



SHSDA: secure hybrid structure data aggregation method in wireless sensor networks

Maryam Naghibi¹ · Hamid Barati¹

Received: 23 April 2020 / Accepted: 24 November 2020 / Published online: 2 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Wireless sensor networks (WSNs) are composed of several nodes, distributed in a geographical region. Limited energy of nodes is the main challenge of WSNs. Hence, it is required to apply different methods to consume less energy for calculations and communications. One method to reduce energy consumption in WSNs is to reduce the number of packets transmitted in the network. Data aggregation technique can cause a decrease in the number of transmitted packets. In fact, the technique combines related data and prevents sending additional packets. In this paper, a secure data aggregation method based on a combination of star and tree structures is suggested. Here, the network is geographically divided into four equal parts, and a stable star structure is formed in each part. In the secure hybrid structure data aggregation (SHSDA) method, each node is assigned a parent for transmitting data. To improve the security of data, the lightweight symmetric encryption is applied, and a key is distributed between each parent node and its children. The encrypted data is sent from leaf nodes to parent nodes, and gradually reaches the root through a star structure. Then the data is transmitted to the base station using the tree structure. The proposed method has been simulated using NS2. The results reveal that the average energy consumption and data delivery delay of SHSDA are less compared with that of conventional methods. Also, SHSDA method causes a rise in packet delivery rate, throughput, and flexibility.

Keywords Wireless sensor network · Data aggregation · Star structure · Encryption

1 Introduction

WSNs are established by the cooperation of sensors through sensing and processing data as well as wireless communication among sensor nodes. The networks have been formed to sense event-driven information and transmit it to the base station for in-depth assessment (Raval et al. 2017; Yousefpour and Barati 2020; Barati et al. 2015). WSNs have had a positive effect on various applications such as environmental monitoring, surveillance missions, health monitoring, home automation, target tracking, traffic monitoring, fire management, agriculture monitoring, industrial failure detection, and energy management (Mittal et al. 2019; Kocakulak and Butun 2017; Rawat and Chauhan 2020). WSNs are often

arranged as a large number of nodes in remote regions, which are insecure and inaccessible for human beings (Vinodha and Anita 2019). Thus it is needed to ensure that the nodes are scattered in a non-uniform manner, and their energy consumption is limited. Also, the nodes should be able to cooperate in WSNs so that they maintain the self-organizing nature of the network (Devi et al. 2020). Hence, it is important to produce an independent and efficient network among sensor nodes to guarantee prolonged network lifetime and controlled energy depletion (Mittal et al. 2019; Dezfouli and Barati 2020). Energy consumption and security are the most important challenges of wireless networks. Hence, a lot of work has been done on energy consumption (Yarinezhad and Hashemi 2019a, b) and security (Jamali and Fotohi 2017; Fotohi 2020; Jamali and Fotohi 2016; Fotohi et al. 2020) in these networks. Security in the network is of specific problems due to man lives are permanently at the condition as in traditional networks the major security concerns include confidentiality, integrity, and availability none of which involves primarily with life security (Fotohi et al. 2016). The data aggregation process minimizes the

✉ Hamid Barati
hbarati@iaud.ac.ir
Maryam Naghibi
m.naghibi@iaud.ac.ir

¹ Department of Computer Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran

number of data packets transmitted through the in-network data processing (Kaur and Munjal 2020; Song et al. 2020). The process is a data transmission mechanism, taking in transmitted data from different nodes. Then redundant packets are identified and are removed to form a single packet. The process leads to a decrease in the number of transmitted packets, which improves energy consumption (Devi et al. 2020; Dehkordi et al. 2020; Sarangi and Bhattacharya 2019). In the literature, there are various methods for deciding communication topologies such as cluster- (Mosavifard and Barati 2020), chain- (Wu et al. 2019), tree- (Ray and De 2017), and tree-cluster-based (Osamy et al. 2018) methods for gathering and transmitting data. The tree-based methods, applying the least number of connections to gather and transmit data, have been viewed as effective. Also, the topology is useful for several applications that include in-network data gathering such as environmental surveillance (Osamy et al. 2019). In a tree-based data aggregating method, all nodes are arranged as a tree. The intermediate node gathers, aggregates, and sends the data sensed by sensor nodes to the base station. The method leads to a decrease in energy consumption and an improvement network lifetime. Thus the technique requires less energy compared with cluster- or grid-based aggregation (John and Jyotsna 2018). Figure 1 illustrates the tree-based data aggregation.

In this paper, a secure hybrid data aggregation method based on a combination of star and tree structures is suggested, called SHSDA. Here, the network is geographically divided into four equal portions, called sections. In each section, the best node in terms of residual energy and centrality is selected as the root of a star structure, and a star structure is established among the nodes. Next, a tree structure is formed between the roots of these star structures. The data gathering process starts from leaves. The sensed data is encrypted by a key shared between the node and its parent and then is sent to the parent. Receiving the data from its

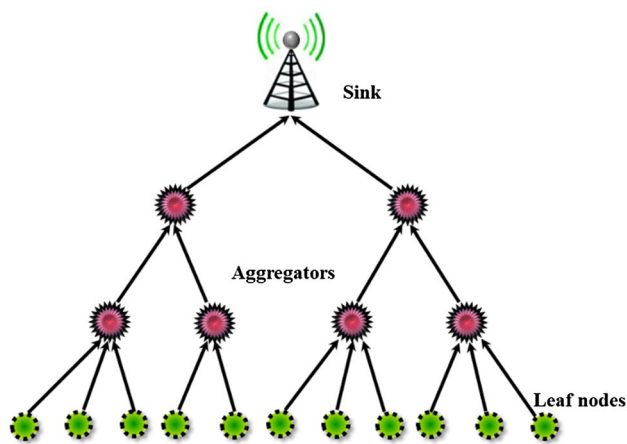


Fig. 1 Tree based data aggregation in wireless sensor networks

children, the parent node decrypts and aggregates it with its own data. Now, the key shared between this node and its parent node is used to encrypt the data. Then the data is sent to the parent. The process continues until the encrypted data is delivered to the root of the star structure. Finally, the root transmits the aggregated data to the base station through the tree structure. There is a shared key between the roots and the base station for encryption as well.

The innovations of SHSDA method are as follows:

- Selecting the parent node in the star structure of each section is based on suitable parameters. Thus a stable structure is formed in each section to transmit data.
- A fault-tolerant star structure is formed in each section so that a grandparent node replaces a dead parent node, and it is not needed to rebuild the star structure.
- On forming star structures, keys are distributed between children and parents. The keys are established based on the exclusive features of each node. In the data gathering phase, encryption is performed hop by hop using these shared keys. This results in a more secure data transmission.
- Due to applying a star structure in each section and a tree structure among the roots of these sections, it is not required to employ route discovery for gathering data.

The rest of the paper is organized as follows. The related work is represented in Sect. 2. Section 3 details the performance of SHSDA. The simulation and analysis of the proposed method is provided in Sect. 4. Section 5 concludes the paper.

2 Related work

In the last decade, several algorithms and protocols have been developed for gathering data in WSNs. Among them, the collection tree protocol (CTP) is widely used for gathering data in WSNs (Sardar et al. 2017). The advantages of the protocol (CTP) are its easier support in recognizing proper nodes and identifying the priority in terms of different parameters over which a protocol is assessed (Sharma et al. 2018). In the following, we discuss some data aggregation algorithms, used in WSNs.

Yarinezhad and Hashemi (2019a) suggested a cluster-based routing algorithm (RFPT) for wireless sensor networks. RFPT includes of two parts, clustering and routing. This routing algorithm decreases and balances the energy consumption in the network by finding suitable paths between each cluster head and the sink.

Farzinavash et al. (2019) suggested EEDMS to reduce the delay of emergency data gathering and energy consumption in WSNs. The algorithm adopts two different methods to

collect emergency and normal data. Emergency information is given priority and is gathered using a spanning tree. However, a mobile sink is employed to gather normal data. In fact, a grid is constructed over the field, and the mobile sink travels over the network to collect the sensed data from cluster heads (CHs).

Gharaei et al. (2019) proposed a scheme for balancing the energy consumption of CHs and cluster members. The scheme applies two mobile sinks, moving on a spiral path among clusters in parallel. They stop in each cluster for a limited time. Reaching each cluster, the mobile sinks move as rapidly as possible among different stop locations to balance the energy consumption of cluster member nodes.

Tabatabaei and Rigi (2019) provided a method, using a sink mobile, to solve the hotspot problem, balance the load, and uniform energy consumption. The method is a reliable routing algorithm based on clustering and mobile sink in WSNs and reduces the delay of reporting. Also, deciding locally and dynamically about the best alternative for faulty CHs causes a rise in the reliability of the network.

Habib et al. (2018) represented a starfish routing algorithm. In this method, a routing backbone is used throughout the network to store the location of the mobile sink. Sensor nodes can access backbone nodes directly to transmit their data to the sink. Data can be delivered to the sink more rapidly due to applying a starfish structure; however, heterogeneous sensors may lead to the inefficiency of the path between sensor nodes and the sink.

Hawbani et al. (2018) suggested STDD protocol to reduce packet delivery delay. The protocol includes three phases: constructing, updating, and maintaining a tree. STDD creates just one main dissemination tree for each mobile sink. On changing the access node, the main tree is updated and maintained so that the shortest path to mobile sinks is guaranteed.

Naghbi and Barati (2020) presented EGRPM to decrease energy consumption and to improve network lifetime in WSNs. In EGRPM, the network is geographically cellularized, and two mobile sinks with a defined movement pattern are used to gather the data of sensor nodes. While the sinks are anchoring, the cells which are adjacent to the anchor point send data directly to the sink, but further cells send it in a multi-hop manner.

Ramesh and Yaashuwanth (2019) proposed TMS approach. The objective of the method is to secure data transmission so authors suggested a lightweight trust management system. The trust-based mechanism prevents attacks and stops malicious nodes from participating in data transmission permanently.

Kumar and Sivagami (2020) offered FSAMR approach to detect malicious nodes. In the method, the fuzzy system is applied to calculate the trust score of each sensor node in the network. Then malicious nodes are detected based on their

trust score. On establishing multi paths between the source and destination, sensed data is sent in a secure manner using the elliptic curve cryptography.

Baburaj (2017) introduced PMMTK approach for secure key distribution in WSNs. In PMMTK, a shared triple-key is used for providing secure communication among cluster member nodes, the CH, and the base station. Also, the nodes are authenticated using the triple-key. Then the encrypted data is gathered and is sent to the base station by the CH.

Hamouid et al. (2020) represented a lightweight and secure tree-based routing protocol (LSTR) for WSNs. In LSTR, the security of data routing from sensor nodes to the base station is provided based on an ID-based authenticated key-agreement protocol. In fact, LSTR provides a compromise between security and resource efficiency.

Uvarajan and Gowri (2020) suggested TAG-GCDA to reduce communication overhead. They proposed a three-fold homogeneous method that allows supervising secure neighbor selection, energy-efficient routing and data aggregation through the greedy approach. The purpose of these processes is to decrease the energy consumption of sensors and improve the network lifetime through less communication and control overhead.

Fotuhi et al. (2020) suggested ASDA-RSA method to protect the network against DoS attacks. The method applies an energy- and distance-based clustering to select appropriate CHs. Also, RSA cryptography algorithm and interlock protocol along with an authentication method were utilized to avoid DoS attacks.

Darabkh et al. (2019) provided a balanced power-aware clustering and routing protocol (BPA-CRP). In this algorithm, a transferring technique, based on a forwarder node, has been introduced to reduce overhead. The role of a forwarder is to collect data from the CH and further forwarders, and then sending it to the next layer or the base station. The protocol applies a policy which handles the death of nodes effectively. In other words, the method is capable to prevent any network disruption or data loss.

Selvi et al. (2019) proposed EATSRA to provide an optimal secure routing in WSNs. In this algorithm, the trust score is employed to detect malicious nodes. The algorithm includes two phases. In the first phase, the secure route is identified based on the trust score. In the second phase, the decision tree algorithm is used to detect the most secure route.

Yarinezhad and Hashemi (2020) proposed an energy-balanced and energy-efficient cluster based routing protocol (EB-CRP) for wireless sensor networks. EB-CRP divides the network operations into two stages, set-up and steady-state. These stages are executed in sequence until the network lifetime ends. The set-up stage itself includes of three steps, bootstrapping, clustering, and routing.

Zhang et al. (2020) proposed EDAGD approach to prolong the wireless sensor network lifetime and to decrease the energy consumption in the data transmission process. The proposed EDAGD includes three algorithms Multihop tree based data aggregation algorithm (MTDA); Entropy-driven aggregation tree-based routing algorithm (ETA) and Gradient deployment algorithm (GDA).

Bongale et al. (2020) suggested an intra-cluster data aggregation method (ICA) with special emphasis on multi hop aggregation for wireless sensor networks. The proposed method operates in rounds and each round includes of three phases namely (1) cluster formation stage, (2) intra cluster aggregation path construction and (3) data transmission phase.

A summary and comparison of the mentioned methods presents in Table 1.

3 Proposed method

Regarding the limitation of energy, energy-efficient data aggregation is viewed as a major challenge of WSNs. Since processing, transmitting, and receiving sensed data consume the most energy, data aggregation is of great significance. Data aggregation plays an important role in gathering sensed data and eliminating redundant data, which results in reduced traffic, limited energy consumption, increased network lifetime, and improved performance. Here, we propose a highly reliable data aggregation method based on a combination of star and tree structures (SHSDA). The proposed method is composed of three phases: constructing star structures, establishing a tree structure, and gathering data. In the first phase, the network is divided into four sections, and a star structure is established among the nodes of each section. In the next phase, a tree structure is constructed among the roots of the star structures. In the final phase, the leaf nodes encrypt and send sensed data to their parents. Receiving data from their children, parents decrypt and aggregate it and their own data. Then the parent node applies the key sharing with its parent to encrypt the data. Now, the data is sent to the parent node. Thus the data is delivered to the root hop by hop. Finally, the root sends data to the base station through the tree structure, formed among roots. While establishing the star and tree structures, each node is aware of the information of its own grandparent, thereby it can replace its dead parent with its grandparent. This results in providing a fault-tolerant structure. The overall flowchart of proposed method is illustrated in Fig. 2

The notations have been illustrated in Table 2.

The details and phases of the proposed method are discussed in the following.

3.1 Network and Energy Consumption Models

SHSDA is consisted of m homogeneous sensor nodes with limited energy. The nodes are fixed and are distributed uniformly in an area of $n \times n$ size. All nodes are equipped with GPS and are aware of their own location in the network field. The data sensed in each section is delivered to the root of the star structure through a sequence of transmissions from children to parent nodes. In each section, there is no difference between the root and other nodes in terms of hardware, processing power, and energy supply; however, the roots are closer to the center of their sections and their residual energy is higher. The roots send data to the base station through a tree structure. The nodes are aware of the base station location, fixed outside of the network field. The energy of sensor nodes can be calculated through a similar model used in a study by Arora et al. (2019). In this model, the energy required to transmit and receive a k -bit message to a distance of d is calculated through Eqs. (1) and (2).

$$E_{Tx}(k, d) = E_{elec} \times k + \epsilon_{amp} \times k \times d^h \quad (1)$$

$$E_{Rx}(k, d) = E_{elec} \times k \quad (2)$$

Here, $h \in \{2, 4\}$ if $d < d_0$ then $h = 2$ else $h = 4$, where E_{Tx} and E_{Rx} indicate the energy required for transmitting and receiving, respectively; and E_{elec} and ϵ_{amp} are the energy required for running the transmitter or receiver circuitry and the amplification energy, respectively.

d_0 is derived from Eq. (3).

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mz}}} \quad (3)$$

where ϵ_{fs} and ϵ_{mz} represent the amplification energy for free space and multi-path, respectively.

Also, sensor nodes data aggregation their own data with the received data and send it to the next node by E_{zn} , calculated as:

$$E_{zn} = u \times E_{Rx} + u \times E_{DA} \times k + k \times E_{Tx} \quad (4)$$

Where u indicates the number of received packets, and E_{DA} is the energy used by each sensor node to aggregate data. Moreover, each route consumes $E(z)$, calculated as:

$$E(z) = \sum_{i=1}^z E_{i,n_z(i)}(k_i, d_i) \quad (5)$$

Where $n_z(i)$ indicates the next hop of node i on path z , and $E_{i,n_z(i)}(k, d)$ is the energy from node i to node $n_z(i)$.

3.2 First phase: constructing star structures

The objective of the phase is to create a regular structure to aggregate and gather data in each section. First, the network

Table 1 Summary and comparison of related works

Ref	Approach	Advantages	Disadvantages
Yarinezhad and Hashemi (2019a)	Decreases and balances the energy consumption in the network by finding suitable paths between each cluster head and the sink	Balancing energy consumption, load balancing	The running time of the proposed algorithm is high for WSNs with more than 3000 sensor nodes and 100 gateways
Farzinvasb et al. (2019)	Collecting emergency data through the spanning tree and collecting normal data using a mobile sink	Reducing delivery delay, balancing energy consumption	long delay in sending normal data
Gharaei et al. (2019)	Collecting data through two mobile sinks with a parallel movement along a spiral path among the clusters	Balancing energy consumption, increasing network lifetime	High cost due to utilizing multiple MSs
Tabatabaei and Rigi (2019)	Applying clustering, a mobile sink and the routing tree for data collection	Reducing delivery delay, increasing reliability	High overhead to update the sink table and MS movement, selecting the cluster heads near the sink and improper distribution of them in the network
Habib et al. (2018)	Sending data to the mobile sink through a routing backbone, consisting of multiple ring- and radial-channels	Reducing delivery delay, fair distribution of data traffic	Inefficient route in case of heterogeneous sensor
Hawbani et al. (2018)	Transferring data to the mobile sink through the shortest path by updating and maintaining the tree	Reducing delivery delay	Not supporting mobile nodes, the high cost of the construction phase
Naghbi and Barati (2020)	Gathering the sensed data in the cells of the network using two mobile sinks with a specific mobility pattern	Reducing delivery delay, increasing packet delivery rate	Not supporting mobile nodes
Ramesh and Yaashuwanth (2019)	Using a lightweight trust management system to allow the secure transmission of data and avoid various attacks	Increasing packet delivery rate	Low resilience and dependability
Kumar and Sivagami (2020)	Providing a secure data transmission on multiple paths between source and destination using the elliptic curve cryptography	Reducing delivery delay	High computational complexity
Baburaj (2017)	Using a shared triple key to authenticate the nodes and allow a secure communication among the cluster member nodes, cluster head and the base station	Increasing throughput, reducing delivery delay	Key connection threat by node capture attacks
Hamouid et al. (2020)	Securing data routing from sensor nodes to the base station based on an ID-based authenticated key-agreement protocol	Reducing communication overhead and energy loss	High computational complexity, not supporting mobile nodes
Uvarajan and Gowri (2020)	Aggregating data in a trustworthy manner using a threefold homogeneous method	Reducing communication overhead	Increasing computational overhead
Fotohi et al. (2020)	Using RSA cryptography algorithm and interlock protocol along with an authentication method to avoid DoS attacks	Increasing throughput and security	Not supporting mobile nodes
Darabkh et al. (2019)	Using a transferring technique, based on a forwarder node, to reduce communication overhead	Load balancing	Unbalanced distribution of nodes load among CHs

Table 1 (continued)

Ref	Approach	Advantages	Disadvantages
Selvi et al. (2019)	Identifying malicious nodes using the trust score, selecting the most secure route from selected routes through the decision tree algorithm	Increasing security, reducing delay	low throughput-Increasing computational overheads
Yarinezhad and Hashemi (2020)	Providing an energy-aware routing algorithm to find the optimal routing tree to balance the load of the sensor nodes	Load balancing, increasing network lifetime	High energy consumption in high round, Running time complexity of the fpt-approximation algorithm
Zhang et al. (2020)	Using three independent algorithms called MTDA, ETA, and GDA to overcome the drawback of random deployment and energy consume dissipation	Increasing network lifetime	Does not support mobile nodes
Bongale et al. (2020)	Providing a new intra cluster based data aggregation strategy using relay nodes	The load among the CHs is more balanced, reduces the energy consumption of the network	Low performance when there is no opportunity for intra cluster data aggregation

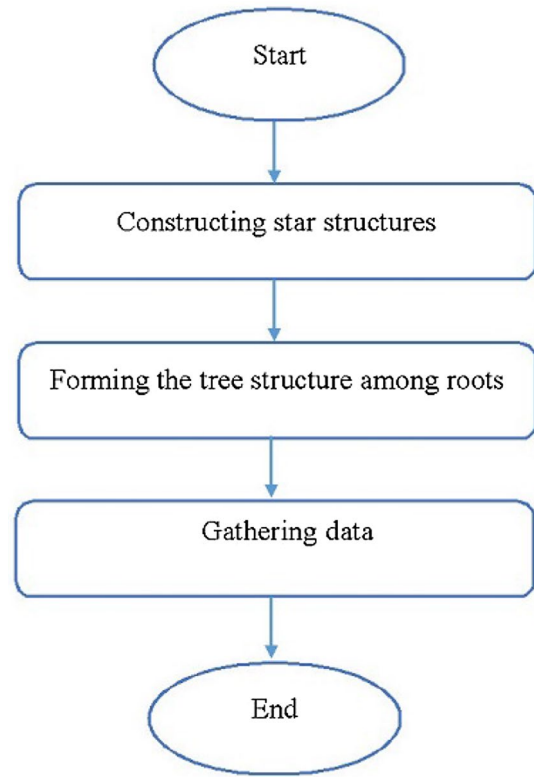


Fig. 2 The overall flowchart of SHSDA

Table 2 Notations

Notation	Description
m	Number of sensor nodes
n	Network length
q_1, q_2, q_3, q_4	Network corners
(x_c, y_c)	Coordinates of the network center
(x_{c_s}, y_{c_s})	Section center
S	Section's number
(x_i, y_i)	Location of the node i
p	Parent node
f	Farthest candidate parent node
L_{Root}	Level of root
L_{Child}	Level of children
R_1	Root of section 1
R_2	Root of section 2
R_3	Root of section 3
R_4	Root of section 4

is divided into four equal sections, and a star structure is established in each section. In each section, the best node in terms of centrality and residual energy is selected as the root of a star structure by the base station. Then the process of forming star structures initiates. The process of dividing the network and forming star structures are discussed in step 1–6.

Step 1: Dividing the network

As shown in Fig. 3, the network is divided into four equal parts. The network is of $n \times n$ size, and the coordinates of the four corners have been depicted in the figure. The coordinates of the network center are derived from Eq. (6).

$$(x_c, y_c) = (a + \frac{n}{2}, b + \frac{n}{2}) \tag{6}$$

where (x_c, y_c) is the network center.

Step 2: Identifying sections number

Based on (x_c, y_c) , nodes derive their section number through Eq. (7).

$$S = \begin{cases} 1, & \text{if } (x_i > x_c \text{ and } y_i > y_c) \\ 2, & \text{if } (x_i < x_c \text{ and } y_i > y_c) \\ 3, & \text{if } (x_i < x_c \text{ and } y_i < y_c) \\ 4, & \text{if } (x_i > x_c \text{ and } y_i < y_c) \end{cases} \tag{7}$$

Where (x_i, y_i) and (x_c, y_c) indicate the node coordinates and the network center, respectively.

Step 3: Transmitting information to the base station

Nodes send a hello packet, containing their ID, residual energy, location, and section number, to the base station. Hello packet has been illustrated in Fig. 4.

Step 4: Selecting the root of sections

The base station divides nodes into four sections based on the sections numbers, which are in hello packets. As shown in Fig. 5, the center of each section is derived from Eq. (8).

$$(x_{c_s}, y_{c_s}) = \begin{cases} (a + \frac{3n}{4}, b + \frac{n}{4}), & \text{if } S = 1 \\ (a + \frac{n}{4}, b + \frac{n}{4}), & \text{if } S = 2 \\ (a + \frac{n}{4}, b + \frac{3n}{4}), & \text{if } S = 3 \\ (a + \frac{3n}{4}, b + \frac{3n}{4}), & \text{if } S = 4 \end{cases} \tag{8}$$

where (x_{c_s}, y_{c_s}) and n denote the coordinates of section center and the network length, respectively.

Then the base station calculates nodes fitness based on the residual energy and centrality through Eq. (9). Now, the base station selects the most proper node in each section as the root and informs it. The information of each root is sent

Node ID	Remaining energy	Location	S
---------	------------------	----------	---

Fig. 4 Hello packet

Fig. 3 Dividing the network

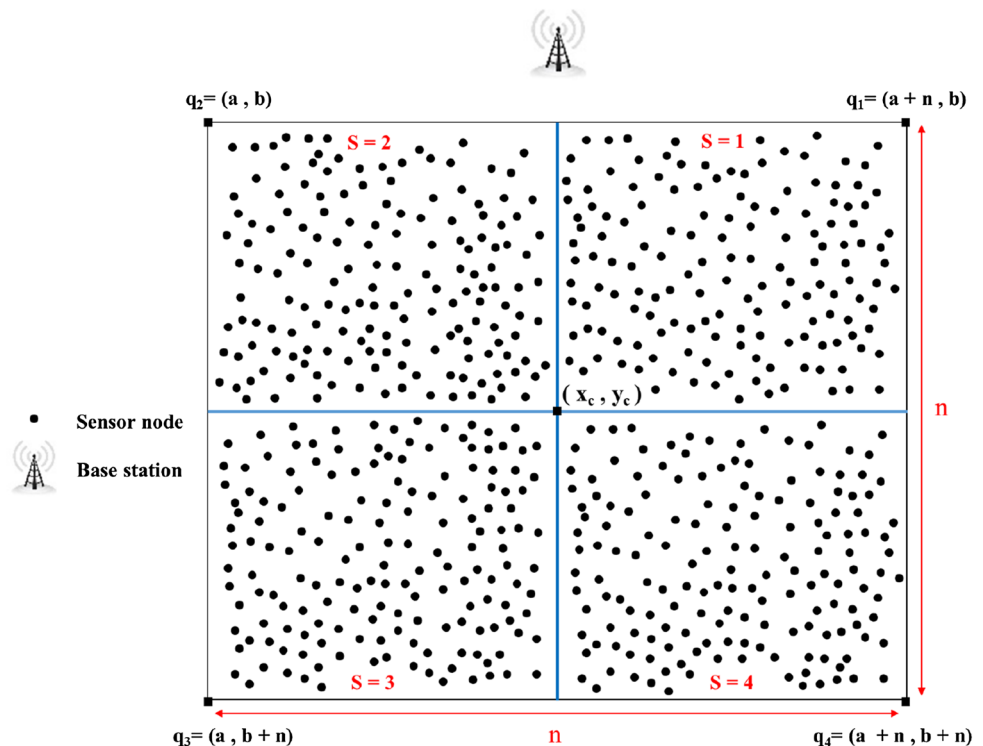
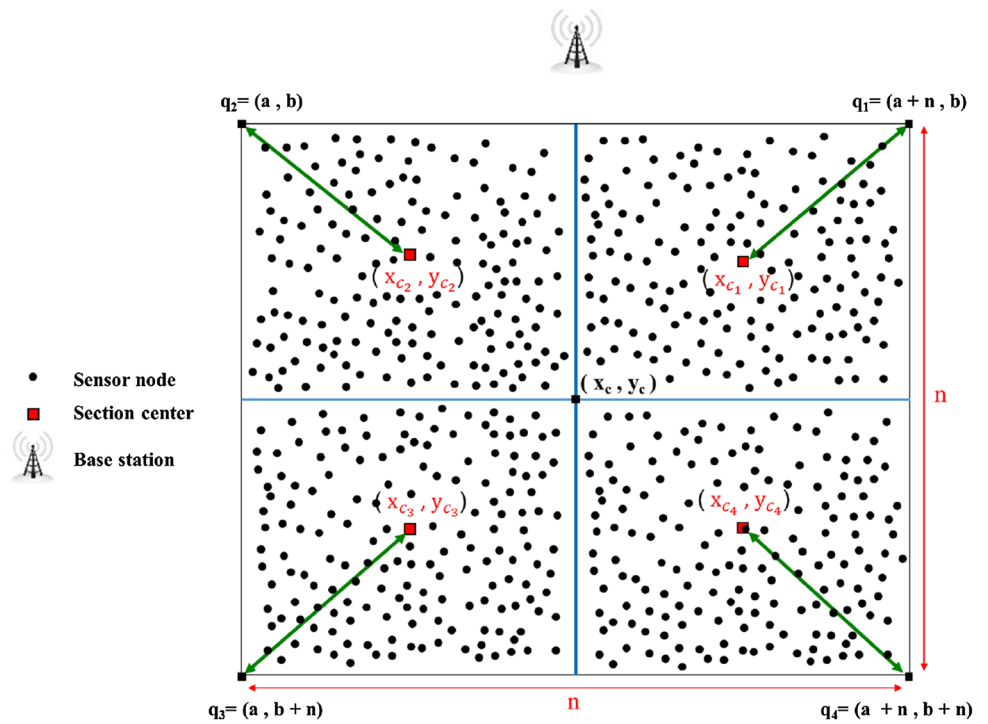


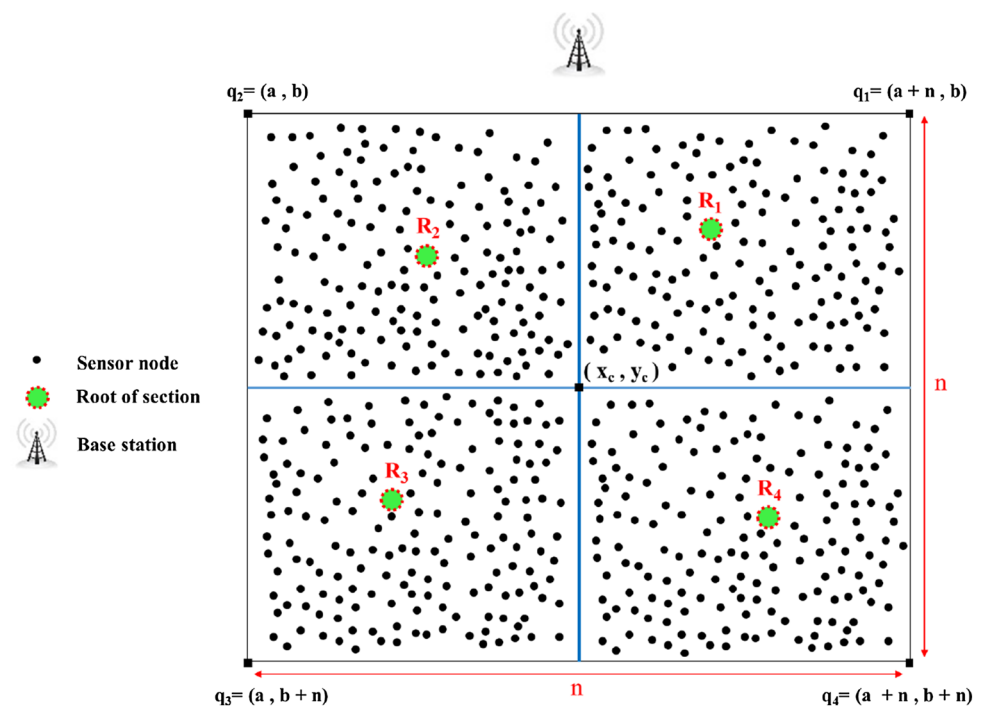
Fig. 5 Determining the center of sections



to other roots by the base station. The information is used for constructing the tree structure among the roots. Figure 6 depicts the network after selecting roots.

$$Fit_{node_i} = \frac{Energy_{current_i}}{Energy_{max_i}} + \left(1 - \frac{\sqrt{(x_i - x_{c_s})^2 + (y_i - y_{c_s})^2}}{\sqrt{(x_{c_s} - x_{q_s})^2 + (y_{c_s} - y_{q_s})^2}}\right) \tag{9}$$

Fig. 6 Selecting the root of sections



where (x_i, y_i) is the location of node i , (x_{q_s}, y_{q_s}) is the coordinate of section corner, (x_{c_s}, y_{c_s}) is the coordinate of section center, S is the section number (derived from Eq. (7)), $Energy_{current}$ is the residual energy, and $Energy_{max}$ is the maximum energy of the node when the battery is fully charged.

Step 5: Distributing shared keys among the roots and the base station

The base station generates four keys and distributes them among the roots of the sections. $Key_1, Key_2, Key_3,$ and Key_4 are assigned to the root of section 1 (R_1), section 2 (R_2), section 3 (R_3), and section 4 (R_4), respectively.

Step 6: constructing star structures

Constructing star structures initiates simultaneously in four sections. In each section, the root ($L_{Root} = 0$) sends a root packet to its neighboring single-hops and introduces itself as their parent.

As shown in Fig. 7, the packet contains the ID, the section number, the ID of the root parent (according to Eq. (10)), the parent key (shared between the root and its parent), and children key (shared between the root and children). These keys are used to improve the security of data aggregation. Section 3.2.1 represents the calculation and distribution of keys among nodes in details.

$$Parent = \begin{cases} BS, & \text{if } S = 1 \\ BS, & \text{if } S = 2 \\ R_2, & \text{if } S = 3 \\ R_1, & \text{if } S = 4 \end{cases} \quad (10)$$

where S is the section number, R_1 and R_2 indicate root of section 1 and root of section 2, respectively.

The nodes receiving the root packet join the root and create level 1 ($L_{child} = 1$). The construction of star structures starts from the center of sections. The nodes of level 1 send an advertisement packet to their single-hop neighbors to advertise themselves as their candidate parents. As shown in Fig. 8, the nodes transmit the energy, the location, the key they share with level 2, the parent’s ID, and the key they share with their parent. Thus if a node fails, the children nodes ($L_{child} = 2$) join their grandparent ($L_{Root} = 0$) and restore the star structure. After sending data to the parent, if the nodes do not receive an ACK message, they detect that their parent is faulty and join their grandparent so that the reliability is improved.

Node ID	S	Parent ID	Parent key	Children key
---------	---	-----------	------------	--------------

Fig. 7 Root packet

Fig. 8 Advertisement packet

Node ID	Remaining energy	Location	S	Parent ID	Parent Key	Children key
---------	------------------	----------	---	-----------	------------	--------------

If the nodes receiving the advertisement packet do not join the star structure, they establish level 2 according to the following conditions.

- If they receive just one advertisement packet, they join the parent node.
- If they receive more than one advertisement packet, they select the fittest candidate parent node to join according to Eq. (11).

When the parents are selected, they share a key with their children. In fact, level 0 and 1 share a key, level 1 and 2 share a key and so on.

$$Fit_{node_p} = \frac{Energy_{current_p}}{Energy_{max_p}} + \left(1 - \frac{\sqrt{(x_i - x_p)^2 + (y_i - y_p)^2}}{\sqrt{(x_i - x_f)^2 + (y_i - y_f)^2}} \right) \quad (11)$$

where $(x_i, y_i), (x_p, y_p), (x_f, y_f), Energy_{current_p},$ and $Energy_{max_p}$ denote the location of the level 2 node, the location of the candidate parent node, the location of the farthest candidate parent node, the current energy of the candidate parent node, and the maximum energy of the candidate parent node when the battery is fully charged.

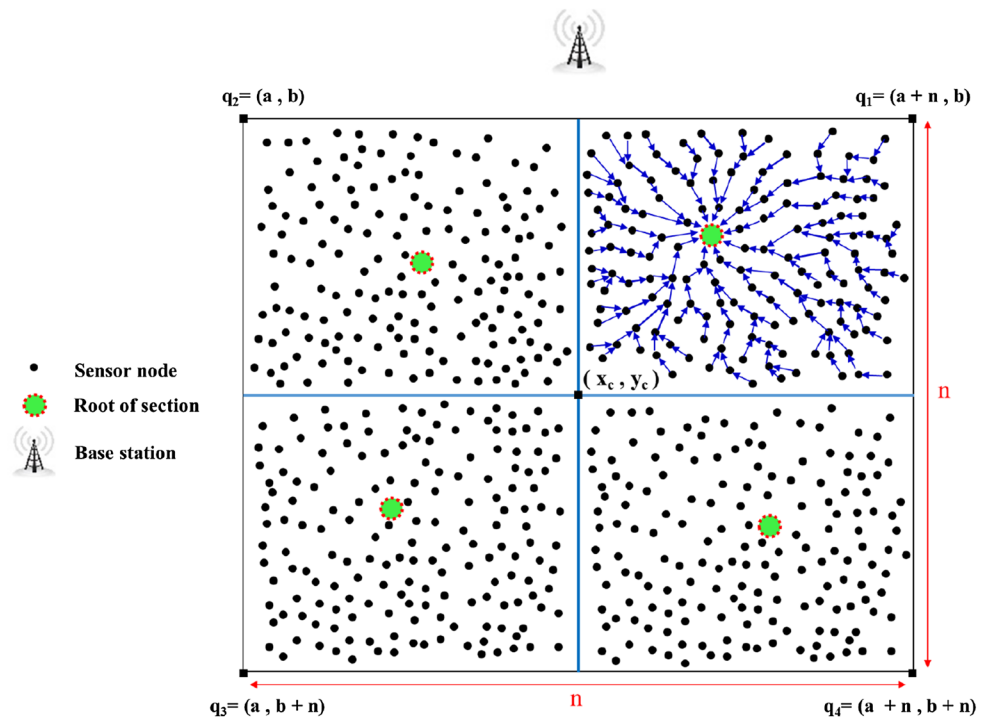
Other levels are established similarly and the process continues until all nodes of a section join the star structure. Figure 9 depicts the establishment of the star structure in section 1.

3.2.1 Calculating keys

Here, the sensor nodes apply rail fence encryption to encrypt their data. Since rail fence encryption involves shared keys, each parent node calculates and sends a key to its children to establish communication with them. The calculation is as follows.

1. The base station produces a random number between 2 and 6 and stores it in the memory of each node—the numbers smaller than 2 make encryption meaningless, and the numbers which are above 6 make the key length long.
2. The least significant byte of the parent node’s ID is divided by 5. Since the remainder is between 0 and 4, the key cannot be long.
3. The result of step 2 is added to the random number, and the sum is the key value.

Fig. 9 Establishment of the star structure in section 1



The flowchart of the first phase has been illustrated in Fig. 10.

The flowchart of constructing star structures in each section has been depicted in Fig. 11. Algorithm 1 shows the pseudo-code for first phase.

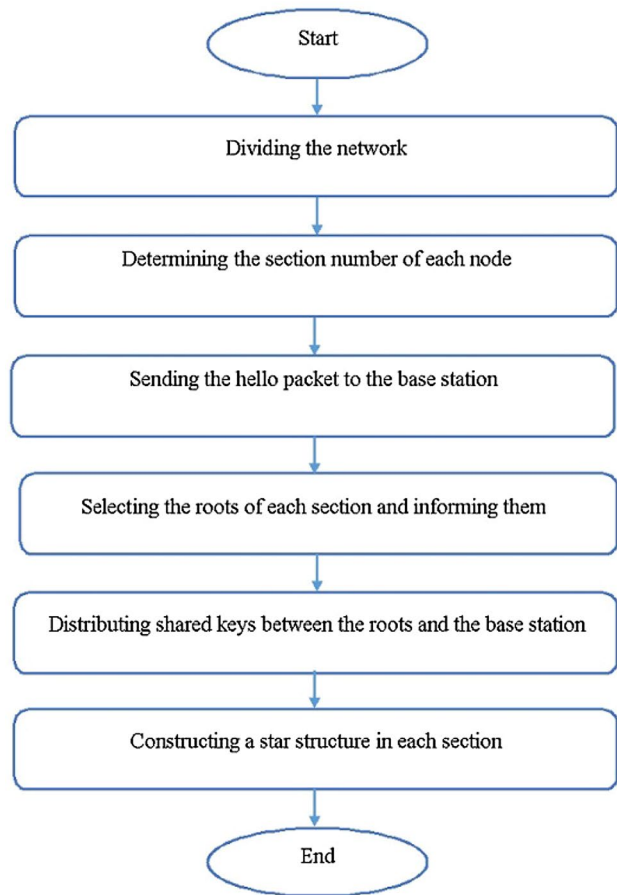


Fig. 10 Flowchart of the first phase

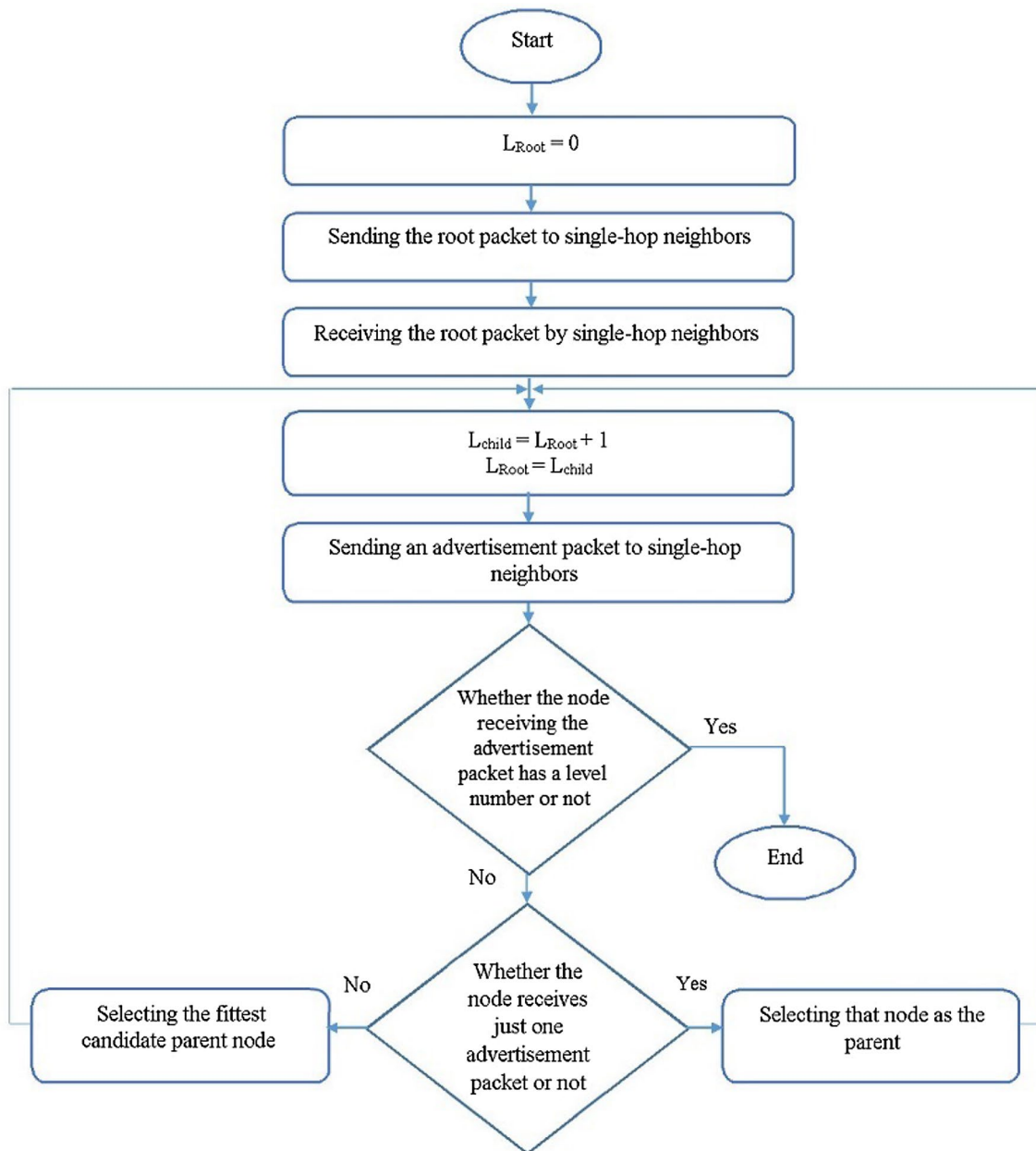


Fig. 11 Flowchart of forming star structures in each section

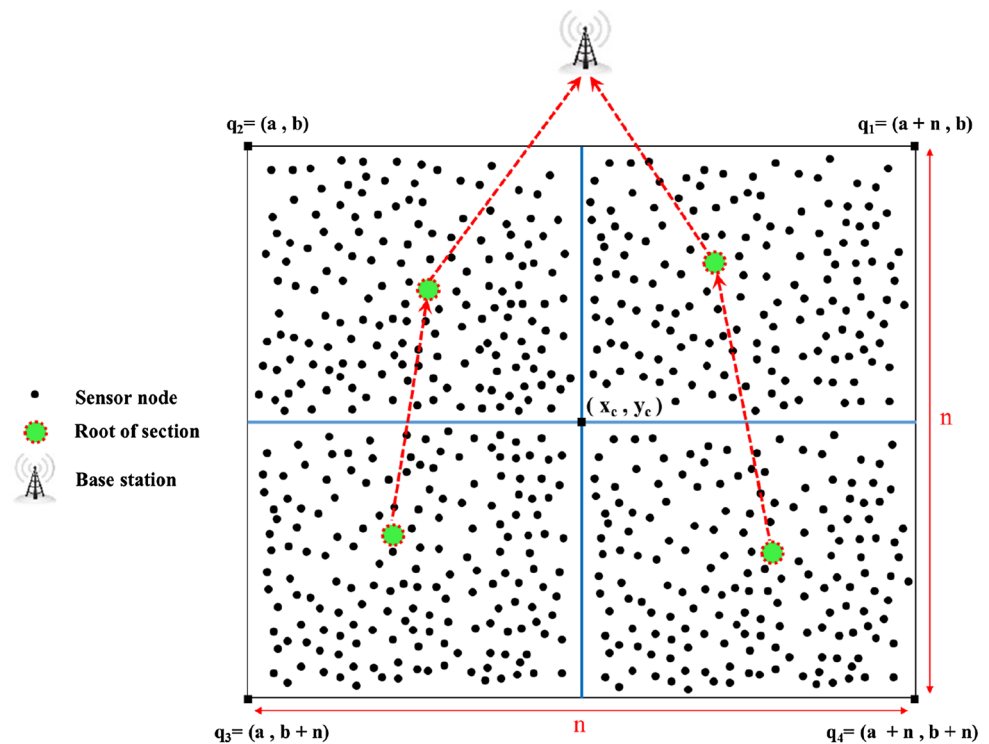
Algorithm 1 Pseudo code for first phase

```

1: Initialization: ( $n*n$ : network size,  $m$ : number of sensor nodes,  $S$ : section number)
2: Obtain the coordinates of the network center and save in  $(x_c, y_c)$ 
3: for ( $i=1$  ;  $i \leq m$  ;  $i++$ ) do
4:   Identifying sections number for  $node_i$ 
5:    $Node_i$  send a hello packet to the base station
6: end for
7: Selecting the root of sections by base station
8: The base station generates four keys
9: Distributes the keys among the roots of the sections
10: for ( $S=1$  ;  $S \leq 4$  ;  $S++$ ) do
11:    $L_{Root} = 0$ 
12:   The root of  $S$ 's section sends a root packet to its neighboring
13:   The nodes receiving the root packet join the root and create level 1
14:   while (All nodes of a section join the star structure) do
15:      $L_{child} = L_{Root} + 1$ 
16:      $L_{Root} = L_{child}$ 
17:     The nodes of level  $L_{child}$  send an advertisement packet
18:     Neighbor nodes received the advertisement packet
19:     if Receive just one advertisement packet then
20:       join the parent node
21:     else if (Receive more than one advertisement packet) then
22:       select the fittest candidate parent node
23:       join to selected parent node
24:     end if
25:   end while
26: end for

```

Fig. 12 Forming the tree structure among roots



Node ID	New parent ID	New parent key
---------	---------------	----------------

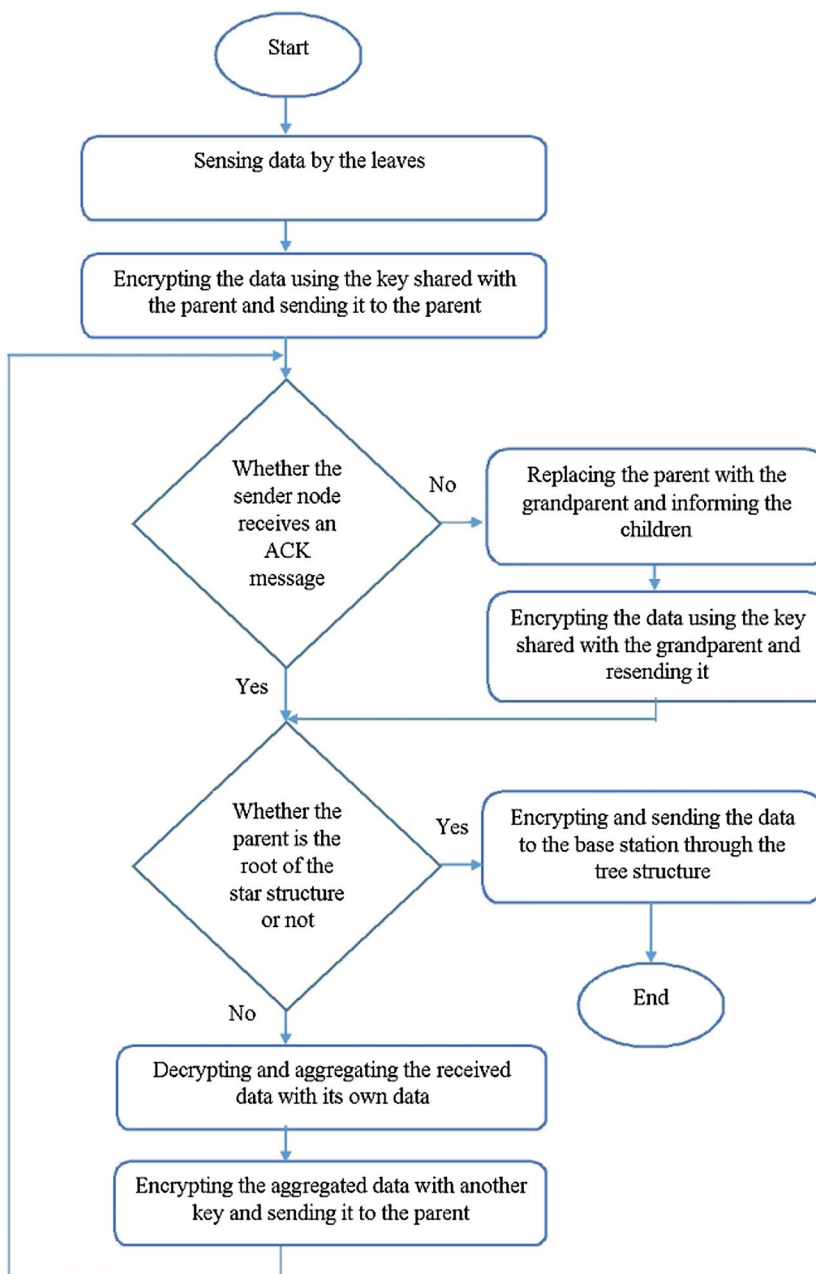
Fig. 13 Update parent packet

3.3 Second phase: forming the tree structure among roots

On forming star structures in sections, a tree structure is constructed to connect the roots to the base station. As shown in Fig. 12, the base station is the root of the tree. The

roots select their parents based on the information they have received from the base station in the fourth step of the first phase. In fact, R_1 and R_2 select the base station as their parent, and R_3 and R_4 choose R_2 and R_1 as their parent, respectively. Clearly, when the root of a section fails, the nodes in level 1 send their data to the root parent. Note that if R_1 or R_2 are faulty, R_4 and R_3 select the base station as their new parent. Then they send an advertisement packet (according to Fig. 13) to their children in level 1 to cause them to update their information on their grandparent.

Fig. 14 Flowchart of data gathering phase in SHSDA



3.4 Third phase: Gathering data

In this phase, the encrypted sensed data is delivered to the roots through a secure structure to be sent to the base station.

The procedure for gathering data is as follows.

Step 1: Sensing data

Leaves sense environmental data.

Step 2: Encrypting and sending data to parent

The data is encrypted in a rail fence manner using the key shared with the parent (See Sect. 3.2.1). Then the data is sent to the parent, and the node waits for an ACK message. If the parent does not send an ACK message, the node detects that its parent is faulty and replaces it with its grandparent. Thus the data is encrypted using the key shared with the grandparent and is resent. The node sends the information of its new parent to its children so that they update their information

on their grandparent. The advertisement packet for updating the new parent has been depicted in Fig. 13.

Step 3: Receiving data by parent

Receiving the data, the parent checks whether it is selected as a root node or not.

- If the node is the root, the data is decrypted and then encrypted using the key shared with its parent, and the data is sent to the base station through the tree structure among roots.
- If the node is not the root, it decrypts the data received from its children using the shared key and then aggregates it. Now, the node initiates step 2.

The flowchart of data gathering phase in SHSDA has been illustrated in Fig. 14. Algorithm 2 shows the pseudo-code for third phase.

Algorithm 2 Pseudo code for third phase

```

1: Leaves sense environmental data
2: Encrypting data using the key shared with the parent
3: Sending the data to the parent
4: L1: The sender node waits for an ACK message
5: if (Do not receive an ACK from parent) then
6:   Replacing the grandparent instead of the parent
7:   The node sends the new parent information to its children
8:   Children update their information
9:   Encrypting the data using new shared key
10:  Resending data to new parent
11: else
12:  if (The parent is the root of star structure) then
13:    Encrypting the data
14:    Sending the encrypted data to base station by tree structure
15:  else
16:    Decrypting the received data
17:    Aggregation with own's data
18:    Encrypting aggregated data and send to parent
19:    Goto L1
20:  end if
21: end if

```

Fig. 15 An example of data encryption in the proposed method

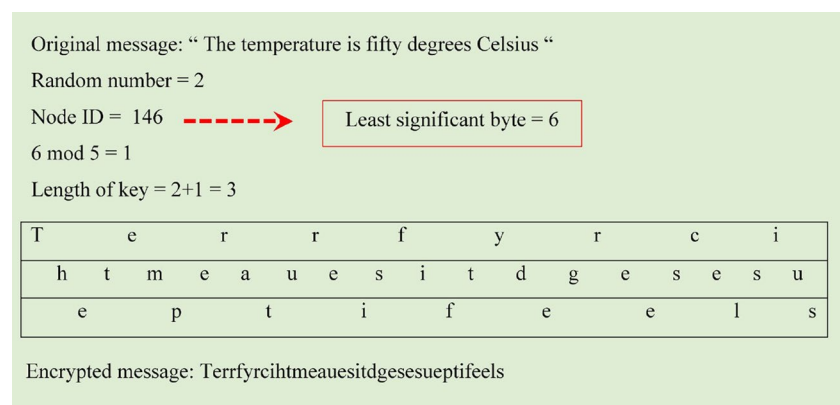


Table 3 Simulation parameters

Parameters	Value
Network area	1500 × 1500 m ²
Number of sensor nodes	500
Initial energy of sensor nodes	2 J
MAC protocol	IEEE-802.11
Nodes distribution model	Random
Data packet size	512 byte
Transmit power	0.660 W
Receiving power	0.395 W
Transmission range	250 m
Constant bit rate	500 kbps
Simulation time	600 s

3.5 Encryption

The main objective of encryption is to design a scheme or protocol to allow completely secure transmission through an insecure channel. In this step, the rail fence encryption,

which is a lightweight symmetric method, has been applied. In fact, the number of horizontal lines equals the number of encryption keys. This is calculated according to Sect. 3.2.1. First, the characters are written diagonally in a downward direction from left to right so that the defined rows are filled. On filling the last row, writing continues in an upward direction. The process continues until the message is completed. If there is an empty space, it is filled by an X, equaling Null. To encrypt the message, it is required to organize the characters in each row from the first to the last (Singh et al. 2019).

An example of data encryption in the proposed method is shown in Fig. 15.

4 Simulation

The section represents the results of the simulation and evaluation of the proposed method. We applied NS 2 tool to simulate the method. The results have been compared with TMS (Ramesh and Yaashuwanth 2019), FSAMR (Kumar

Fig. 16 Average energy consumption in proportion to different node numbers

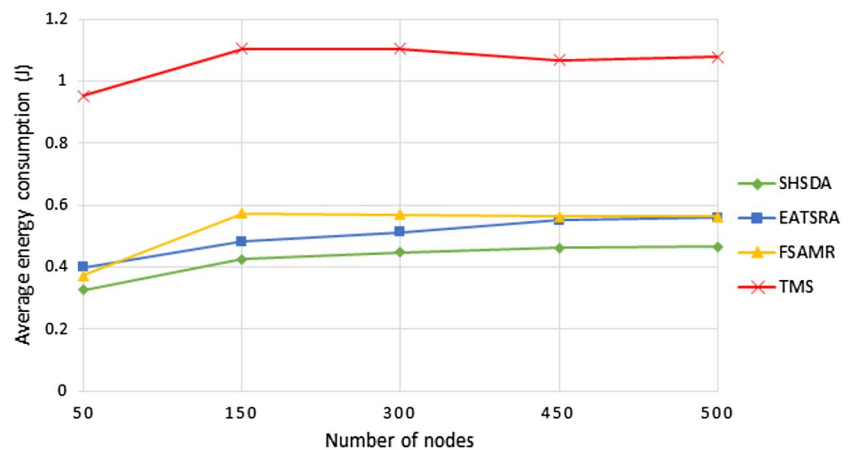


Fig. 17 Average energy consumption at different times

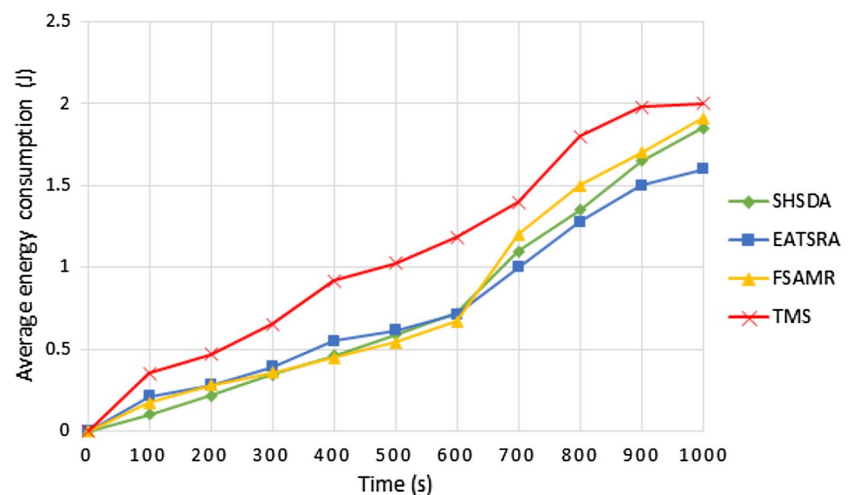
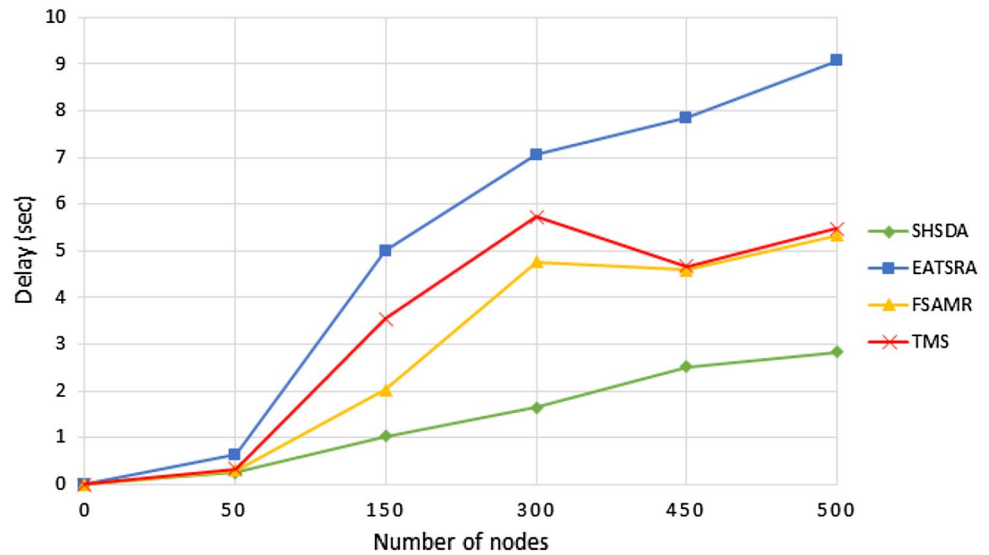
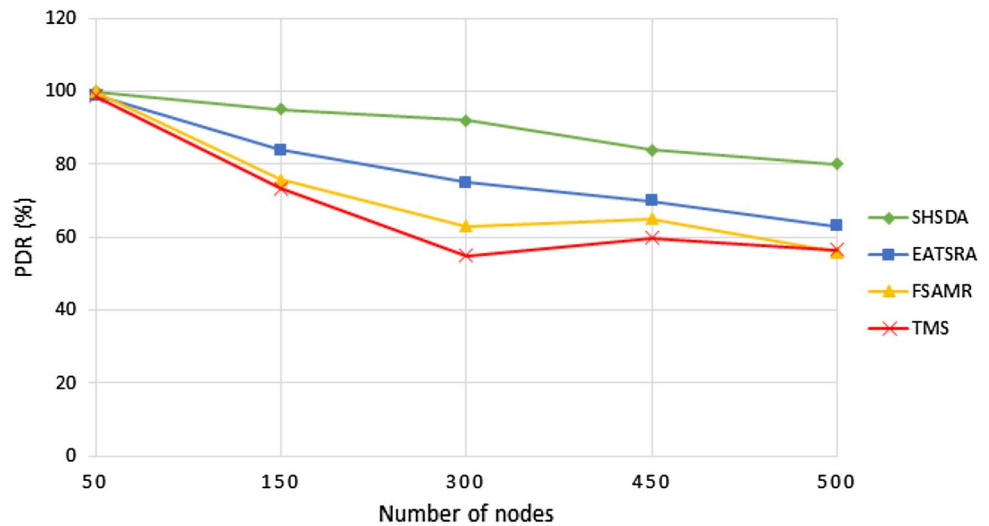


Fig. 18 Average end to end delay**Fig. 19** Packet delivery rate

and Sivagami 2020) and EATSRA (Selvi et al. 2019) protocols. Table 3 lists the simulation parameters.

Figure 16 illustrates the average energy consumption of the network nodes in relation to the number of nodes. The strategies used for the proposed method decreased the energy consumption of SHSDA compared with other methods. Here, the network is divided into some parts to transmit data to the base station. Next, a star structure is established in each section, and the node with the highest energy level and shortest distance from the center is selected as the root. Also, each node selects the closest node with the highest energy level as its parent. Thus a regular stable structure is formed for transmitting data in each section. In fact, the data is sent to the root step by step, which leads to consuming less energy. Then the data is delivered to the base station through a tree structure, which is formed among the roots. The structure is able to restore itself, which results in proper energy

distribution. In addition, employing star and tree structures eliminates the process of path discovery and improves the energy consumed for transmitting information. Therefore, the proposed method requires less energy compared with other methods.

In Fig. 17, the average energy consumption of the network nodes per unit time has been illustrated. In each section, a node is selected as the root, gathering the data of the section to be sent to the base station. The nodes are connected to the root in a star shape. When the root of the section dies, the nodes of level 1 are selected as the root of the section and send data to the base station. The process continues hierarchically with the nodes of other levels. Thus when the network starts its operation, the energy consumption is less compared with other methods; however, over time, the nodes die, and more nodes connect the base station in a single-hop manner. This results in a slight rise in the average energy

Fig. 20 Throughput

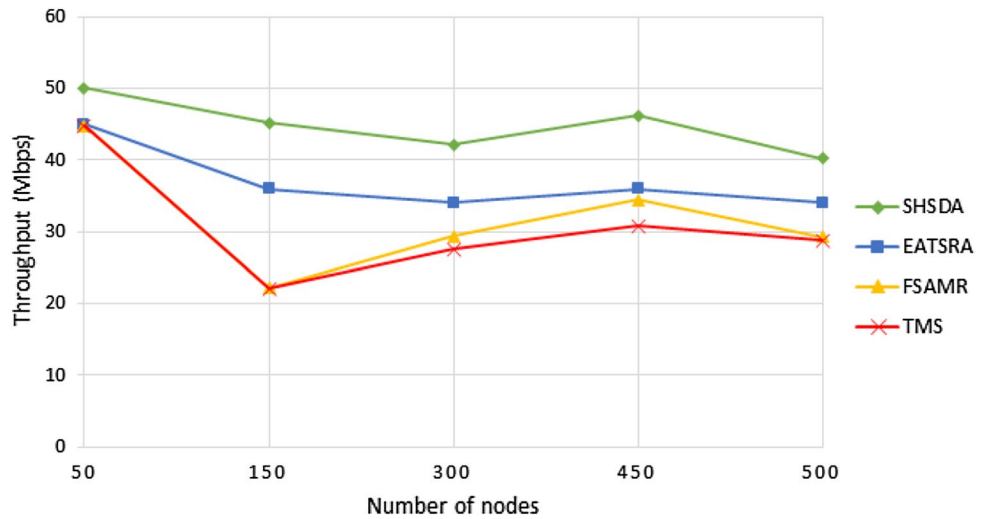


Fig. 21 Flexibility

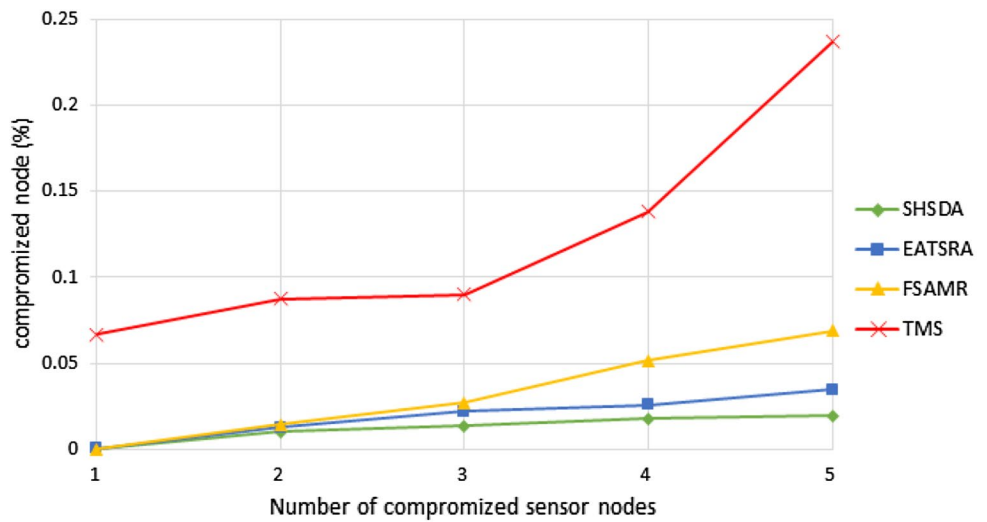
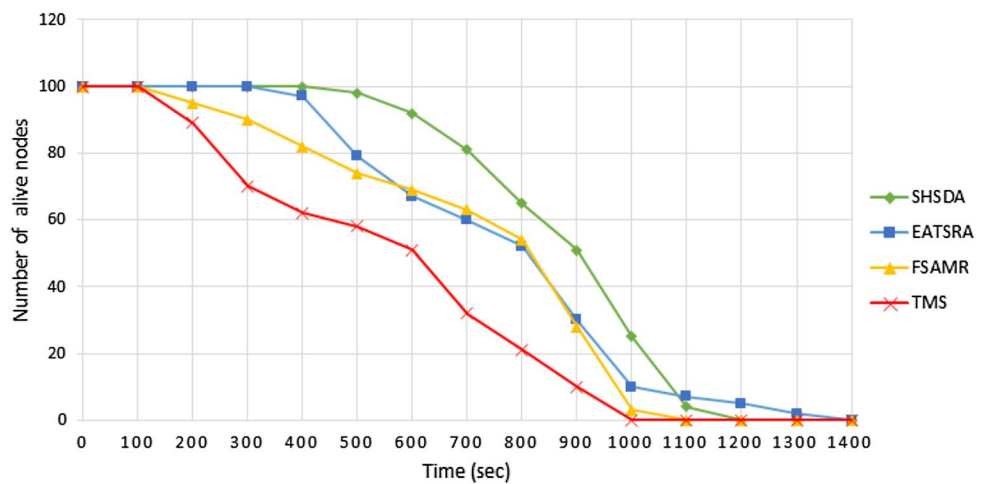


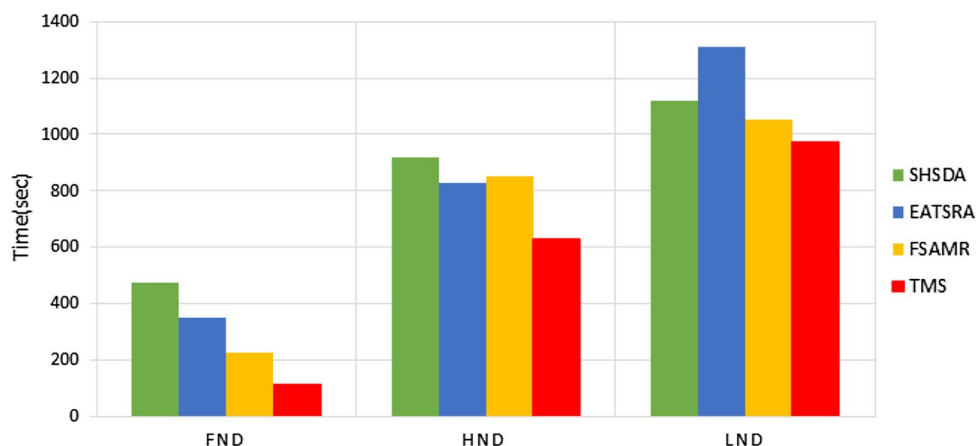
Fig. 22 Alive nodes



consumption of the network nodes, which is negligible since the connection between alive nodes and the base station is maintained.

End-to-end delay is the time required to deliver a packet from the source to the destination node. The end-to-end delay in the proposed method has been depicted in Fig. 18.

Fig. 23 Network lifetime



Here, a star structure is used in each section, and a tree structure is formed among the sections. Both star and tree structures are regular and fault-tolerant. In fact, when a parent node dies, the grandparent node replaces it so that it is not needed to rediscover paths. Thus data is sent to the base station through a regular fault-tolerant structure. Also, the method applies a lightweight symmetric encryption method, in which the encryption delay is so small. Generally, the delay of the proposed method is less compared with other methods.

Packet delivery rate is defined as the ratio of the received data packets to the transmitted data packets and is expressed in percentage. A comparison between packet delivery rates has been illustrated in Fig. 19. In the proposed method, a regular stable structure is used in each section as well as among the sections to transmit data to the base station. In fact, the network is divided into partitions, and the node with the highest score in terms of energy level and centrality is selected as the root of each section. The star and tree structures are fault-tolerant so that the grandparent node replaces the dead parent node to avoid extra cost and path rediscovery. The stable path created between the source and the base station improves packet delivery rate. Therefore, as shown in Fig. 19, the packet delivery rate in this method is superior to other methods.

Throughput is the amount of data transmitted from the source to the destination within a specific time. Various factors such as route stability, delay, and malicious nodes have effects on throughput. As shown in Fig. 20, the throughput of our proposed method is higher than that of other methods. Also, Fig. 16 illustrates that SHSDA has a regular stable structure and consumes less energy. Figures 18 and 19 show that our proposed method provides less delay but a higher packet delivery rate compared with other methods. In this way, the throughput of SHSDA is higher and can deliver more data compared with other methods.

Fig. 21 depicts the flexibility of the proposed method. Here, due to using a shared key, locally distributed between

each parent and its children, if the attacker discovers the key, it just has access to the information of two nodes. In fact, the data is encrypted and decrypted hop by hop so that discovering the encryption key does not have a significant effect on changing data. Therefore, the flexibility of SHSDA is improved in the presence of malicious nodes.

The comparison between the proposed method and other methods in terms of the number of alive nodes and the network lifetime has been depicted in Figs. 22 and 23. In this comparison, the number of nodes equals 100. Figure 22 shows the number of alive nodes while implementing the proposed method, and Fig. 23 illustrates the network lifetime in terms of the death of the first node, half of the nodes and all nodes. In the proposed method, the structure of the network has been organized as a star, and the packets are sent to the nodes of higher levels over short distances. Since the highest level of energy consumption is required for transmitting data, the structure causes nodes to consume less energy. Also, the collaborative structure of the network leads to balancing energy consumption. As shown in Figs. 22 and 23, the nodes lifetime is increased, which results in improved network lifetime. However, the roots of cells die and single-hop communication is increased overtime so that there is a steep rise in the number of dead nodes. The figures illustrate that in the proposed method the lifetime of the last node is shorter compared with EARSRA method.

5 Conclusion

One challenge of WSNs is the method of transmitting information to the base station and keeping it confidential. This paper suggests a secure data aggregation structure based on a combination of the star and tree structure. Here, the network is geographically divided into four equal parts. A regular and stable star structure is established in each part to send data. In fact, the star structure transmits data hop by hop, which results in a decrease in the energy consumption of

section members. Then the data is transmitted from the roots of each section to the base station through a tree structure, which is regular and fault-tolerant. Thus when the parent node dies, the grandparent node replaces it. Also, a shared key, used for encryption and decryption of data, is locally distributed between the parents and their children. This leads to having a flexible method in the presence of malicious nodes. The results of the simulation reveal that our proposed method (SHSDA) is superior to other methods in terms of the average energy consumption, data delivery delay, packet delivery rate, throughput, and flexibility. Further studies can enhance the proposed method through adjusting it with mobile WSNs, using mobile sinks with proper mobility models and employing a dynamic structure to manage the added or removed nodes.

References

- Arora VK, Sharma V, Sachdeva M (2019) ACO optimized self-organized tree-based energy balance algorithm for wireless sensor network. *J Ambient Intell Humaniz Comput* 10(12):4963–4975
- Baburaj E (2017) Polynomial and multivariate mapping-based triple-key approach for secure key distribution in wireless sensor networks. *Comput Electric Eng* 59:274–290
- Barati H, Movaghar A, Rahmani AM (2015) EACHP: energy aware clustering hierarchy protocol for large scale wireless sensor networks. *Wirel Pers Commun* 85(3):765–789
- Bongale AM, Nirmala CR, Bongale AM (2020) Energy efficient intra-cluster data aggregation technique for wireless sensor network. *Int J Inf Technol*. <https://doi.org/10.1007/s41870-020-00419-7>
- Darabkh KA, El-Yabroudi MZ, El-Mousa AH (2019) BPA-CRP: a balanced power-aware clustering and routing protocol for wireless sensor networks. *Ad Hoc Netw* 82:155–171
- Dehkordi SA, Farajzadeh K, Rezazadeh J, Farahbakhsh R, Sandrasegaran K, Dehkordi MA (2020) A survey on data aggregation techniques in IoT sensor networks. *Wirel Netw* 26(2):1243–1263
- Devi VS, Ravi T, Priya SB (2020) Cluster based data aggregation scheme for latency and packet loss reduction in WSN. *Comput Commun* 149:36–43
- Dezfooli NN, Barati H (2020) A distributed energy-efficient approach for hole repair in wireless sensor networks. *Wireless Netw* 26:1839–1855
- Farzinvas L, Najjar-Ghabel S, Javdzadeh T (2019) A distributed and energy-efficient approach for collecting emergency data in wireless sensor networks with mobile sinks. *AEU Int J Electron Commun* 108:79–86
- Fotohi R, Ebazadeh Y, Geshlag MS (2016) A new approach for improvement security against DoS attacks in vehicular ad-hoc network. *Int J Adv Comput Sci Appl* 7(7):10–16
- Fotohi R (2020) Securing of Unmanned Aerial Systems (UAS) against security threats using human immune system. *Reliab Eng Syst Saf* 193:106675
- Fotohi R, Nazemi E, Aliee FS (2020) An agent-based self-protective method to secure communication between UAVS in unmanned aerial vehicle networks. *Veh Commun* 26:100267. <https://doi.org/10.1016/j.vehcom.2020.100267>
- Fotohi R, Firoozi Bari S, Yusefi M (2020) Securing wireless sensor networks against denial-of-sleep attacks using RSA cryptography algorithm and interlock protocol. *Int J Commun Syst* 33(4):4234
- Gharaei N, Bakar KA, Hashim SZM, Pourasl AH (2019) Inter-and intra-cluster movement of mobile sink algorithms for cluster-based networks to enhance the network lifetime. *Ad Hoc Netw* 85:60–70
- Habib MA, Saha S, Razzaque MA, Mamun-or-Rashid M, Fortino G, Hassan MM (2018) Starfish routing for sensor networks with mobile sink. *J Netw Comput Appl* 123:11–22
- Hamouid K, Othmen S, Barkat A (2020) LSTR: lightweight and secure tree-based routing for wireless sensor networks. *Wirel Personal Commun* 112:1479–1501
- Hawbani A, Wang X, Kuhlani H, Karmoshi S, Ghoul R, Sharabi Y, Torbosh E (2018) Sink-oriented tree based data dissemination protocol for mobile sinks wireless sensor networks. *Wirel Netw* 24(7):2723–2734
- Jamali S, Fotohi R (2016) Defending against wormhole attack in MANET using an artificial immune system. *New. Rev Inf Netw* 21(2):79–100
- Jamali S, Fotohi R (2017) DAWA: defending against wormhole attack in MANETs by using fuzzy logic and artificial immune system. *J Supercomput* 73(12):5173–5196
- John N, Jyotsna A (2018) A survey on energy efficient tree-based data aggregation techniques in wireless sensor networks. In 2018 international conference on inventive research in computing applications (ICIRCA), IEEE, pp 461–465
- Kaur M, Munjal A (2020) Data aggregation algorithms for wireless sensor network: a review. *Ad Hoc Netw* 100:102083
- Kocakulak M, Butun I (2017) An overview of Wireless Sensor Networks towards internet of things. In 2017 IEEE 7th annual computing and communication workshop and conference (CCWC), IEEE, pp 1–6
- Kumar AR, Sivagami A (2020) Fuzzy based malicious node detection and security-aware multipath routing for wireless multimedia sensor network. *Multimed Tools Appl* 79:14031–14051
- Mittal N, Singh U, Salgotra R (2019) Tree-based threshold-sensitive energy-efficient routing approach for wireless sensor networks. *Wirel Pers Commun* 108(1):473–492
- Mosavifard A, Barati H (2020) An energy-aware clustering and two-level routing method in wireless sensor networks. *Computing* 102:1653–1671
- Naghbi M, Barati H (2020) EGRPM: energy efficient geographic routing protocol based on mobile sink in wireless sensor networks. *Sustain Comput Inform Syst* 25:100377
- Osamy W, Khedr AM, Aziza A, El-Sawya A (2018) Cluster-tree routing scheme for data gathering in periodic monitoring applications. *IEEE Access* 6:77372–77387
- Osamy W, El-sawya AA, Khedr AM (2019) SATC: a simulated annealing based tree construction and scheduling algorithm for minimizing aggregation time in wireless sensor networks. *Wirel Pers Commun* 108(2):921–938
- Ramesh S, Yaashuwanth C (2019) Enhanced approach using trust based decision making for secured wireless streaming video sensor networks. *Multimed Tools Appl* 1–20
- Raval G, Bhavsar M, Patel N (2017) Enhancing data delivery with density controlled clustering in wireless sensor networks. *Microsyst Technol* 23(3):613–631
- Rawat P, Chauhan S (2020) Probability based cluster routing protocol for wireless sensor network. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-020-02307-1>
- Ray A, De D (2017) Performance evaluation of tree based data aggregation for real time indoor environment monitoring using wireless sensor network. *Microsyst Technol* 23(9):4307–4318
- Sarangi K, Bhattacharya I (2019) A study on data aggregation techniques in wireless sensor network in static and dynamic scenarios. *Innovations Syst Softw Eng* 15(1):3–16
- Sardar TH, Khatun A, Khan S (2017 December) Design of energy aware collection tree protocol in wireless sensor network. In 2017

- IEEE international conference on circuits and systems (ICCS), IEEE, pp 12–17
- Selvi M, Thangaramya K, Ganapathy S, Kulothungan K, Nehemiah HK, Kannan A (2019) An energy aware trust based secure routing algorithm for effective communication in wireless sensor networks. *Wirel Pers Commun* 105(4):1475–1490
- Sharma V, Kumar R, Kumar N (2018) DPTR: distributed priority tree-based routing protocol for FANETs. *Comput Commun* 122:129–151
- Singh K, Johari R, Singh K, Tyagi H (2019 October) Mercurial cipher: a new cipher technique and comparative analysis with classical cipher techniques. In: 2019 International conference on computing communication and intelligent systems (ICCCIS), IEEE, pp 223–228
- Song H, Sui S, Han Q, Zhang H, Yang Z (2020) Autoregressive integrated moving average model-based secure data aggregation for wireless sensor networks. *Int J Distrib Sens Netw* 16(3):1550147720912958
- Tabatabaei S, Rigi AM (2019) Reliable routing algorithm based on clustering and mobile sink in wireless sensor networks. *Wirel Pers Commun* 108(4):2541–2558
- Uvarajan K P, Gowri Shankar C (2020) An integrated trust assisted energy efficient greedy data aggregation for wireless sensor networks. *Wirel Pers Commun* 114:813–833
- Vinodha D, Anita EM (2019) Secure data aggregation techniques for wireless sensor networks: a review. *Arch Comput Methods Eng* 26(4):1007–1027
- Wu H, Zhu H, Zhang L, Song Y (2019) Energy efficient chain based routing protocol for orchard wireless sensor network. *J Electric Eng Technol* 14(5):2137–2146
- Yousefpoor MS, Barati H (2020) DSKMS: a dynamic smart key management system based on fuzzy logic in wireless sensor networks. *Wirel Netw* 26:2515–2535. <https://doi.org/10.1007/s11276-019-01980-1>
- Yarinezhad R, Hashemi SN (2019a) Solving the load balanced clustering and routing problems in WSNs with an fpt-approximation algorithm and a grid structure. *Pervasive Mobile Comput* 58:p101033
- Yarinezhad R, Hashemi SN (2019b) Exact and approximate algorithms for clustering problem in wireless sensor networks. *IET Commun* 14(4):580–587
- Yarinezhad R, Hashemi SN (2020) Increasing the lifetime of sensor networks by a data dissemination model based on a new approximation algorithm. *Ad Hoc Netw* 100:102084
- Zhang J, Lin Z, Tsai PW, Xu L (2020) Entropy-driven data aggregation method for energy-efficient wireless sensor networks. *Inf Fusion* 56:103–113

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.