Received: November 17, 2020. Revised: January 23, 2021.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Secure Data Aggregation Using Ranking Strategy and Intelligence Similarity Function (SDARIS) for Wireless Sensor Networks

Reshma Siddegowda[1]* **Shaila Kalanath[2]** **Venugopal Kuppanna Rajuk[3]**

[1]*Visvesvaraya Technological University – Research Resource Center, Belagavi, India*
[2]*Vivekananda Institute of Technology, Bangalore, India*
[3]*Bangalore University, India*
* Corresponding author's Email: author@fit.ac.jp

Abstract: The environmental monitoring application is widely used in agriculture, green house farming, floriculture to monitor temperature range in different zone. This application is an embedded application fabricated with wireless sensors, smart phones, protocols, etc. Therefore, with respect to sensor nodes, the battery lifetime and security are the greater challenge in Wireless Sensor Network. In our work, we have designed a protocol using ranking strategy and intelligence similarity function. Initially, the clusters are formed based on the radius range. The SDARIS protocol focused on training the outer nodes of the cluster to acquire balanced cluster. The outer nodes are trained using ranking strategy and are assigned based on Euclidian distance from outer node to cluster head's and node density of the clusters. This ranking strategy assist to balance cluster size by assigning nodes evenly among all clusters. In addition to balance a cluster size, the SDARIS concentrates on aggregates and secure data during transmission using intelligence similarity function and session keys. The session key is generated for every interval, so the intruder node fails to identify the session key and thus mitigates intruder intervention in the network. Thus, the collaboration of ranking strategy, intelligence similarity function and session key concept in SDARIS protocol improve the overall performance of the network. Hence, SDARIS protocol achieved greater performance with increment of 12% data accuracy, 10% of cluster accuracy, 15% network lifetime and 26% throughput when compared with existing protocol.

Keywords: Cluster accuracy, Communication costs, Data accuracy, Environmental monitoring, Machine learning.

1. Introduction

The emerging technologies are embedded with sensors to fabricate smart devices. The trade-off between battery energy and behaviour of the sensor nodes are of concern during fabrication. Sensor nodes are bounded with resource constrained parameters and possessed different challenges viz, security, data aggregation, energy, availability, operating system, etc. [1-3].

The energy and security challenges have been discussed various algorithms [4-7] based on cluster approach that balances energy resource uniformly among the network and achieves greater energy efficiency. The concept of aggregating data and guaranting data delivery to the base station are discussed in [8, 9]. These algorithms focused on clustering and security of the data, but are not connected on accurate data delivery.

Thus, in our work we have adapted ranking strategy and intelligence similarity function. The different types of training techniques in WSNs, viz, Node localization, high dimensionality reduction and Support vector machine (SVM) respectively [10-12].

1.1 Motivation

It is required to aggregate the data accurately so as to balance the energy discharged throughout network lifetime. Then, aggregated data is forwarded to the base station. Therefore, during data transmission security becomes a challenging task. The SDARIS protocol is proposed to increase data accuracy, cluster accuracy, network lifetime, throughput and packet delivery ratio.

1.2 Contribution

The SDARIS protocol is designed with Ranking Strategy. This involves intelligence training methodology for sensor nodes. The sensor nodes are adopted spontaneously for different network size and topology. The spontaneous behaviour of sensor nodes works together to balance the energy and training is provided to sensor nodes. The SDARIS protocol uses intelligence similarity function to reduce redundant data transmission and assigns session key for each interval to mitigate intruder intervention in the network. Thus, intelligent behaviour of sensor nodes improves network connectivity of a node and packet delivery is guaranteed.

1.3 Organization

The review of existing work is discussed in Section 2 and background work is presented in section 3. Problem definition is stated in section 4. Section 5 depicts and explore the system diagram and even discussed about the mathematical model of proposed work. The performance evaluation and comparison with existing protocol is presented in Section 6. Section 7 is encapsulated with conclusions.

2. Related work

The review of energy protocols related to secure and balanced routing for data transmission is discussion in this section.

Deshpande et al., [13] discussed different similarity measures. Jacques et al., [14] proposed filtering technique which aggregated data by eliminating redundant information using set collected joins and similarity functions. Filtering techniques is applied periodically on stored data to preserve scarce energy. Thus, it saves energy and increases network lifetime. But filtering techniques is based on Jaccard similarity which fails to achieve data accuracy.

Hamid et al., [15] discussed on comprehensive review of humidity sensors. Yin et al., [16] designed a model based on spatial clustering and Principal Component Analysis (PCA). This model compresses data before transmission to the CH and CH is selected on energy basis to minimize energy consumption. The spatial correlation is used to group the sensor nodes to form a cluster. Thus, it prolongs network lifetime, but considering energy parameter alone does not balance the energy in the network. It leads to increased network disconnection and chances are more to introduce malicious node while electing CH by pretending that, it possesses maximum energy.

Diwakaran et al., [17] developed a data-aware energy conservation approach to aggregate data. The CH is responsible to aggregate the data and compress data using PCA technique. The sensor nodes monitor the network for every round. If the sensor node encounters any difference in actual data and predicted data, then the difference details are communicated with cluster head (CH). Then, cluster head uses PCA technique to compress the received data. Thus, it reduces redundant data during data transmission. This model violates data if the difference of actual data is less with predicted data even though if the sensed actual data is important to notice.

Kuila et al., [18] presented clustering scheme for data aggregation and communication of aggregated data. This scheme uses load balanced clustering concept. Therefore, it achieves greater network lifetime. But the scheme does not focus on cluster heads residual energy. Rostami et al., [19] addressed various existing homogeneous and heterogeneous clustering algorithms to balance energy consumption in the network.

Mahnaz et al., [20] proposed a clustering algorithm using fuzzy logic residual energy of cluster head to determine the distance and density of a node connected to cluster head. This algorithm is focused on cluster formation and improves network lifetime But, this protocol is not concerned with security aspect of the data.

Yuan et al., [21], presented data density correlation degree (DDCD) for clustering to aggregate data in Wireless Sensor Network. DDCD correlation consists of 3 functions that verifies data density correlation degree of a sensor node. Cluster formation based on data density is done and verified with the results. Finally, clusters are merged with outer nodes. According to the information of DDCD preserved sensor node routing table. Therefore, it achieves greater network connectivity. This algorithm uses naming concept for nodes based on the action of sensor nodes that leads to an overhead to identify nodes each time and increased the time complexity.

3. Background

Ihsan et al., [22] uses cosine similarity function to form clusters and to eliminate data. This protocol uses Inter Quartile paradigm to remove an outlier. Thus, SOMDA accomplish maximum Network Lifetime and minimal energy utilization.

In our work, we used ranking strategy and intelligence similarity function. The sensor nodes are trained to adoptable changes in the network. Thus, sensor nodes spontaneously form the clusters and eliminate redundancy. Hence, SDARIS protocol acquire prolonged network lifetime, data accuracy,

Received: November 17, 2020. Revised: January 23, 2021.

packet delivery ratio, throughput and cluster accuracy.

4. Problem definition

The smart device is fabricated with 'N' sensor nodes, protocols and other hardware components. It is a challenging task to prolong the nodes battery life. Hence, it is required to adopt train the outer node to spontaneously connect to respective cluster to balance cluster size to utilize energy in the balanced manner. It is required to identify fraud data introduced by intruder.

The main objective of SDARIS protocol are:

(i) Balance energy utilization in the network.

(ii) Mentoring sensor nodes using neural networks.

(iii) Mitigate fraud data introduced by the intruder.

Assumption:

(i) All nodes initial energy is same.

(ii) Moderate temperature range is considered between 30°C - 35°C.

5. System and mathematical model

The SDARIS protocol is proposed to guide sensor nodes to cluster itself in a large sparse sensor network using Artificial Neural Network approach. The SDARIS system model is shown in Figure 1. which consists of two different phases.

1. Clustering Formation using Ranking Strategy (CFRS) Phase.

2. Intelligent and Secured Data Aggregation (ISDA) Phase.

Lemma 1: Let $n = \{n_1, n_2, \cdots n_p\}$ and $G = \{ G_1, G_2, \cdots G_m\}$ where $\{n \in G : \|d\| \approx r \dashv | \|d\| < r\}$

Proof: Set n is composed of p sensor nodes which are clustered into m clusters as $G_1, G_2, \cdots G_m$. Initially, CH is elected randomly, and cluster formation is done based on the distance parameter from CH. The nodes which are closest to the specified radius range *'r'* from the cluster head will be grouped together to form clusters.

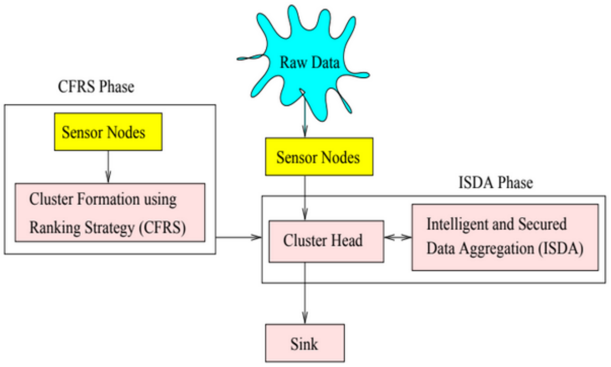

Figure. 1 SDARIS system flow diagram

Clustering formation using ranking strategy (CFRS) phase

As per Lemma 1, the sensor nodes are grouped to form clusters initially. With reference to Fig. 2, it is shown that several sensor nodes are situated outside the cluster in the network. This phase is focused to train outer nodes to connect to respective clusters. The cluster is selected based on the Euclidian Distance (ED), Density of a nodes (DN) in each cluster, Rank of Euclidian Distance (R_{ED}) and Rank of Density of a Node (R_{ND}).

Euclidian distance (ED)

The distance from each outer node to Cluster Head (CH) is computed using Eq. (1) so as to connect the nearest cluster.

$$ED = \sum_{i\leftarrow1}^{k}\sum_{j\leftarrow1}^{m}\sqrt{(x_i - x_{CH_j})^2 + (y_i - y_{CH_j})^2} \tag{1}$$

Density of a node (DN)

The density of a node in each cluster is computed to obtain the cluster information which has low density. This information helps to connect an outlier node to lower density cluster to balance the cluster size in the network. So, that balance the network lifetime and energy usage.

where k = number of outer nodes in the network and m = number of clusters.

Rank of euclidian distance (RED)

The rank is assigned to computed Euclidian distance value using Eq. (1). The rank criteria are considered by using,

$$\begin{matrix} R_{ED_1} \\ R_{ED_2} \\ \vdots \\ R_{ED_n} \end{matrix} = \begin{matrix} First\_\min(ED) \\ Second\_\min(ED) \\ \vdots \\ max(ED) \end{matrix} \tag{2}$$

The R_{ED_1} = first rank assigned to the node which possess minimum distance towards representative cluster among all the clusters in the network. R_{ED_2} = second rank and so on.

Received: November 17, 2020. Revised: January 23, 2021.

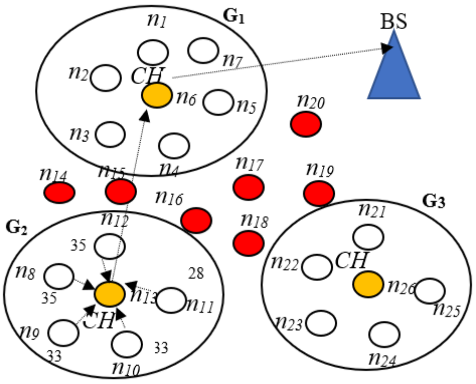

Figure. 2 Example of SDARIS network architecture

Rank of density of a node (RND)

The rank is assigned to clusters, based on the preserved information about Density of a Node (DN). The selection criteria are done using,

$$\begin{matrix} R_{DN_1} \\ R_{DN_2} \\ \vdots \\ R_{DN_n} \end{matrix} = \begin{matrix} First\_\max(DN) \\ Second\_\max(DN) \\ \vdots \\ min(DN) \end{matrix} \tag{3}$$

where, R_{DN_1} is considered as a first rank that is assigned to the cluster which incur maximum nodes, R_{ED_2} is the second rank and so on.

Based on the above calculated parameters, outer node is trained to select the cluster itself based on the trade-off between two ranks associated to the clusters.

Lemma 2:

$$\boldsymbol{n \, \varepsilon \, G \leftrightarrow G} = \begin{cases} \mathbf{G(R_D(n)) \Leftrightarrow D(R_D(n)) - D(R_D(n-1)) \leq \alpha} \\ \quad \mathbf{and \; ND(R_{ND}(n)) < ND(R_{ND}(n-1))} \\ \mathbf{G(R_D(n-1)) \Leftrightarrow (R_D(n)) - D(R_D(n-1)) \leq \alpha} \\ \quad \mathbf{and \; ND(R_{ND}(n)) > ND(R_{ND}(n-1))} \end{cases}$$

Illustration 1:

In Fig. 2, let *N=25, n = {n₁, n₂, n₃, n₄, n₅, n₆, n₇, n₈, n₉, n₁₀, n₁₁, n₁₂, n₁₃, n₁₄, n₁₅, n₁₆, n₁₇, n₁₈, n₁₉, n₂₀, n₂₁, n₂₂, n₂₃, n₂₄, n₂₅, n₂₆}, G= {G₁, G₂, G₃}, r=100m.*

Proof: Consider *n* as an outer node belonging to cluster G, if and only if cluster possess a trade-off between R_D and R_{ND}, where R_D is a Distance Rank and R_{ND} is a Node Density Rank. Let *D* $(R_D(n))$ and $D(R_D(n\text{-}1))$ be 1st minimum distance and 2nd minimum distance between outer node and neighbor cluster respectively. The *ND* $(R_{ND}(n))$ and *ND* $(R_{ND}(n\text{-}1))$ are 1st minimum node density and 2nd minimum node density among outer node's neighbor clusters. Whereas α is the threshold distance difference among the Closest Clusters for outer node. The $D(R_D(n))\text{-}D(R_D(n\text{-}1)) \leq \alpha$ computes the threshold difference between neighbor clusters. The cluster which lies within α range is considered as a Closest Cluster (CC). Then, the node density of CC's is identified. Finally, this lemma states that the node belongs to cluster which possess minimum Node Density among CC.

The process of selecting nodes to form a cluster is as follows:

Step 1: Electing cluster which exhibit maximum rank from R_{ED} located within the α difference of *ED* among all clusters is performed as discussed in Lemma 2.

Step 2: Minimum rank from R_{ND} among the selected cluster is presented as in Lemma 2.

So, there is a trade-off between distance and density of a node in clusters.

Illustration 2:

In Fig. 2, let N=25, n = {n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}, n_{11}, n_{12}, n_{13}, n_{14}, n_{15}, n_{16}, n_{17}, n_{18}, n_{19}, n_{20}, n_{21}, n_{22}, n_{23}, n_{24}, n_{25}, n_{26}}, G={G_1, G_2, G_3}, r=100m.

Step 1: Consider a network composed of 26 sensor nodes and 3 clusters. In the initial phase CH's are randomly elected and *r* range is specified in the network. Clusters are formed based on the resource parameter *r*. The nodes which are closest and within the coverage range of r is grouped into one cluster as discussed in Lemma 1. Therefore, the cluster formation is performed as shown:

G_1= {n_1, n_2, n_3, n_4, n_5, n_6, n_7},
G_2 = {n_8, n_9, n_{10}, n_{11}, n_{12}, n_{13}},
G_3 = {n_{21}, n_{22}, n_{23}, n_{24}, n_{25}, n_{26}}

Step 2: The node which lies outside the coverage area is treated as neurons. Then, the training is given for each neuron to gain an intelligence for detecting a cluster the neuron belongs to. As shown in Fig. 2, the Neuron Set (NS) is given as:

NS={n_{14}, n_{15}, n_{16}, n_{17}, n_{18}, n_{19}, n_{20}}

Table 1. Coordinates and distance to clusters of outer nodes

Node ID	**X**	**Y**	**D_{CH1}**	**D_{CH2}**	**D_{CH3}**
n_6 (G_1 – CH)	150	450	-	-	-
n_{13} (G_2 – CH)	100	200	-	-	-
n_{26} (G_3 – CH)	350	200	-	-	-
n_{14}	80	300	165.5	101.8	287.9
n_{15}	110	310	145.6	110.4	264.0
n_{16}	150	290	160.0	102.9	219.3
n_{17}	220	340	130.3	184.3	191.0
n_{18}	220	270	193.1	138.9	147.6
n_{19}	300	345	183.0	247.0	153.3
n_{20}	290	400	148.6	275.8	208.8

Table 2. Cluster ID (CID), density of a node (DN) of each cluster, euclidian distance (ED), rank of DN (rank$_{DN}$) and rank of ED (rank$_{ED}$)

Node ID	CID	DN	ED	Rank$_{DN}$	Rank$_{ED}$
n$_{14}$	1	8	165.52	2	2
	2	6	101.80	3	3
	3	9	287.92	1	1
n$_{15}$	1	8	145.60	2	2
	2	6	110.45	3	3
	3	9	264.00	1	1
n$_{16}$	1	8	160.00	2	2
	2	6	102.95	3	3
	3	9	219.31	1	1
n$_{17}$	1	8	130.38	2	3
	2	6	184.39	3	2
	3	9	191.04	1	1
n$_{18}$	1	8	193.13	2	1
	2	6	138.92	3	3
	3	9	147.64	1	2
n$_{19}$	1	8	183.09	2	2
	2	6	247.03	3	1
	3	9	153.37	1	3
n$_{20}$	1	8	148.66	2	3
	2	6	275.86	3	1
	3	9	208.80	1	2

Step 3: In the Fig. 2, nodes *{n$_{14}$, n$_{15}$, n$_{16}$}* neither belongs to cluster *G$_1$* nor *G$_2$*. Assume,

$$R_1(n_{14}, CH(G_1)) = 120m$$
$$R_2(n_{14}, CH(G_2)) = 140\ m$$
$$R_1(n_{15}, CH(G_1)) = 105m$$
$$R_2(n_{15}, CH(G_2)) = 110\ m$$
$$R_1(n_{16}, CH(G_1)) = 102m$$
$$R_2(n_{16}, CH(G_2)) = 150\ m$$

As per Lemma 2 from Table 2 - for node n$_{14,}$ Consider,

$$Rank_{ED_1} = 287.9236$$
$$Rank_{ED_2} = 165.5294$$
$$Rank_{ED_1} - Rank_{ED_2} \leq 50$$

∴ Consider $Rank_{ND_1}$ and $Rank_{ND_2}$ of corresponding $Rank_{ED_1}$ and $Rank_{ED_2}$, since two clusters Euclidian distance difference from node n_{14} is less than 50. Hence, these two clusters are considered with their respective density of node and train the node to elect the cluster which has low node density to balance the energy utilization.

$$Rank_{ND_1} = 9$$
$$Rank_{ND_2} = 8$$

Therefore, the above ranks prove that n_{14} selects cluster *G$_2$* and get connected. Hence, similarly rank computed for *n$_{14}$* and *n$_{16}$* and proves that both the nodes are strongly correlated with cluster *G$_2$* and *n$_{15}$* is correlated with cluster *G$_1$*. Hence, clusters *G$_1$* and *G$_2$* are reformed as

$$G_1 = \{n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_{14}, n_{16}\},$$
$$G_2 = \{n_8, n_9, n_{10}, n_{11}, n_{12}, n_{13}, n_{15}\}$$

Intelligent and secure data aggregation (ISDA)

The given network is now active to monitor the temperature. All sensor nodes in the cluster, monitors the environment temperature range and forwards it to cluster head. Sensor nodes establishes keys for every session and the session expires for every 2 minutes. In addition to sensed data, sensor nodes forwards established session key to cluster head. Cluster Head is trained with intelligence to identify false data injected by malicious node and to mitigate redundant data collected from sensor nodes. Cluster Head is also responsible to remove outliers of the temperature range using Interquartile.

The procedure of training Cluster Heads using intelligence to aggregate data and for secure data transmission is as follows:

Step 1: The random weight is assigned for each cluster initially, and obtain the variation as,

$$Y = \sum_{i\leftarrow1}^{m}(X_i - W_i) \quad (4)$$

Step 2: Interquartile is computed using,

$$InterQuartile = Maximum - Minimum \quad (5)$$

Step 3: The upper environment temperature range is computed as,

$$U_r = Maximum + 1.5InterQuartile \quad (6)$$

Step 4: The lower environment temperature range is obtained using,

$$L_r = Minimum - 1.5 * InterQuartile \quad (7)$$

Cluster _Id	CID	CH R$_E$	Node _ID	Session Key	Old Session Key

Figure. 3 Sensor node's routing table

Cluster_Id	Node_Id	Old Session Key	Session Key	Residual Energy

Figure. 4 Cluster head's routing table

Table 3. Simulation parameter

Parameter	Values
Number of Nodes	50,100,200
Simulation Topology	1000m*1000m
Traffic	CBR
Transmission Range	40m
Number of clusters	16
Initial energy	1J
Data packet size	64
Energy Consumed during Transmission	0.016J
Energy consumed during Reception	0.018J
Energy consumed during Idle condition	0.0005J
Simulation Time	20000 second

Fig. 3 maintains old and present session key along with basic details. If session key of node generated data is matched then, session key is stored in CH's routing table (shown in Fig. 4) for respective node. Then, data is considered as a trusted data else data is considered as a false data injected by malicious node. Finally, the aggregated data forwards to the base station. The base station utilizes aggregated data as per the requirement of particular application.

6. Simulation and performance evaluation

The SDARIS protocol is implemented using network simulator 2 (NS2). The simulation is performed for various network size from 50 nodes to 100 nodes. The nodes and network configuration are presented in Table 3.

6.1 Performance metrics

1. ***Data Accuracy:*** It is the exact amount of unique data aggregated during data aggregation.
2. ***Packet Delivery Ratio (PDR):*** It is the fraction of amount of data delivered to destination with respect to time.
3. ***Cluster Accuracy:*** It is defined as how exactly outer node incur to the balanced and efficient cluster.
4. ***Network Lifetime:*** It is defined as the duration of network connectivity.
5. ***Throughput Rate:*** It is the rate of data packet transmitted per unit time.

6.2 Performance evaluation

Table 4 shows actual result of data accuracy when SDARIS protocol is applied. The protocol uses intelligence similarity approach to aggregate data. This approach successfully eliminate data for different data size whereas SOMDA uses cosine similarity to eliminate redundant data and it is not suitable for different data sizes. Thus, SDARIS protocol achieves 12% greater efficiency than SOMDA protocol because SDARIS protocol applied for large set of data size and it aggregates data accurately with the use of intelligent similarity function. Fig. 5 exhibits the comparison graph of SDARIS and SOMDA protocol.

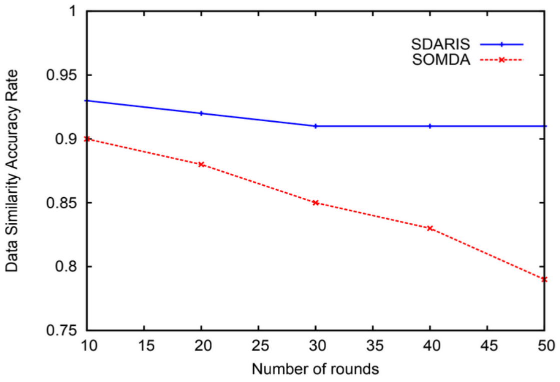

Figure. 5 Data accuracy

The cluster accuracy analysis results are numerically represented in Table 4 and graphically represented in Fig. 6. The intelligence self-learning and organization technique is adopted for clusters formation using density of a node (ND), Euclidian distance ED, Rank of distance between nodes and nearest clusters (R_{ND}). The trade-off between R_{ND} and R_{ED} select the nearest cluster. This ranking exhibit exact cluster accuracy and 10% improvement over SOMDA protocol. SOMDA protocol degrades the network performance as it uses cosine similarity which is not appreciable for large scale networks.

The SOMDA and SDARIS protocol's network lifetime comparison result is shown in Fig. 7. The SDARIS algorithm is focused on balanced energy utilization *i.e.*, during cluster formation the outer nodes are guided in such a way that the outer nodes join to a cluster which consists of less number of nodes in the cluster, so that nodes are uniformly disturbed among the network. Therefore, network connectivity extends for longer time. But in SOMDA protocol, the cosine similarity function is used to form a cluster. The cosine similarity function is not suitable for large sparse network. It degrades the network performance when the network size is large. Whereas in SDARIS protocol, the CFRS algorithm incorporated ranking strategy focused on evenly organizing the cluster size. Another disadvantage of SOMDA protocol is, it's not focused on security aspect. Hence, SDARIS protocol also compared with SAR protocol on aspect of security. The SAR protocol mitigated grey hole attack, black hole attack and worm hole attack. The SAR compromises with

Received: November 17, 2020. Revised: January 23, 2021.

Table 4. Cluster accuracy, data accuracy, network lifetime and live nodes

Number of Rounds	Clustering Accuracy in percentage		Data Similarity Accuracy Rate		Network Lifetime in percentage		
	SOMDA	SDARIS	SOMDA	SDARIS	SAR	SOMDA	SDARIS
10	90	93	0.90	0.93	90	90	96
20	89	91	0.88	0.92	73	72	90
30	88	90	0.85	0.91	45	58	70
40	85	90	0.83	0.91	15	29	40
50	80	90	0.79	0.91	2	5	20

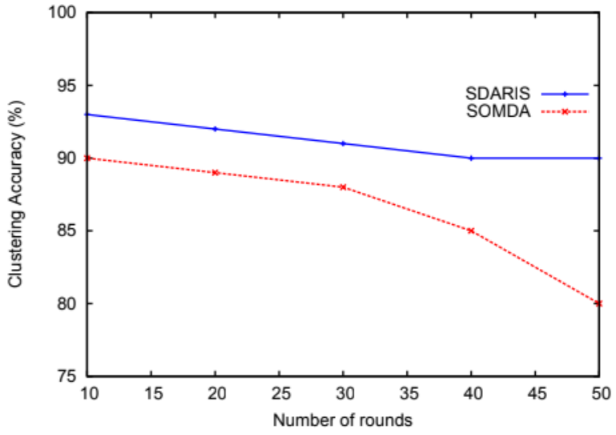

Figure. 6 Cluster accuracy

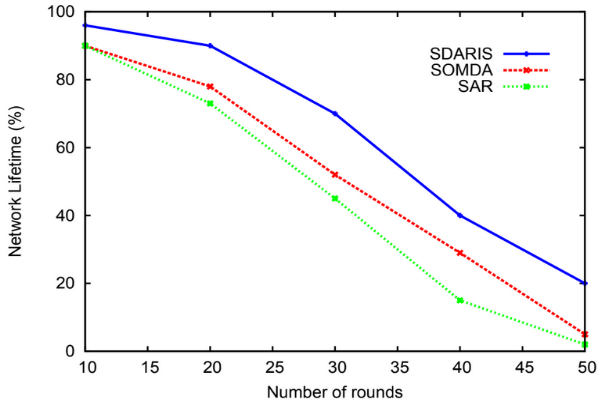

Figure. 7 Network lifetime

Table 5. Packet delivery ratio

Number of Malicious Nodes	Data Length in Bytes		
	SAR	SOMDA	SDARIS
2	0.98	0.97	1
4	0.98	0.97	1
6	0.98	0.97	1
8	0.98	0.97	1
10	0.98	0.97	1
12	0.98	0.97	1

Table 6. Throughput

Number of Nodes	Data Length in Bytes		
	SAR	SOMDA	SDARIS
50	86.152	85.67	92.23
100	120.653	110.35	170.75
150	135.263	125.45	172.45
200	150.635	140.35	173.65
250	160.593	148.45	179.22
300	170.364	156.35	182.35

SOMDA protocol has not focused on securing data during transmission. The SOMDA protocol has not incorporated appropriate algorithm for security wherein it is focused only on cluster accuracy and data accuracy. The SDARIS protocol is also compared with SAR protocol which discovers route using ‘quantifiable guarantee of security’. Thus, SAR protocol concentrates on security with increased latency. The SDARIS protocol focused both on latency and energy utilization during data transmission. It utilizes more energy for computation to provide security for the data. Hence, SDARIS protocols shows 15% improvement of network lifetime over SOMDA protocol and 17% improvement over SAR protocol.

The SDARIS protocol delivers sensitive information successfully to the base station. The false data injected by intruder is strongly mitigated with the keys generated by cluster head for every time interval. The cluster head detects the key that is assigned to sensor nodes during the interval along with data sent by sensor nodes. Therefore, this strategy guarantees packet delivery. Whereas

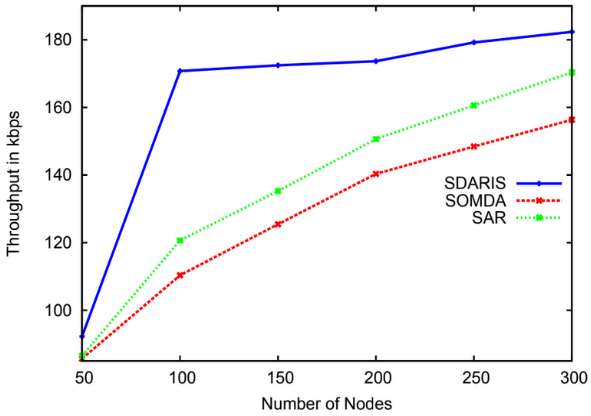

Figure. 8 Throughput

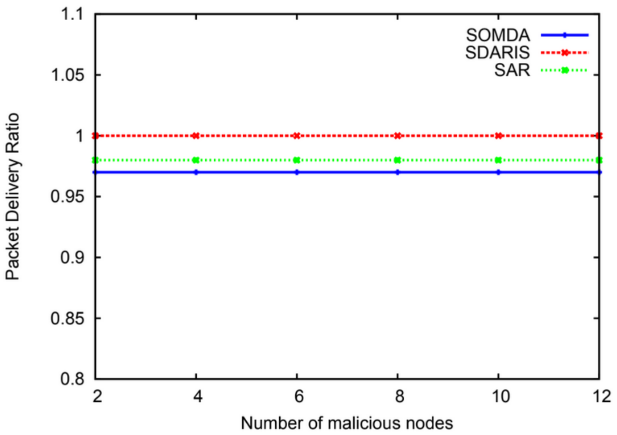

Figure. 9 Packet delivery ratio

energy efficiency and security. Hence, SDARIS protocol exhibits better PDR when compared with SOMDA and SAR protocols as shown in Fig. 9.

7. Conclusions

The SDARIS protocol achieved greater network connectivity and guaranteed data transmission in Wireless Sensor Networks. The intelligence learning approach adaption to sensor nodes are developed based on rank between the nodes and nearest cluster in the network, Euclidian distance and density of a node. The outliers of the temperature range are eliminated by computing the Inter Quartile. Then, the data is aggregated by the cluster head using similarity technique. Finally, this aggregated data is transmitted to the base station. The cluster head provides security for the data before data transmission to the base station with session key establishment process in the specified time interval to mitigate internal and injection of false data in the network. Therefore, SDARIS protocol exhibits greater performance with increment of 12% data accuracy, 10% of cluster accuracy, 15% network lifetime and 26% throughput compared with existing protocol. The cluster heads in SDARIS protocol is not accessible to all cluster members with one hop distance since the outer nodes connected to respective clusters are lies far away from cluster head. Therefore, this work can be further extended with appropriate centroid algorithm to elect a cluster head which are accessible to all the cluster member.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

"Conceptualization - Dr. Shaila K and Dr. Venugopal K R; Methodology – Reshma S; Software – Reshma S; Validation – Reshma S, Dr. Shaila K and Dr. Venugopal K R; Formal Analysis, Investigation, Resources, Data Curation, Writing—Original Draft Preparation – Reshma S; Writing – Review, Editing, and Supervision – Dr. Shaila K; Visualization – Dr. Venugopal K R."

References

[1] P. Gope, A. K. Das, N. Kumar, and Y. Cheng, "Lightweight and Physically Secure Anonymous Mutual Authentication Protocol for Real-Time Data Access in Industrial Wireless Sensor Networks", *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 9, pp. 4957-4968, Sept. 2019, doi: 10.1109/TII.2019.2895030.

[2] S. Pundir, M. Wazid, D. P. Singh, A. K. Das, J. P. C. Rodrigues, and Y. Park, "Intrusion Detection Protocols in Wireless Sensor Networks Integrated to Internet of Things Deployment: Survey and Future Challenges", *IEEE Access*, Vol. 8, pp. 3343-3363, 2020, doi: 10.1109/ACCESS.2019.2962829.

[3] A. Mallikarjuna, V. Reddy, A. V. U. Phani Kumar, D. Janakiram, and G. Ashok Kumar, "Wireless Sensor Network Operating Systems: A Survey" *International Journal of Sensor Networks*, Vol. 5, No. 4, pp. 236-255, 2009.

[4] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and Challenges of Wireless Sensor Networks in Smart Grid", *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 10, pp. 3557-3564, 2010, doi: 10.1109/TIE.2009.2039455.

[5] A. E. Mohammed, M. A. Elrazik, M. E. Bakry, S. Q. Hasan, A. Q. Hasan, and S. Zaid, "Challenges in Wireless Sensor Networks", *International Journal of Advanced Research in Computer Science & Technology (IJARCST)*, Vol. 4, No. 4, pp. 22-27, 2016.

[6] H. Farman, H. Javed, B. Montrucchio, M. Khan, and S. Ali, "Energy Efficient Hierarchical Clustering Approaches in Wireless Sensor Networks: A Survey", *Wireless Communications and Mobile Computing*, Vol. 2017, pp. 1-14, 2017.

[7] V. Geetha, V. K. Pranesh, and T. Sushma, "Clustering in Wireless Sensor Networks: Performance Comparison of LEACH & LEACH-C Protocols Using NS2", In: *Proc. of 2nd International Conf. on Computer, Communication, Control and Information Technology (C3IT-2012)*, Vol. 14, pp. 163-170, 2012.

[8] A. Aseeri and R. Zhang, “Secure Data Aggregation in Wireless Sensor Networks: Enumeration Attack and Countermeasure”, In: *Proc. of IEEE International Conf. on Communications (ICC),* pp. 1-7, 2019, doi: 10.1109/ICC.2019.8761889.

[9] S. B. Othman, A. Trad, H. Youssef, and H. Alzaid, “Secure Data Aggregation in Wireless Sensor Networks”, *12th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET),* pp. 55-58. 2013.

[10] D. K. Praveen, T. Amgoth, C. S. Rao and Annavarapu, “Machine Learning Algorithms for Wireless Sensor Networks: A Survey”, *Information Fusion,* Vol. 49, pp. 1-25, 2019, doi: https://doi.org/10.1016/j.inffus.2018.09.013

[11] S. Banihashemian, F. Adibnia, and M. A. Sarram, “A New Range-Free and Storage-Efficient Localization Algorithm using Neural Networks in Wireless Sensor Networks”, *Wireless Personal Communications*, Vol. 98, No. 1, pp. 1547–1568, 2018.

[12] F. Zhu and J. Wei, “Localization Algorithm for Large Scale Wireless Sensor Networks Based on Fast-SVM”, *Wireless Personal Communications,* Vol. 95, No. 3, pp. 1859–1875, 2017.

[13] R. Deshpande, B. V. Sluis and C. L Myers, “Comparison of Profile Similarity Measures for Genetic Interaction Networks”, *PLoS One,* Vol. 8, No. 7, pp. 1-11, 2013.

[14] J. M. Bahi, A. Makhoul, and M. Medlej, “Data Aggregation for Periodic Sensor Networks using Sets Similarity Functions”, In: *Proc. of 7th International Wireless Communications and Mobile Computing Conf.,* pp. 559-564, 2011 doi: 10.1109/IWCMC.2011.5982594.

[15] F. Hamid, W. Rahman, and H. Mohd, “Humidity Sensors Principle, Mechanism, and Fabrication Technologies: A Comprehensive Review”, *Sensors,* Vol. 14, No. 5, pp. 7881-7939, 2014.

[16] Y. Yin, F. Liu, X. Zhou, and Q. Li, “An Efficient Data Compression Model Based on Spatial Clustering and Principal Component Analysis in Wireless Sensor Networks”*, Sensors,* Vol. 15, No. 8, pp. 19443-19465, 2015.

[17] S. Diwakaran, B. Perumal, and K. V. Devi, “A Cluster Prediction Model-Based Data Collection for Energy Efficient Wireless Sensor Network”, *The Journal of Supercomputing,* Vol. 75, pp. 3302–3316, 2018.

[18] P. Kuila and P. K Jana, “Approximation Schemes for Load Balanced Clustering in Wireless Sensor Networks”, *Journal of Supercomputing,* Vol. 68, pp. 87-105, 2013.

[19] A. S. Rostamil, M. Badkoobe1, F. Mohanna, H. keshavarz, A. Asghar, R. Hosseinabadi, and A. K. Sangaiah, “Survey on Clustering in Heterogeneous and Homogeneous Wireless Sensor Networks”, *The Journal of Supercomputing*, Vol. 74, pp. 277-323, 2018.

[20] M. Toloueiashtian and H. Motameni, “A New Clustering Approach in Wireless Sensor Networks using Fuzzy System”, *The Journal of Supercomputing,* Vol. 74, pp. 717–737, 2018.

[21] F. Yuan, Y. Zhan, and Y. Wang, “Data Density Correlation Degree Clustering Method for Data Aggregation in WSNs”, *IEEE Sensors Journal,* Vol. 14, No. 4, pp. 1089 - 1098, 2014.

[22] Ullah and H. Y. Youn, “A Novel Data Aggregation Scheme Based on Self-Organized Map for WSNs”, *The Journal of Supercomputing,* Vol. 75, pp. 3975–3996, 2019.

[23] S. Archana and A. S. Salvan, “SAR Protocol Based Secure Data Aggregation in Wireless Sensor Networks”, In: *Proc. of Ninth International Conf. on Intelligent Systems and Control (ISCO)* pp. 1-6, 2015.