A Secure Aggregation Routing Protocol with Authentication and Energy-Saving on Data Mining and Big Data

Ying Wang, Bo Liu, and Fengyin Li(✉)

School of Computer Science, Qufu Normal University, Rizhao 276826, China

Abstract. In the era of big data, with the mining of the value of big data, wireless sensor network has been more and more widely used. However, due to the limited and uneven resource distribution of the sensor equipment, the service life of the wireless sensor is seriously affected. Moreover, the problem of privacy leakage in wireless sensor networks is becoming more and more serious. In order to solve the resource-constrained and privacy-leaking problems of nodes, we propose a secure convergent routing protocol with authentication and energy-saving. First, a new cost function calculation method is proposed. Aiming at the resource limitation of sensor nodes, the communication cost calculation function of each node is designed by synthesizing the residual energy of the node, the distance from the sink node and the current link quality, to measure the communication cost of wireless sensor nodes dynamically. The data transmission path is designed by changing the cost function. Secondly, a key agreement protocol with authentication function is designed. In order to solve the problem of privacy leakage in the communication process, a trusted registration node is used to register the node and distribute the authorization information, mutual authentication and session key negotiation are carried out, a key agreement scheme with authentication function is designed for safe and efficient key distribution. Finally, through performance analysis, we verify that our protocol consumes less total energy than the other two related protocols.

Keywords: Big data · Wireless sensor network · Energy saving · Authentication · Key agreement

1 Introduction

With the rapid development of social network and mobile Internet technology, the collection, storage, analysis and release of all kinds of data become convenient and fast. Organizations such as health care, insurance companies, e-commerce companies, social networking sites, and telecom operators publish industry data for research, and a variety of big data analytics companies, big data analytics competitions, and more. In the era of big data, businesses are hard at work

Y. Tan et al. (Eds.): DMBD 2021, CCIS 1453, pp. 192–204, 2021.
https://doi.org/10.1007/978-981-16-7476-1_18

mining data to analyze usage patterns and improve product features or user experience [3].

With the value mining of big data, the information related to individuals and enterprises will increase significantly. Data transmission and data privacy protection in big data and data mining have also aroused widespread concern. In big data environments, data is collected by sensors, which collect data to make more informed decisions that help organizations develop in an internet environment. Wireless technology enables sensors to be connected to each other to form a "Large network" [9]. Sensors with sensing data, communication and processing capabilities are connected to each other, which is a typical wireless sensor network. Wireless sensor networks are used in many fields, such as agriculture to monitor environmental temperature and humidity, traffic network, traffic flow and so on. They can also be deployed in many places where people can not reach, such as deserts, so there's a military component [6].

However, wireless sensor networks also have some limitations [10], mainly facing the following two challenges:

(1) Sensor equipment has the characteristics of micro-size and therefore also has the nature of limited resources. Energy consumption is a very outstanding problem in WSN communication.
(2) In some cases in a real-time data environment, external adversaries can directly access real-time data from sensor nodes.

In recent years, many people have made some research contributions on how to save energy consumption of sensor nodes. In 2019, Arpita Mallick et al. proposed a multi-hop routing protocol that transfers data from body area sensor networks (BAN) to cellular devices or display coordinators. In the hospital scenario, there can be multiple display coordinators near one of Ban's coordinators. In addition, in a hospital, multiple patients may have wearable or implantable sensors and form multiple BAN configurations nearby [4]. In 2020, Muhammad Ilyas et al. proposed a three-tier cluster-based routing protocol for wireless sensor networks. This protocol utilizes three-layer clustering mechanism, which is a center-based clustering protocol [1]. Researchers have also done some technical research on the issue of data privacy breach. In 2020, Naswan S et al. proposed an anonymous access authentication scheme for wireless sensor networks in large data environments, using a set of lightweight symmetric encryption and hash function to prevent existing known attacks, perfect forward secrecy provides two way authentication and complete anonymity [7]. 2021, Inam ul haqa et al. have proposed an efficient multi server architecture based on Hasche's authenticated key protocol scheme, which uses a simple hash operation to achieve key sharing and identity authentication [2].

All of the above studies give rise to the enlightenment of our scheme. In order to solve the problem of unbalanced energy consumption of sensor nodes, some nodes have to bear the problem of high energy consumption resources, and the problem of privacy leakage in data transmission between sensor nodes, we propose a secure converged routing protocol with authentication and energy-saving functions. The main contributions of this article are as follows.

(1) A new calculation method of cost function is proposed.
The cost function proposed in this paper can be used to effectively classify the cost consumption level of a sink node when transmitting data by selecting a node from all sensor nodes as a cluster "Leader" node, a reasonable path is planned for the realization of energy-saving converging routing protocol based on cost function.
(2) A new key agreement scheme with authentication function is proposed.
The proposed key agreement protocol is used for efficient key distribution in communication between two parties, which realizes data privacy of communication between nodes and ensures data security.
(3) The performance of the proposed protocol and other existing schemes is analyzed.
We compare the performance of the proposed scheme with the existing schemes, and clearly verify the efficiency of our proposed protocol in terms of total energy consumption.

This article is organized as follows. In Sect. 2, we introduce the basics used in this article. In Sect. 3, a new multi-layer energy-saving convergent routing protocol with authentication key agreement is proposed. In Sect. 4 provides a performance comparison of the proposed scheme with existing related schemes. In Sect. 5 summarizes the full text.

2 Preliminary

2.1 WSN

The network environment itself has the openness, if can not carry on the effective supervision to the network environment, the entire Internet system will certainly be disorderly [5]. Based on this, it is necessary to implement constraint management for the load network environment. Wireless Sensor Networks (WSN) is a kind of distributed Sensor Networks, the end of which is the Sensor that can sense and examine the outside world. Sensors in WSN communicate wirelessly, so network settings are flexible, device locations can be changed at any time, and wired or wireless connections can be made with the Internet. A multi-hop self-organizing network formed by wireless communication. The wireless sensor network (WSN) not only connects the wireless network with the wired network, but also senses and checks the state of the external sensors through flexible network settings, and then makes device changes, it's also a great way to avoid the dangers of the internet. Wireless network sensors can also collect and transfer data in the internet environment, not only reduce the cost of network transmission, but also simplify the form of network deployment. In the process of data transmission, the dynamic collection and detection of network resource information can be realized through the operation of internal sensor nodes, and the online monitoring of information transmission can be realized [11].

2.2 Cluster Structure in WSN

In wireless sensor networks, cluster-based routing protocol is called hierarchical routing protocol. The idea of hierarchy-based or cluster-based routing protocol is to divide the network into many different clusters based on certain attributes, including cluster head (CLH) and cluster member node (CM), cluster heads can communicate directly with Sink nodes or form a higher-level network, in which each cluster head node is regarded as a normal node and then clustered until the last node is left in the network. The cluster head node is responsible for the management of the cluster members and acts as the leader or coordinator. The collection and processing of the data of the cluster members are also carried out at the cluster head node. Meanwhile, the cluster head is responsible for the data transmission between clusters, data is taken from a cluster member node (CM), aggregated and forwarded to a base station (BS) or Sink node. The choice of CLH is based on important parameters such as residual energy and the distance to the sink node (SN). The CLH role rotates according to the level of the node. The level of a node is determined by the increase or decrease of important parameters. Now the most popular algorithm is to improve the energy balance of the network, the energy consumption of the network is evenly distributed to all nodes, the network is divided into periodic rounds. The advantage of hierarchical routing protocol is that it is scalable, can adapt to the dynamic changes of the network, and is suitable for large-scale networks [8]. Cluster-based WSN is divided into two-tier and three-tier hierarchies as shown in Fig. 1.

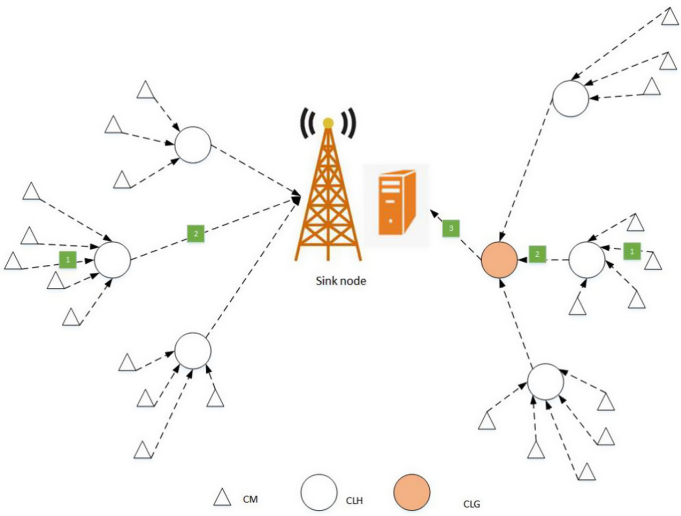

Fig. 1. Cluster-based WSN two-level structure and three-level structure.

3 Proposed Protocol

In order to solve the problems of high power consumption, network lifetime, low network throughput and privacy leakage in secret data transmission, we propose an energy-saving routing protocol with authentication and agreement key. This paper is organized as follows, in Sect. 3.1 gives the network model of the scheme, our scheme is implemented on the ordinary Internet of Things Network, in Sect. 3.2 we carry out the network deployment of the sensor nodes, and collected deployment information for the sink node. In Sect. 3.3, we first design a cost function to calculate a cost for each sensor node. The sensor node is chosen for its high energy and good link quality. The node near the sensor node becomes CLH (Cluster Head), CLG (Cluster Gateway) node. In Sect. 3.4, after selecting CLH node and CLG node, we use K-means clustering algorithm to divide sensor nodes near CLH node into the same cluster, and these cluster member nodes send information to CLH node, CLH then sends the aggregated information to the CLG node, and finally the CLG node sends the information to the Sink node to construct a three-layer cluster structure in Sect. 3.5. In Sect. 3.6, we design a communication scheme based on authenticated key agreement for two-node secret data transmission. Finally, a secure aggregation routing protocol with authentication and energy-saving functions is implemented.

3.1 Network Model

The sensor nodes monitor the physical and chemical reaction from various intelligent network environments, and then transmit these changes to the sink node (SN) for decision making and processing. The network model scenario studied in this paper is an ordinary Internet of things model, as shown in Fig. 2. The function of the realization is the process of the sensor node passing the collected information to the sink node.

3.2 Network Deployment and Initializing

At this stage, in order to monitor the various data, we begin with sensor deployment, where we deploy micro-sensor equipment over a two-dimensional target area, as shown in Fig. 2. These sensors have different properties and have different capacities, sizes and capabilities. All sensor nodes with heterogeneous properties are deployed randomly. They are stationary and do not move when deployed.

After deployment, each node determines its position according to its latitude and longitude, its energy status and its link status.

After the deployment and preparation work is completed, the pre-processing stage is entered, and the sensor node performs message broadcasting and message reply.

After the sensor nodes are deployed, the Sink node (SN) communicates with all the deployed nodes via a broadcast initiated greeting packet ($Init\_Hello_{Pkt}$). $Init\_Hello_{Pkt}$ contains information about the SN node, such as SN_{id} and its own location.

Fig. 2. General structure model of wireless sensor networks.

After receiving $Init\_Hello_{Pkt}$ from SN, each node calculates the distance to SN node dtance and packages its remaining energy R.E, link quality LQ into the $Reply_{pkt}$ for response. $Reply_{pkt}$ contains important information about all deployed nodes, such as N_{id}, residual energy R.E, link mass LQ, distance to SN node d_{tance}.

After receiving the $Reply_{pkt}$ from the node, SN responds with the confirmed packet (Ack_{Pkt}), which ensures that the message has been successfully received. Ack_{Pkt} contains the information of SN node, such as SN_{id} and its own location.

Therefore, at this stage, SN obtains important information of the deployed node (N_{id}, residual energy R.E, link quality LQ, distance to SN node d_{tance}).

3.3 Cost Function

After SN obtains the information, the cost function (C.F) for each deployed node is calculated based on the information obtained from the deployed node. Based on the results of C.F, the decisions of CLH, CLG and CN are made. Sink nodes identify CLH, CLG nodes, broadcast to all nodes, send a notification packet, notify all nodes which nodes are selected, and public at the Sink node. In addition, SN deploys a CN (Check-up Node) to monitor CLH and CLG. If the number of forwards for CLH and CLG does not exceed a fixed number of packets, CN broadcasts messages in the cluster to stop responding to them and blacklists the relevant CLH or CLG. CN node should also monitor the energy consumption of CLH and CLG nodes in real time. If these two nodes run out of energy, SN node will be notified to update CLH and CLG nodes dynamically to ensure the effective implementation of the protocol. CN is selected from CLH by Sink nodes based on the cost function C.F chooses CN according to the game

theory, compares two games in the selected CLH node, chooses the low cost as the CN node.

In addition, due to CLH, CLG nodes need to consume a lot of energy, in order to extend the service time of the nodes, we provide an energy collection module for CLH, CLG nodes, that allows them to harvest energy from their surroundings or from radio signals around them.

The cost function (C.F) is calculated in three parts, including the distance from each node to Sink node, the residual or total energy of each node, and the quality of each node in the overall network node (link quality).

3.4 Cluster Construction Based on Cost Function

According to the above four sections, Sink node calculates the cost function of all nodes based on the results of the preprocessing section. According to the cost function, CLH and CLG nodes are selected. Next we build the cluster. The Clustering Method is used to cluster all the data elements, and then the three-layer cluster routing protocol is formed, which can reduce the data transmission cost of large-scale WSN.

In order to classify the network into hierarchical clusters, we adopt the idea of K-means clustering algorithm, and take the distance as the standard of similarity measurement between data objects, that is, the smaller the distance between data objects, the higher the similarity between them, the more likely they are in the same cluster.

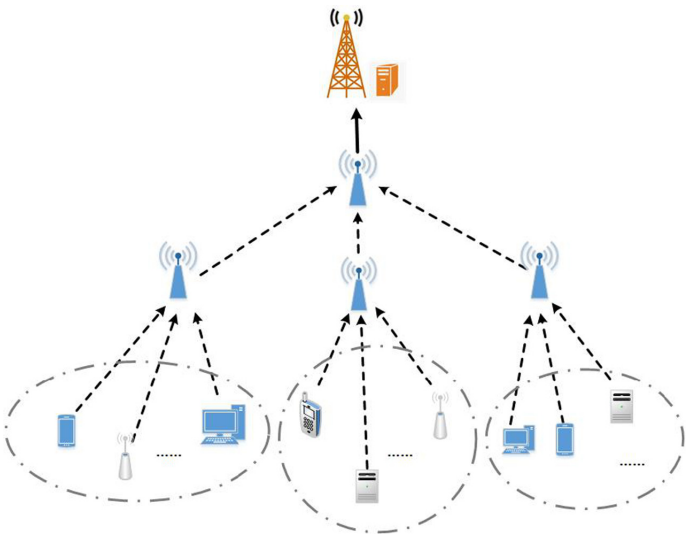

Fig. 3. Three-tier cluster structures.

3.5 Energy-Saving Aggregation Routing Protocol Based on Cost Function

Finally, our system model is constructed as shown in Fig. 3.

In this paper, we construct a three-layer cluster routing protocol. The sensor node wants to transmit data to the Sink node. The sensing data is forwarded by CM (Cluster Member) to CLH, which forwards it to CLG, and then the aggregated data is further forwarded to SN.

3.6 Anonymous Communication Scheme Based on Energy-Saving Routing Protocol

After clustering, the idea of on-demand routing discovery can be used when CM nodes want to send secret information to CLH nodes. When the CM node needs to send packets to the CLH node, it sends a routing request message RREQ to all its neighbors. When an intermediate node receives it, the intermediate node adds its address to the packet, sends it to its neighbor, and so on. When a request reaches the destination CLH node, it generates the route reply message RREP. The reply is unicast along the reverse path already built by the intermediate node to the CM source node. The CM node thus gets a multi-hop link to the CLH node, see Fig. 4.

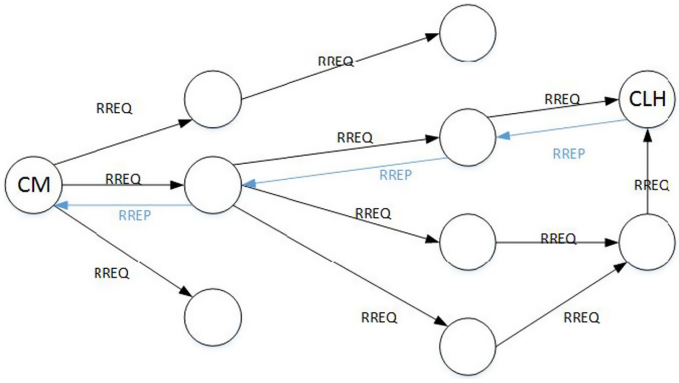

Fig. 4. The communication process between the CM node and the CLH node.

But when the two communicate, because the Internet is a public channel, malicious users can intercept ongoing communications and wreak havoc. In order to reduce such activities and ensure the security of the online communication, we design an identity authentication key protocol scheme to protect the privacy and security of the node communication. With this mechanism, two communication entities validate each other and then build a session key to protect future communication between them.

Registration of CLH. During this process, CLH_j node is registered in RN node. CLH_j sends a registration request and RN feeds back to CLH_j two keys, h(Q) and $h(HID_j||Q)$.

Figure 5 shows the detailed process of the CLH node registration phase.

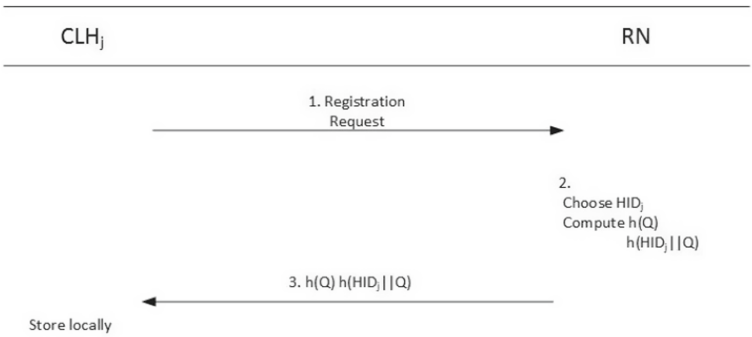

Fig. 5. Registration of CLH.

Registration of CM. This process is the CM_i node registration phase, in the Secure Channel, CM_i provides RN with its own identity ID_i and password PW_i. After receiving the command, RN will merge the two parts into a smart chip, and hide the $h(HID_j||Q)$ key generated by CLH_j node with CM_i identity information to generate A_{ij}, which will be stored in the chip and fed back to the CM_i node.

Figure 6 shows the detailed process of the CM node registration phase.

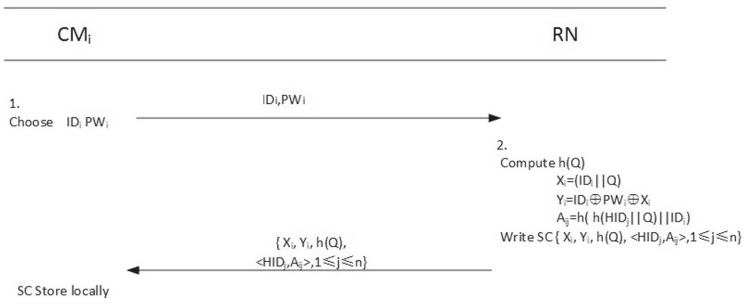

Fig. 6. Registration of CM.

Validation of CM. The process is to verify the validity of CM_i node with password for smart chip. Before the CM_i request begins communication, the smart chip first validates the CM_i node by providing its own $< ID_i^*, PW_i^* >$. The smart chip computes $X_i^* = Y_i \oplus ID_i^* \oplus PW_i^*$ and verifies $X_i^* \stackrel{?}{=} X_i$. If the condition is met, the CM_i node password and identity information is consistent with the registered node in RN, the node is valid and verified. If the condition is not met, the smart card rejects the CM_i node.

Key Agreement with Authentication. If the smart chip is verified, the smart chip provides the CM_i node with the A_{ij}, which contains the $h(HID_j||Q)$ key. The CM_i finds the CLH_j it wants to communicate with and encrypts a random number α_i. This ensures that only CLH_j can decipher the random number. At the same time, the identity information is sent to CLH_j together with the verification message V_1 generated by the random number. After receiving the message, CLH_j decrypts the random number α_i with its own key and verifies the correctness of the message V_1. If the verification is successful, a random number α_j is generated for the later session key SK calculation. Similarly, a validation message, V_2, will be passed to CM_i. CM_i carries out the identification of CLH_j and the calculation of session key SK, and finally the confirmation of SK.

Figure 7 shows the detailed process of authentication and key agreement between CM and CLH nodes.

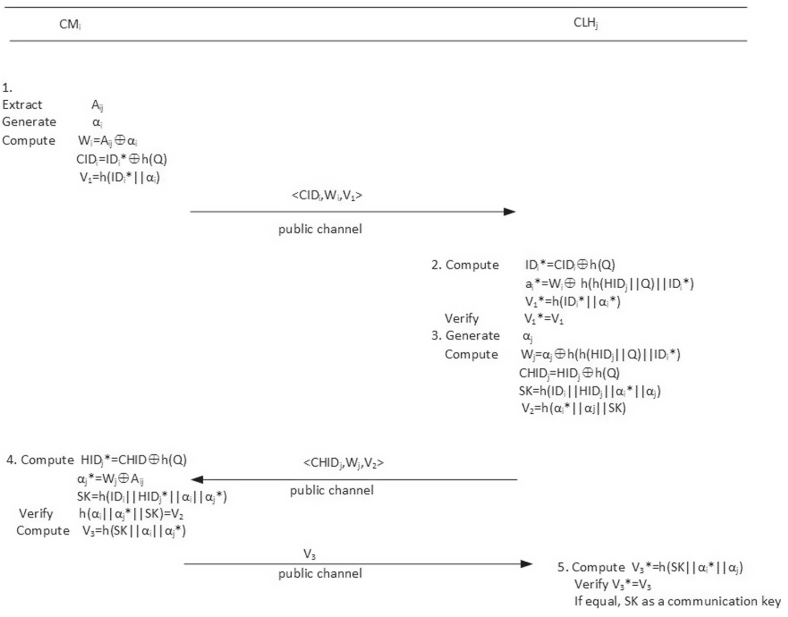

Fig. 7. Key agreement with authentication function.

4 Security Analysis

We assume that we have compromised a pair of cluster head node CLH_j^C and cluster member node CM_i^C. Using these leaked confidential information nodes, the enemy can not impersonate any other node, nor can it compute the session key of any other participant, as follows: Adversary A can get $\{h(Q), h(HID_j||Q)\}^c$ from cluster head, and adversary A can get HID_j from cluster member.

Can not impersonate any legitimate CLH node: In all nodes, $h(Q)$ is known, but $h(HID_j||Q)$ is different. To know $h(HID_j||Q)$, A either knows Q from RS (which is impossible), or it can reverse-deduce it from $A_{ij}^c = h(h(HID_j||Q)||ID_i{}^c)$ (which is also impossible).

Can not impersonate any legitimate CM node: To impersonate a member of the legal cluster, adversary A needs to calculate the A_{ij}. Suppose the adversary obtains $< CID_i, W_i, V_1 >$ from the landing phase, relevant to the calculation of A_{ij}, of these four parameters only W_i, $W_i = A_{ij} \oplus \alpha_i$. Because of the unavailability of α_i, A_{ij} is not computable.

Can not calculate any session key: The session key is calculated as $SK = h(ID_i||HID_j||\alpha_i||\alpha_j)$. A_{ij} is not obtainable, so α_i and α_j are not obtainable.

5 Performance Analysis

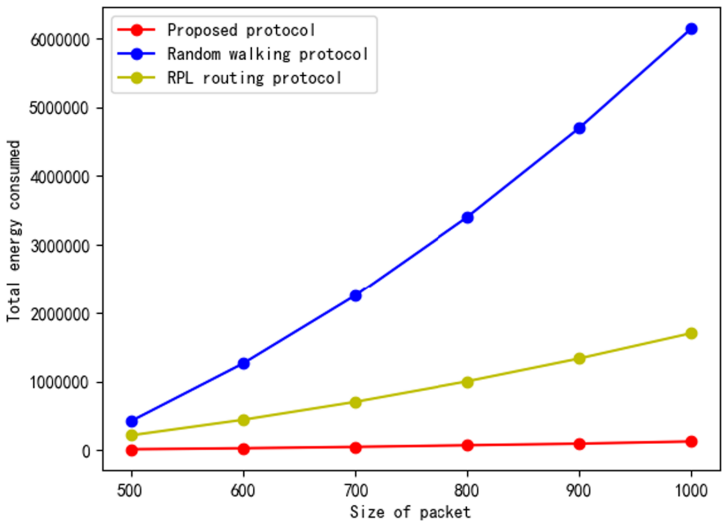

Fig. 8. Total energy consumed.

In this section, we analyze and compare this scheme with the other two schemes. We set all the member nodes at the bottom to send packets to the SN node, the

completion of the transmission is considered as the end of a round of data transmission. We calculate the total energy consumption of each node from the total energy consumption, according to the established routing protocol, calculate the total energy consumption of all member nodes to SN node.

Figure 8 shows the energy consumption of a round trip from all the member nodes we calculated to the sink node. As can be seen from Fig. 8, the difference in total energy consumption between protocols increases as the number of packets to be transmitted increases after the information is transmitted. When the packet reaches 1000 bits, as can be seen from the graph, the total energy of RPL is 200 times that of our proposed protocol, and the total energy consumption of RPL is 66 times that of our proposed protocol. To sum up, in terms of total energy consumption, our proposed protocol is lower than other protocols.

6 Conclusion

With the advent of the big data era, it is conceivable that the collection and analysis of data is very important. As a data collection device, wireless sensor network has greatly promoted the development of big data and data mining technology. This paper focuses on a secure convergent routing protocol with authentication and energy-saving functions. This paper proposes a solution to the problems of excessive energy consumption, unbalanced resource consumption and data privacy leakage in wireless sensor networks. The designed cost function selects a suitable "Leader" node of the data transmission cluster by calculating the communication loss, helps to plan the data transmission path, and solves the problem that some nodes consume too much energy due to the uneven resource consumption. The key agreement protocol is designed by using one-way Hasche function and XOR operation to hide and restore the parameters, using these parameters to calculate the shared key. The privacy protection of the secret data is realized, and the problem that the secret data is easy to be divulged when two nodes communicate is solved. The key agreement protocol with authentication function is designed. We authenticate first and then negotiate to prevent the third party from maliciously impersonating. Through the implementation of the above three parts, we construct a secure converging routing protocol with authentication and energy-saving functions. At last, through the performance analysis, the total energy consumption of the proposed protocol is compared with that of the two related protocols. Our work extends the network lifetime to some extent, increases the number of active nodes, reduces energy consumption, and increases the privacy of secret data transmission to some extent.

References

1. Ilyas, M., Ullah, Z., Khan, F.A., Chaudary, M.H., Durrani, H.: Trust-based energy-efficient routing protocol for internet of things-based sensor networks. Int. J. Distrib. Sens. Netw. **16**(10), 155014772096435 (2020)

2. Iuh, A., Jian, W.A., Yz, A., Sm, B.: An efficient hash-based authenticated key agreement scheme for multi-server architecture resilient to key compromise impersonation. Digital Commun. Networks **7**(1), 140–150 (2021)
3. Lekhwar, S., Yadav, S., Singh, A.: Big data analytics in retail (2019)
4. Mallick, A., Saha, A., Chowdhury, C., Chattopadhyay, S.: Energy efficient routing protocol for ambient assisted living environment. Wireless Personal Commun. **109**(3), 1333–1355 (2019)
5. Mishra, P.K., Verma, S.K.: Ffmcp: feed-forward multi-clustering protocol using fuzzy logic for wireless sensor networks (wsns). Energies **14**(10), 2866 (2021)
6. Moon, S.H., Park, S., Han, S.J.: Energy efficient data collection in sink-centric wireless sensor networks: a cluster-ring approach. Comput. Commun. **101**, 12–25 (2016)
7. Nashwan, S.: Aaa-wsn: anonymous access authentication scheme for wireless sensor networks in big data environment - sciencedirect. Egyptian Informatics Journal (2020)
8. Ullah, M.F., Imtiaz, J., Maqbool, K.: Enhanced three layer hybrid clustering mechanism for energy efficient routing in IoT. Sensors **19**(4) (2019)
9. Ullah, Z., Ahmed, I., Ali, T., Ahmad, N., Niaz, F., Cao, Y.: Robust and efficient energy harvested-aware routing protocol with clustering approach in body area networks. IEEE Access, p. 1 (2019)
10. Wei, L., Yuwang, Y.: Energy-efficient routing protocol in mobile ad hoc networks. Comput. Eng. Design **039**(010), 3013–3017 (2018)
11. Wenkang Zhou, X.W.: WSN routing algorithm based on AHP and FIS. In: Computer Engineering, pp. 1–11 (2020)