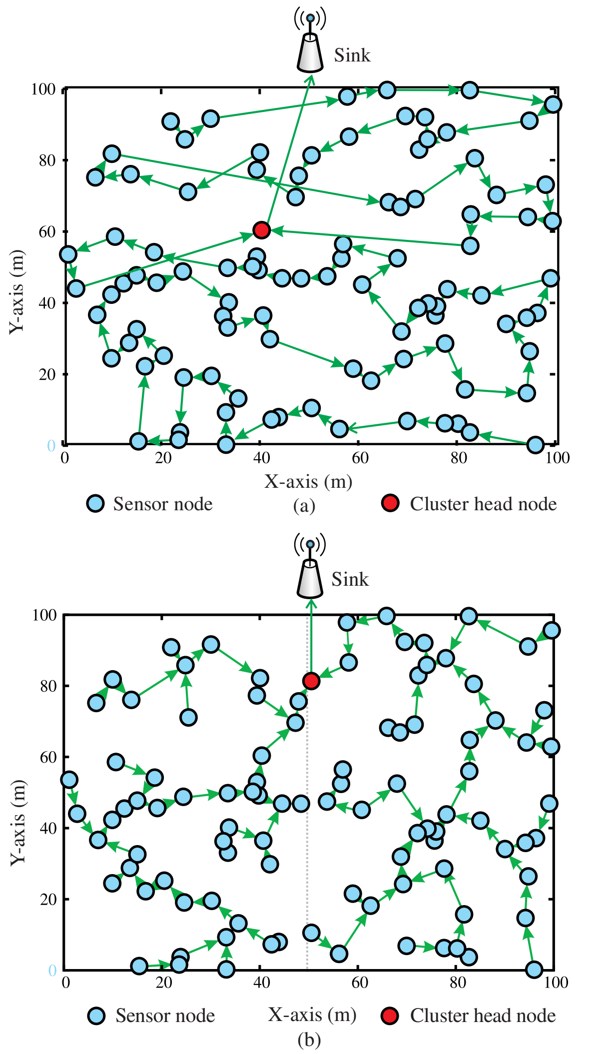

Fig. 4. The network topology with: (a) PEGASIS and (b) CRSDF protocols

Algorithm 2 Chain Formation Phase

Input: N sensor nodes, CLN
Output: CHAIN

CHAIN = {CLN} *// the chain begin by leader node*
cloestNode = {the closest node to the CLN in N};
CHAINP = {*cloestNode* } *// the chain begin by the closest node to the sink*
$N = N \setminus$ {*cloestNode,* CLN};
while (N ≠ ∅) **do**
  Find node *i* in *N* and node *j* in CHAIN, whose $d_{i,j}$ is minimum;
  Find node *k* in CHAINP, whose $d_{i,k}$ is minimum;
  if ($d_{i,j} < d_{i,k}$) **then**
    Connect node *i* to node *j* in CHAIN;
    Append node *i* into CHAIN;
  else
    Connect node *i* to node *k* in CHAINP;
    Append node *i* into CHAINP;
  endif
  Discard node *i* in *N*
end while
Connect CHAINP to CHAIN by connecting the node *cloestNode* with CLN
Create time slots according to the TDMA schedule for all member node
Broadcast CHAIN and TDMA information to the network
return CHAIN;

C. Phase 3: Data Fusion

After constructing the chain, CLN generate a TDMA schedule and broadcast them to all member nodes in its chain. The sensor nodes start measuring, aggregating, and compressing data as in Algorithm 3 follow:

Algorithm 3 Data Fusion Phase

Input: The list of data packets Lp
Output: a single data packet with a small size

if (node *i* in {CHAIN} is a middle node or CLN) **then**
if (the data packet needs to decompress) **then**
for each packet *p* in {list of data packets}
for every 7 bits in compressed data packet p in the decoder **do**
  Y = read the high 4 bits as side information
  s = read the lower 3 bits
  Search value *X* in subset Z_s, which has Hamming distance with *Y* is less than or equal to 1;
  Append *Y* and *X* into data packet *pp* decompression;
end for
  Delete the packet p in {list of data packets}
  Update the packet *pp* into {list of data packets}
end for
end if
//Data fusion based on D-S
  for each packet *p* in {the list of data packets} **do**
    Calculate the mass function $m_i(A)$ based on Equation (9)
  end for
  Fuse data for all packets as Equation (13)
end if
// This code for data lossless compress as DSC
  for every 8 bits in the data packet **do**
    Y = read the high 4 bits as side information
    X = read the lower 4 bits as original data
    Calculate $s = X*G^T$
    Append *Y* and *s* into a new data packet compression p
  end for
return: {a single packet p}

D. Phase 4: Data Transmission

After completing the first two phases above, the data packets are processed and transmitted in CRSDF starting at the furthest node in the chain. Nodes will transfer the observed data to the next node in the time slots designated by the TDMA mechanism along the chain. With the chain constructed in CRSDF have more than two end nodes as shown in Fig. 4(b). The middle nodes or CLN will receive the data, decompress these data packets and aggregate them with their own data, re-compress, and then it forwards this single packet to the next node along the chain or the sink. After a period of time, the next round will be restarted by reselecting CLN and reconstructing the chain for a new round.

VI. Experimental And Performance Evaluation

A. Simulation Parameters

To evaluate the performance of CRSDF, we have simulated CRSDF, LEACH, and PEGASIS in the network simulator ns-2.34 [21, 22, 23] and C/C++ using the scenario with some parameters that are indicated in Table 2, [10, 13].

TABLE II. THE SIMULATION PARAMETERS

No. Item	Parameters Description	Value
1	Size of simulation field	100m ×100m
2	Number of sensor nodes (*N*)	100 nodes
3	Energy consumption: two ray ground model (E_{towray})	10 pJ/bit/m^2
4	Energy consumption: free space model (E_{friis})	0.0013pJ/bit/m^4
5	Energy consumption: Electric circuit (E_{elec})	50 nJ/bit
6	Energy consumption: Data fusion (E_{DF})	5 nJ/bit/packet
7	The initial energy of node (E_{init})	2J
8	Packet size	1024 bytes
9	Simulation time	3600s
10	Sink location	49,175

B. Simulation results

Fig. 5 displays the simulation result of the total number of dead nodes in the network according to the number of rounds of four schemes: LEACH [10], PEGASIS [2], IEEPB [11], and our proposed CRSDF. We can see that the CRSDF has a longer network lifespan than LEACH and PEGASIS and IEEPB. The network using the LEACH protocol stops working at about 100 rounds, the PEGASIS prolongs its life to about 1400 rounds, and IEEPB can preserve the lifecycle to about 1650 rounds, whereas the CRSDF protocol can lengthen its lifespan to about 1750 without compression and 1900 rounds with compression. This improvement is due to the reduction of the distance transmission from the CHs node to the sink device and the avoidance of long links in the chain, the more rounds, the longer the network lifetime.

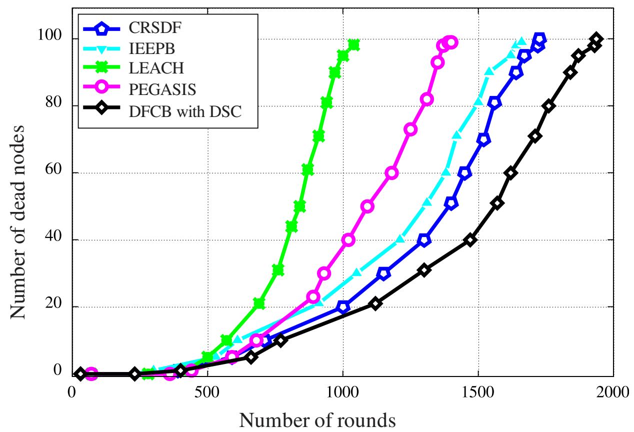

Fig. 5. Number of nodes alive during the simulation time in (rounds)

In Fig. 6, we describe the energy consumption of four protocols according to the simulation time (rounds). Based on the results displayed in Fig. 6, we can observe that the energy consumption nodes in the network running the proposed protocol is less than about 50%, 30%, and 20% compared to LEACH, PEGASIS, and IEEPB with data compression, respectively.

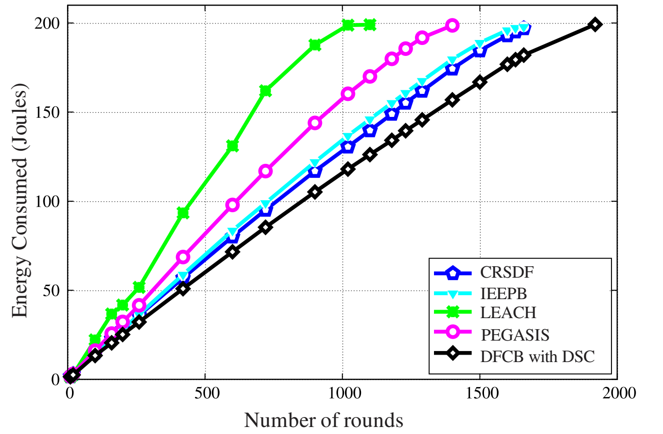

Fig. 6. Energy consumption of three protocols

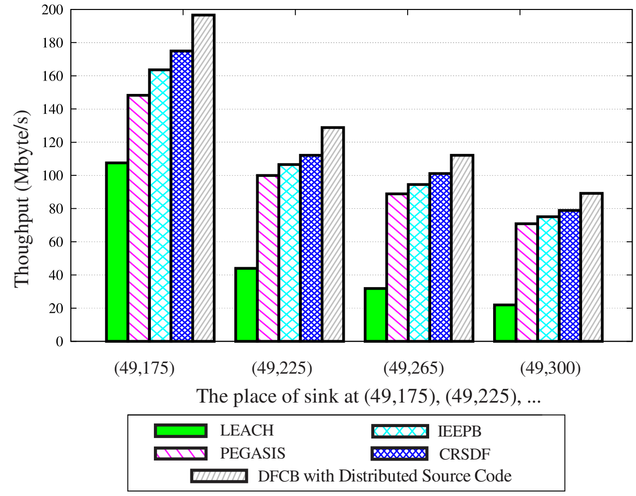

Fig. 7. The network throughput with the different places of the sink

As shown in Fig. 7, the network throughput is described as the amount of data packets transported success from the overall sensor nodes to the sink device in unit time. We can observe that network throughput will decrease if the distance from network nodes to the sink device is because CHN nodes consume a lot of energy for transmitting data over links with far distances, however, CRSDF still has in better performance terms of throughput as compared to the other three protocols mentioned.

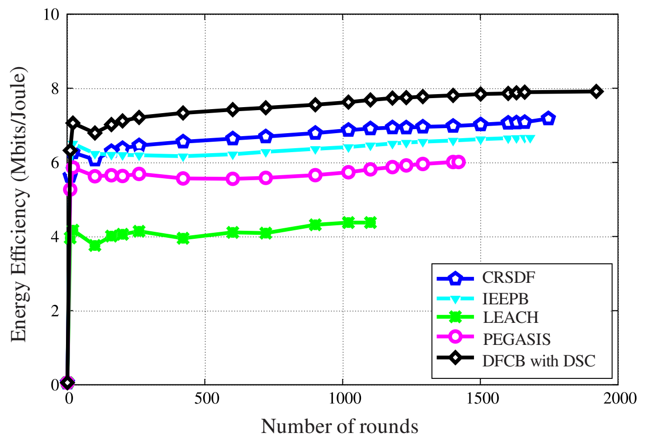

Fig. 8. The number of messages received at the sink

Fig. 8 presents the energy efficiency in Megabits per Joule according to the total number of rounds in running various routing protocols. We can observe clearly that our proposed protocol achieves higher energy-efficient than the other three protocols: LEACH, PEGASIS, and IEEPB about 30%, 20%, and 15%, respectively. The reason for this enhancement is the CRSDF minimizes the energy consumption in the communication process among nodes by decreasing the data packet size before forwarding it to the next node within the chain and the sink.

VII. Conclusion

In this paper, we presented collaboration among a chain-based routing scheme and data fusion to improve energy efficiency, namely CRSDF, which can extend the network lifespan by constructing a chain connecting the nearest neighbor nodes and fusing, compressing sensed data in order to reduce the number of bits in the data packet before sending to the sink device. Simulation results display that the energy efficiency of CRSDF is better than LEACH, PEGASIS, and IEEPB about 30%, 20%, and 15%, respectively. In the future, we study energy efficiency hierarchical routing protocol for applications IoT based on wireless sensor networks.

Acknowledgment

This work was supported by Hung Yen University of Technology and Education under grant code UTEHY.L.2022.01.

References

[1] Yang, X., Pei, X., Chen, G., Li, T., Wang, M., & Wang, C. (2019). A Strongly Unforgeable Certificateless Signature Scheme and Its Application in IoT Environments. *Sensors*, 19(12), 1 – 27.

[2] S. Lindsey and C. S. Raghavendra, "Pegasis: Power-efficient gathering in sensor information system," in *IEEE Aerospace Conference Proceedings*, March 2002, pp. 1125 – 1130.

[3] L. Zi, W. Chen, X. Liu, and X. Chen, "A novel chain-based routing protocol, branchain, in wireless sensor networks," *International Journal of Embedded Systems*, vol. 11, no. 3, pp. 259–268, 2019.

[4] M. Malleswari, B. A. Krishna, N. Madhuri, and M. K. Chowdary, "Implementation of modified huffman coding in wireless sensor networks," *International Journal of Computer Applications*, vol. 177, no. 1, pp. 14–17, 2017.

[5] K. Sayood, *Lossless Compression Handbook*. California, USA: Academic Press, 2003.

[6] D. Slepian and J. K. Wolf, "Noiselesscoding of correlated information source," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[7] P. Kumar and V. Gupta, "Data compression using distributed source coding in wireless sensor network," *International Journal of Electronics Communication and Computer Technology*, vol. 2, no. 1, 2011.

[8] K.-C. Lan and M.-Z. Wei, "A compressibility-based clustering algorithm for hierarchical compressive data gathering," *IEEE Sensors Journal*, vol. 17, no. 8, pp. 2550–2562, 2017.

[9] I. Ullah, J. Youn, and Y.-H. Han, "Multisensor data fusion based on modified belief entropy in dempster–shafer theory for smart environment," *IEEE Access*, vol. 9, pp. 37 813–37 822, 2021.

[10] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE Transactions on wireless communication*, vol. 1, no. 4, pp. 660–670, 2002.

[11] F. Sen, Q. Bing, and T. Liangrui, "An Improved Energy-Efficient PEGASIS-Based Protocol in Wireless Sensor Networks," in *Eighth International Conference on Fuzzy Systems and Knowledge Discovery*. Shanghai, China, July 2011, pp. 2230–2233.

[12] H. Wu, H. Zhu, L. Zhang, and Y. Song, "Energy efficient chain based routing protocol for orchard wireless sensor network," *Journal of Electrical Engineering & Technology*, vol. 14, p. 2137–2146, 2019.

[13] S. S, S. E, and D. D, "Enhanced energy efficient routing for wireless sensor network using extended power efficient gathering in sensor information systems (e-pegasis) protocol," *Procedia Computer Science*, vol. 194, p. 89–101, 2021, the 18th International Learning & Technology Conference.

[14] H. Duong-Viet and V. Nguyen-Dinh, "Df-swin- sliding windows for multi-sensor data fusion in wireless sensor networks," in *The 9th International Conference on Knowledge and Systems Engineering (KSE)*. Hue, Vietnam, October 2017, pp. 54–59.

[15] C. M. Sadler and M. Martonosi, "Data compression algorithms for energy-constrained devices in delay tolerant networks," in *Proceedings of the 4th international conference on Embedded networked sensor systems*. Boulder, Colorado,USA, November 2006, pp. 265–278.

[16] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80 – 94, 2004.

[17] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[18] E. D. C. Bezerra, A. S. Teles, L. R. Coutinho, and F. J. da Silva e Silva, "Dempster–shafer theory for modeling and treating uncertainty in iot applications based on complex event processing," *Sensors*, vol. 21, no. 5, pp. 1–26, 2021.

[19] Tan, N, D.,& Viet, N, D. "SCBC: Sector-Chain Based Clustering Routing Protocol for Energy Efficiency in Heterogeneous Wireless Sensor Network", Advanced Technologies for Communications (ATC), Ho Chi Minh, Vietnam, 2015, pp. 314-319.

[20] Van-Hau, N, Nam, V. H., Linh, D, M., and Quy, V. K., (2021), An improved agent-based AODV routing protocol for MANET, EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 8(27), 2021.

[21] Tan, N, D., Quy, K, V., Hung, P, N., & Vinh, L, V. (2021). Sector Tree-Based Clustering For Energy Efficient Routing Protocol in Heterogeneous Wireless Sensor Network, International Journal of Computer Networks & Communications (IJCNC), 13(5), 57-74.

[22] W. Heinzelman, "MIT uAMPS LEACH ns Extensions," https://blog.katastros.com/a?ID=01700-10dfaae3-e3c2-4f2f-8235-608c3010d995 (accessed: Feb 29, 2022), 2004.

[23] VINT Project, "The network simulator - NS2," http://www.isi.edu/nsnam/ns (accessed: Feb 2, 20212, 1997.