Link Aware Aggregation Query with Privacy-Preserving Capability in Wireless Sensor Networks

Yunfeng Cui[1,3], Wenbin Zhai[1,3], Liang Liu[1,3](✉), Youwei Ding[2], and Wanying Lu[1,3]

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
{cyfnuaa,wenbinzhai,liangliu,wanyinglu}@nuaa.edu.cn

[2] School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing, China
ywding@njucm.edu.cn

[3] Key Laboratory of Civil Aviation Intelligent Airport Theory and System, Civil Aviation University of China, Tianjin, China

Abstract. In wireless sensor networks (WSNs), users often submit spatial range queries to obtain statistical information of an area in the network, such as the average temperature and the maximum humidity of an area. The existing privacy-preserving aggregation query algorithms depend on pre-established network topology, and maintaining network topology requires lots of energy. In addition, these algorithms assume that the nodes between the communication radius can perform perfect communication, which is impractical. Aiming to solve these problems, this paper proposes a link aware aggregation query algorithm with privacy-preserving capability, that is, Reliable Spatial Range Data Aggregation Query with Privacy-Preserving (RPSAQ). RPSAQ first divides the query area into multiple sub-areas, and each sub-area is divided into multiple grids according to the network topology and link quality. Under the condition of ensuring node-perceived data privacy, RPSAQ collects sensing data of nodes by traversing the grids in the query area, which not only reduces the packet loss rate and energy consumption of sensor nodes, but also ensures the sensing data's privacy. The experiment results show that RPSAQ outperforms the existing privacy protection algorithms in terms of packet transmission, energy consumption and query result quality.

Keywords: Wireless sensor network · Spatial range aggregation query · Link quality · Privacy preserving

1 Introduction

Wireless sensor networks (WSNs) are data-centric, and users often submit spatial range queries to obtain statistical information of an area in the network,

J. Chen et al. (Eds.): EISA 2022, CCIS 1641, pp. 209–224, 2022.
https://doi.org/10.1007/978-3-031-23098-1_13

such as the average temperature and the maximum humidity of an area in a forest. Therefore, wireless sensor networks have broad applications in the fields of national defense, military, medical and environmental monitoring.

Sensor nodes are battery powered, and the energy of the battery is limited. Also, in many cases, battery replacement is difficult. Many studies indicate that the energy consumed by sensor nodes is mainly used in data transmissions. Therefore, in order to extend the service life of WSNs, the number of data transmissions needs to be reduced. In addition, due to open deployment and wireless communication, the sensing data of nodes is at risk of being captured. In the application scenario with information confidentiality requirements, the security of the sensor network information needed to be improved. Therefore, it is necessary to study the energy-efficient and privacy-preserving spatial range aggregation query processing technique to solve these two problems.

The existing spatial range aggregation query algorithms have been proposed, which can be divided into two categories according to the topology they depend on: cluster-based and tree-based. However, these algorithms are susceptible to node movements, node failures, and the surrounding environment. Moreover, maintaining network topology incurs the energy consumption of network infrastructure due to the frequent changes of network topology.

Aiming to solve these problems, this paper proposes a link aware aggregation query algorithm with privacy-preserving capability called RPSAQ. RPSAQ is divided into three stages. First, the query message and a random number randomly generated are sent to a node in the query area using geographic routing protocol [11]. Then, the node is used as a starting node, and the query area is divided into several sub-areas. Each sub-area is dynamically divided into multiple grids according to the network topology and link quality. A query node is elected in each grid, which is responsible for sending query message to all nodes in the next grid. The nodes in the grid use the point-by-point overlay strategy to aggregate the sensing data according to the established route, and transmit the partial query result after aggregation to the node in the next grid. This above process is repeated until all nodes in the query area are accessed and the final query result is thus generated. Finally, the final aggregation result will be transferred back to the sink using geographic routing protocol.

According to the real-time network topology and link quality, RPSAQ dynamically divides the grid and selects a query node for each grid. Therefore, the impact of network topology changes on the quality of query results is avoided. Based on link quality information, RPSAQ selects communication links with low packet loss rate to distribute query information, and aggregate sensing data, which avoids multiple retransmissions of data packets and reduces energy consumption. The experiment results show that RPSAQ outperforms the existing algorithms in terms of energy consumption, the number of packet transmissions and query result quality. The main contributions of this paper are as follows:

1. We propose a route-based and infrastructure-free dynamic data collection protocol, which does not depend on the pre-established topology.

2. The method proposed in this paper only needs the nodes in the query area to participate in query processing, and does not require all nodes in the network, thus saving energy consumption.
3. The method proposed in this paper provides efficient data aggregation and takes link quality into account.

The organization of this paper is as follows. Section 2 summarizes the related work, and Sect. 3 introduces the preliminaries. Section 4 proposes a link aware aggregation query algorithm with privacy-preserving capability. An analysis of our work and experimental results is presented in Sect. 5. Section 6 summarizes the paper and presents future research directions.

2 Related Works

The window aggregation query has been extensively studied, and the existing algorithms can be divided into two categories. (1) Tree-based algorithms [2,4,5,8]: these algorithms rely on preconfigured topology and assume that communication between nodes is safe. (2) Route-based algorithms [7,22,23,25]: the query route of these algorithms is dynamically generated in the query process, which reduces the impact of network topology changes on query processing, but they assume that the communication model of the node is an ideal disk graph.

Because WSNs have the characteristics of self-organization and multi-hop, the extensive application of wireless sensor faces serious data leakage problems. Limited resources also bring a series of challenges for the extensive application of wireless sensor network. Data aggregation is an important way to reduce energy consumption. In recent years, many secure data aggregation schemes have been proposed.

Some existing schemes [9,19–21,24] take their base station (BS) as the root node and organize their nodes into a tree structure. In [9], He et al. proposed data aggregation privacy protection technology called SMART based on data fragmentation. Each node divides its sensing data into several fragments to hide its original sensing data and sends the data fragments to different intermediate nodes. After distributing the data fragmentation, the final aggregation result is finally derived at the base station. Considering that the data sent by the non-leaf node to its parent node is the result of the aggregation of the sub-tree where it is rooted rather than the original data, there is no need to protect the privacy of the data sent by the non-leaf node. On the basis of SMART, Wang et al. [21] proposed a method for fragmenting the sensing data of the leaf nodes called PECDA. The data fragments of the leaf nodes are sent to the neighbor nodes through the secure channel to protect the privacy of the leaf nodes.

In order to avoid data loss, some schemes [16,26] adopt a ring topology or a layered model, and group nodes into one layer or in the same ring according to the number of hops of the node to the BS. By grouping the nodes of layer x, their grouping is then sent to any node of layer $x - 1$ in their transport range. Therefore, there are multiple parent nodes between nodes and BS, and there are multiple paths between nodes and BS.

In addition, encryption strategy is often used in wireless sensor network to protect the sensing data. In order to protect the privacy of the original data, the data is encrypted and then transmitted to the next hop node [13]. After arriving at the next hop node, the encrypted data will be decrypted and then the aggregation operation will be completed with the sensing data of the next hop node. This process is repeated until it reaches BS. Since the encrypted data will be decrypted in each hop, the intermediate node can easily get the original sensing data, and there is a security threat in the way of hopping encryption. In order to overcome this shortcoming, some schemes [1,12,14,15,18,20,27] proposed that aggregation nodes use homomorphic encryption strategy to directly aggregate ciphertext data, and other nodes cannot decrypt in the transmission process. And all nodes in the sensor network encrypt the sensing data using the secret key shared with the sink node.

The privacy protection strategy in the case of tree-based data aggregation relies on the constructed distribution routing tree. When the node moves, the network topology changes frequently, which lead to an increase in the cost of maintaining the routing tree. The application of encryption strategy reduces the dependence on topology and increases the resource consumption. Both assume that the query area is a full network.

Both tree-based and itinerary-based spatial range queries assume that communication between nodes is secure. The tree-based query scheme relies on the constructed topology. Although the itinerary-based query scheme avoids the dependence on the topology, the algorithm assumes that the communication model of nodes is an ideal disk graph.

Aiming to solve these problems, this paper proposes a link aware aggregation query algorithm with privacy-preserving capability called RPSAQ. According to the real-time network topology and link quality, RPSAQ dynamically generates query routes. By using link quality information, links with low packet loss rate are selected to distribute query information and aggregate perceived data, which avoids multiple retransmissions of data packets and reduces energy consumption.

3 Preliminaries

3.1 IWQE Protocol

Itinerary-based Window Query Execution (IWQE) [25] is an itinerary-based spatial range aggregation query processing algorithm. As shown in Fig. 1, IWQE divides the entire query area $ABCD$ into two sub-areas $AEFD$ and $EBCF$. And an itinerary is built to traverse all the nodes in these sub-areas. The nodes on the query itinerary are query nodes (such as $S_1, S_2, \cdots, S_9$). Each query node is responsible for collecting the sensing data of its neighbors, and sending query messages and partial query results to the next query node. Nodes other than query nodes in the query area are data nodes (such as a and b) Each data node is responsible for sending its local sensing data to their query nodes (such as the temperature and humidity of the query area). The detailed process of IWQE is as follows:

1. Once receiving a query request submitted by a user, the sink S sends the query message Q_m to the first query node S_1 within the query area using geographic routing protocol [11].
2. Node S_1 broadcasts the query message Q_m to its neighbors after receiving it.
3. The neighbors of node S_1 in the query area send their sensing data to S_1 successively after hearing the query message.
4. After receiving all the sensor data of its neighbors in the query area, node S_1 aggregates them with its local data to calculate partial query result, which will be sent to the next query node of S_1. And S_1 selects S_2 from its neighbors to be the next query node.
5. The process is repeated by other query nodes in the query area. When it stops at node S_9 which is the last query node, S_9 will calculate the final query result and return it back to the sink using geographic routing protocol.

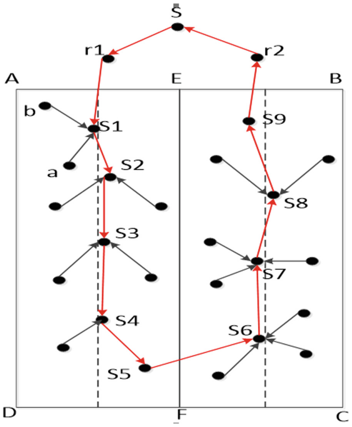
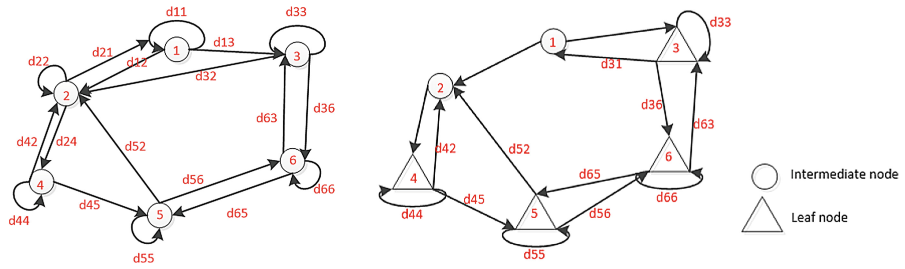


Fig. 1. The IWQE.

IWQE dynamically generates query itinerary according to the real-time network topology, which weakens the impact of network topology changes on query processing. However, IWQE assumes that wireless communication between nodes is the ideal disk model, which is impractical. And this assumption brings several problems: (1) The communication cost between query node S_i and its next query node S_{i+1} is high. As shown in Fig. 1, the node S_1 chooses S_2 from its neighbors as the next query node, which is the furthest one in the vertical direction. In the real network, the greater the distance between nodes, the poorer the link quality. Therefore, in order to send partial query results and query messages to S_2, node S_1 needs continuous retransmissions due to the lossy communication link between node S_1 and S_2, which brings a large amount of energy consumption. (2) Collecting the sensing data of neighbor nodes consumes a large amount of energy. The distance between node a and node b is denoted as $Dist(a, b)$ and



(a) (b)

Fig. 2. The process of the improved algorithms & the process of SMART

the probability that node a successfully transmits messages to node b is denoted as $P(a, b)$. In Fig. 1, data node a is in the communication range of both S_1 and S_2. And $Dist(a, S_1) \gg Dist(a, S_2)$, $P(a, S_1) \ll P(a, S_2)$. In IWQE, the sensing data of node a is collected by query node S_1. However, if node S_2 collects the sensing data of node a, the energy consumption can be significantly reduced.

3.2 SMART Protocol

In [9], He et al. proposed the privacy protection protocol SMART based on data slicing. Figure 2(a) shows the process of SMART, which can be divided into three stages: data slicing, data mixing, and data aggregation. Each node slices its own data into three pieces, keep a piece for itself, and then transmits the rest two pieces to its neighbor nodes. After all pieces are received, each node mixes its own piece and the received pieces to obtain a new result.

On the basis of SMART, [6,10,28] proposed some improved sensor network privacy protection protocols. Figure 2(b) shows the process of these improved algorithms. They do not need to slice all the nodes in the query region. Only the leaf nodes fragment their data in the data slicing stage. As shown in Fig. 2(b), the leaf node n_3 slices its own data into three pieces: d_{31}, d_{33} and d_{36}. Node n_3 keeps d_{33} for itself, and transmits d_{31} and d_{31} to n_1 and n_6 respectively.

Above privacy-preserving data aggregation algorithms can aggregate sensing data to obtain the desired query results without compromising the privacy of sensing data. However, these algorithms all rely on pre-established topology trees. Because the corresponding topology will change frequently due to node failures and node movements, maintenance of the topology will incur additional energy consumption.

4 RPSAQ

4.1 Basic Idea

Suppose all sensor nodes can get their own locations through localization algorithm, and they broadcast their location information regularly. Therefore, all nodes can get the location information of their neighbor nodes. And each node uses link estimation algorithm to calculate the link quality between the node and all its neighbor nodes.

As shown in Fig. 3, RPSAQ divides the query area into several grids, and selects a query node in each grid, which is responsible for broadcasting query messages to all nodes in the next grid. Based on the link quality between nodes, RPSAQ sets the grid size and the query node reasonably to reduce the packet loss rate and the energy consumption of distributing the query message, and ensures the quality of the query results. The query processing of RPSAQ can be divided into three stages:

1. The query message and a random number randomly generated are sent to a node in the query area using geographic routing protocol which takes link quality into account.
2. The query message is sent to all grids in the query region, and the sensing data of all nodes in the query region is collected and aggregated to generate the final aggregation result.
3. The final query results generated in the second stage are returned to the sink through the geographic routing protocol which takes link quality into account.

The detailed process of RPSAQ is shown in Fig. 3. It is assumed that RPSAQ is used to calculate the average temperature of area Q_a. Without loss of generality, the query area is a rectangle. After receiving the query request from the user, the sink sends the query message Q_m and the random number γ to the query area Q_a through multiple relay nodes by using geographic routing protocol which takes link quality into account [17]. The last relay node R_1 selects node q_1 from its neighbor nodes as the first query node of the query region Q_a. After Q_m and γ arrive at node q_1, γ is added to q_1's sensing data D_{q_1} to obtain $D^{'}_{q_1}$. Then node q_1 sets the size of grid g_2 and selects the query node q_2 in grid g_2 (referring to section $IV.B$). Also, node q_1 sends the query message Q_m to all nodes in grid g_2 and select node a from its neighbor nodes as the starting node of grid g_2. After receiving the query message Q_m and partial aggregation result $D^{'}_{q_1}$, node a adds $D^{'}_{q_1}$ to its sensing data D_a to obtain $D^{'}_a$ and sends partial aggregation result $D^{'}_a$ to the next node in grid g_2. This process is repeated until node c's partial aggregation result $D^{'}_c$ arrives at the query node q_2 in grid g_2. Then query node q_2 add $D^{'}_c$ to its sensing data $D^{'}_{q_2}$ to obtain partial aggregation result $D^{'}_{q_2}$, and sets the size of grid g_3 and selects the query node q_3 in the grid g_3. Also, node q_2 sends the query message Q_m to all nodes in grid g_3 and select node e from its neighbor nodes as the starting node of grid g_3. After receiving the query message Q_m and partial aggregation result $D^{'}_{q_2}$, node e repeats what node a does in grid g_2 until partial aggregation result reaches the query node q_3. This process is repeated until the last query node q_n in the query area Q_a obtains the final aggregated result $D^{'}_{q_n}$. Finally, node q_n sends $D^{'}_{q_n}$ back to the sink via the geographic routing protocol. The sink subtracts the random number γ from the returned aggregation results $D^{'}_{q_n}$ to obtain the real aggregation results in the query region.

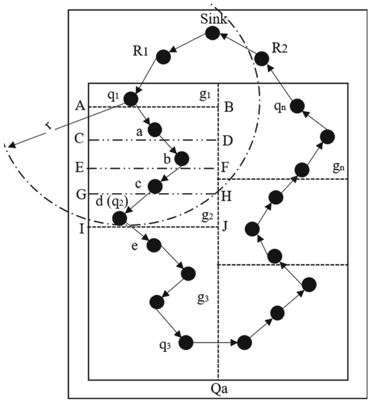


Fig. 3. The execution procedure of the RPSAQ algorithm

The entire dynamic query process of RPSAQ does not depend on the preconstructed topology, which aggregates the sensing data while ensuring the privacy of the nodes' sensing data. Taking node q_0 as an example, the data sent by node q_0 to node a is $D^{'}_{q_1}$, and node a cannot infer the sensing data of node q_0 because of random number γ.

4.2 How to Set the Grid Size and the Query Node

Suppose RPSAQ divides the query area into n grids $g_i(i \in [1, n])$, the set of nodes in grid g_i is denoted as $NS_i(i \in [1, n])$, the corresponding query node in g_i is $q_i(i \in [1, n])$, the number of nodes in g_i is represented by $|NS_i|$. The query node $q_{i-1}(i \in [2, n])$ selects its neighbor nodes along the data transmission direction as next potential query nodes for grid g_i and the set of potential next query nodes is denoted as ξ_i. For example, in Fig. 3, node a, b, c, d, R_1 and the sink are neighbor nodes of the query node q_1 and node a, b, c and d are along the transmission direction of q_1's sensing data. Therefore, the set of q_1's potential next query nodes is $\xi_2 = \{\text{a}, b, c, d\}$. In addition, to ensure normal communication between neighbor nodes, we define as follows: $Dist(q_{i-1}, q_i) \leq r, \forall i \in [2, n]$, which refers to that the distance between each query node q_{i-1} and the next query node q_i is less than or equal to the maximum communication radius of the node r.

Based on the constructed link model, this paper proposes a distributed heuristic Grid Setup Algorithm (GSA) to set the grid size and select the query node for each grid. In GSA, we define the average energy consumption to aggregate all sensing data in grid g_i to the next query node q_i as follows:

$$F(g_i, q_i) = \frac{E_c(g_i, q_i)}{|NS_i|} \tag{1}$$

$E_c(g_i, q_i)$ refers to the total energy consumption to aggregate all sensing data in grid g_i to the next query node q_i, which includes two parts: (1) the total energy consumption that each node in grid g_i receives the query message Q_m

broadcasted by the query node q_{i-1}; (2) the total energy consumption that each node in grid g_i calculate partial aggregation result and sends it to the next node.

In GSA, the query node q_{i-1} constructs the set of potential next query nodes $\xi_i = \{\alpha_j\}, j \in [1, \gamma]$. If node α_j is selected as the next query node by node q_{i-1}, then the corresponding grid g_{α_j} is also determined. Based on $F(g_i, q_i)$ and ξ_i, the query node q_{i-1} respectively evaluates the average energy consumption to aggregate all sensing data in grid g_i to each potential query node α_j. If min $F_i(g_{\alpha_j}, \alpha_j), \alpha_j \in \xi_i$, then the query node q_{i-1} will select node α_j as the next query node q_i and its corresponding grid is g_{q_i}. Take Fig. 3 as an example, the query node q_1 constructs the set of potential next query nodes $\xi_2 = \{a, b, c, d\}$. And the grids corresponding to node a, b, c and d are $g_a = ABCD$, $g_b = ABEF$, $g_c = ABGH$ and $g_d = ABIJ$ respectively. Then based on $F(g_i, q_i)$ and ξ_2, the query node q_1 respectively evaluates the average energy consumption to aggregate all sensing data in grid g_2 to each node in ξ_2. If $F(g_d, d) < F(g_i, i), \forall i \in [a, b, c]$, then q_1 will select node d as the next query node q_2 and its corresponding grid $g_2 = ABIJ$.

4.3 Method of Handing Voids

As shown in Fig. 4, after the query node a collects the sensing data of the node in the grid $ABCD$, it needs to send the query message Q_m and partial query results $D_a^{'}$ to the nodes in region $EFGH$. As the region $CDEF$ under the grid $ABCD$ does not have the neighbor nodes of node a, $CDEF$ becomes a void region, which leads to the interruption of query processing. To avoid this problem, a geographic routing protocol is used to bypass the void region.

As shown in Fig. 4, the query node of grid $ABCD$ is a. The region $CDEF$ is characterized as a void region because a has no neighbors in $CDEF$, which will interrupt RPSAQ algorithm at node a. In order to address this issue, after data collection is finished in grid $ABCD$, its query node a will take the center of rectangular region $EFGH$ as the target location (where the distance between E and G) and use the geographic routing protocol to send the partial result and the query message to a node in $EFGH$.

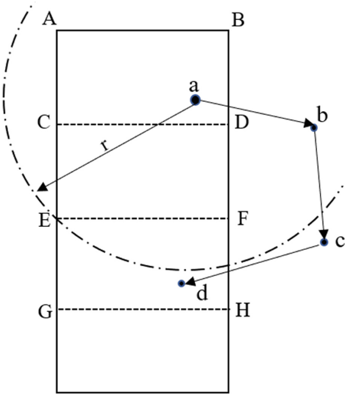


Fig. 4. Bypassing the void region

4.4 Privacy Analysis

The existing spatial range aggregation query algorithms usually utilize key mechanism to protect the privacy of sensing data. According to Eschenauer and Gligor [6], assume there are K secret keys in the secret key pool, and each sensor node in the sensor network randomly selects k secret keys from the pool. The probability of at least one identical key between any two neighbors is defined as $P_{connect} = 1 - \frac{((K-k)!)^2}{(K-2k)!2K!}$ and the probability that the third party has the same key is defined as $P_{overhear} = \frac{k}{K}$.

In the SMART protocol, when establishing a secure connection, node a and b can establish a secure connection because a and b have the same secret key value d_{ij}. The privacy leakage of sensing data mainly includes two aspects: (1) The third-party node has the same secret key with the probability $P_{overhear}$; (2) During the secret key preallocation stage, each node randomly selects k secret keys from the secret key pool with K secret keys and then establishes a connection if the adjacent nodes have the same secret keys. The third party can guess the secret key, and the correct guess is $\frac{1}{C_K^k}$. Therefore, SMART and PECDA have the possibility of privacy data leakage during data aggregation. In RPSAQ, if an attacker wants to get the original data of node a, it needs to get the aggregation result sent to the node a and the data transmitted to a's next node after node a completes the aggregation. However, RPSAQ adopts a route-based dynamic aggregation method, which increases the difficulty of the attacker guessing the next routing node and avoids the problem of sensing data leakage caused by secret key leakage in SMART and PECDA.

5 Analysis of Experimental Results

This section analyzes our proposed RPSAQ through simulation experiments. RPSAQ, SMART and PECDA are implemented in the simulator [3]. The geographic routing protocol which takes link quality into account [17] is used to

Table 1. Link model parameters

Parameter	Value
p	1
f	50 byte
P_t	0 dBm
d_0	1 m
$PL(d_0)$	55 dBm
η	4
σ	4
P_n	−105 dBm

Table 2. Simulation parameters

Parameter	Value
Area covered	100 m × 100 m
Node communication radius	11 m
Number of nodes	640
Sensing data size	110
Query data size	22

send the query message to the query area and return the final query result to the query origination node.

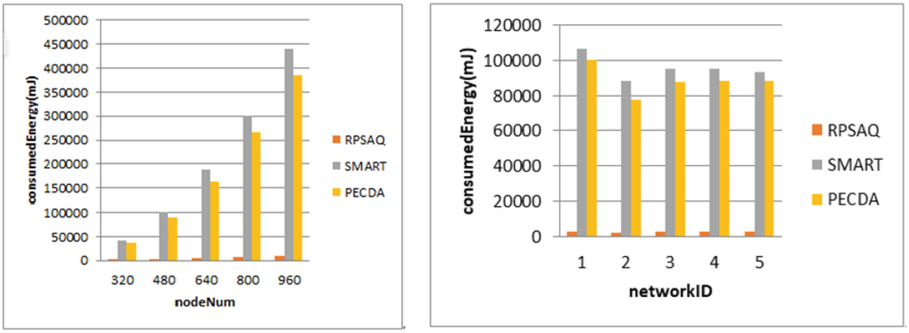
According to the literature [29], the typical *MICA2* mote satisfies the following log normal path loss model.

$$
\begin{aligned}
ppr(d) &= \left(1 - \frac{1}{2} e^{-10\left(\frac{r(d)}{10}\right) \times \frac{1}{1.28}}\right)^{p \times 8f} \\
r(d) &= P_t - PL(d_0) - 10\eta \lg\left(\frac{d}{d_0}\right) + N(0,\sigma) - P_n
\end{aligned}
\tag{2}
$$

where $ppr(d)$ means the proportion of successful packet transmissions when the distance between the sending node and the receiving node is d. p denotes the coding rate and f is the size of the data frame. r(d) denotes the *SNR*(signal to noise ratio) of the receiving node when the distance between the sending node and the receiving node is d. P_t is the transmit power of the sending node. $PL(d_0)$ represents the power loss for the referenced distance d_0. η is the path loss exponent. $N(0,\sigma)$ represents the Normal Random Variable with expectation 0 and variance σ. P_n is the noise floor. The parameters of *LNM* link quality model in our experiments are shown in Table 1. The link qualities between nodes in our experiments are all generated by *LNM* model. In addition, Table 2 summarizes the default parameters used in our simulations.

5.1 Energy Consumption

This set of experiments analyzes the effects of node density and different topologies on the energy consumption of RPSAQ, SMART and PECDA. Since the link quality is not taken into account in SMART and PECDA, the data sent by the data node to the query node is not retransmitted when the sensing data is lost. In order to make there algorithms comparable, a retransmission mechanism is introduced in SMART and PECDA.

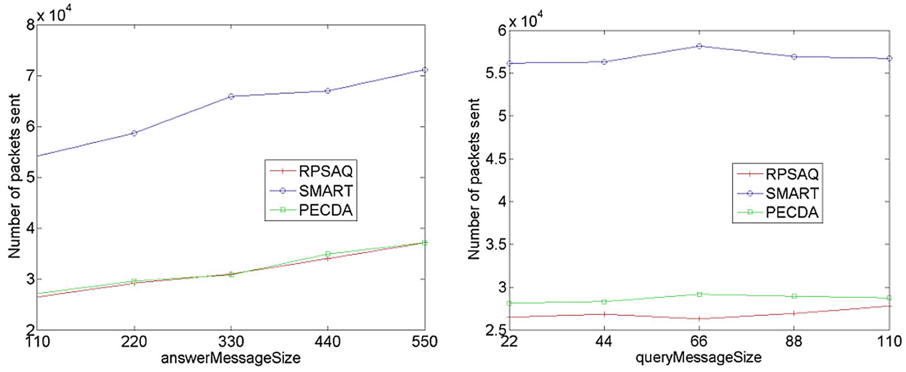


(a) (b)

Fig. 5. Influence of the number of network nodes on energy consumption & influence of different network nodes on energy consumption

The effect of the number of network nodes on energy consumption is shown in Fig. 5(a). It can be seen that as the number of nodes in the network becomes larger, PECDA and SMART consume significantly more energy than RPSAQ. This is because SMART and PECDA are based on data slicing to protect data privacy. In SMART, each node slices its own data into three pieces, keeps a piece for itself, and then transmits the rest two pieces to its neighbor node. The energy consumed in slice data distribution takes up a large proportion of the total energy consumption. In PECDA, only the leaf nodes fragment their sensing data in the data slicing stage. However, distributing slice data also consumes energy to some extent. Moreover, RPSAQ sets the grid size and selects the query node according to the link quality in the network, which reduces the number of packet retransmission and thus saves energy.

The effect of different network topologies on energy consumption is shown in Fig. 5(b). It can be seen from Fig. 5(b) that the change of the network topology has little effect on RPSAQ and the energy consumption of RPSAQ remains low. However, the energy consumption of SMART and PECDA remain high in all cases. This is because SMART and PECDA rely on pre-constructed topologies and maintaining the topology results in a large amount of energy consumption. By contrast, route-based RPSAQ outperforms topology-dependent SMART and PECDA in terms of energy consumption.



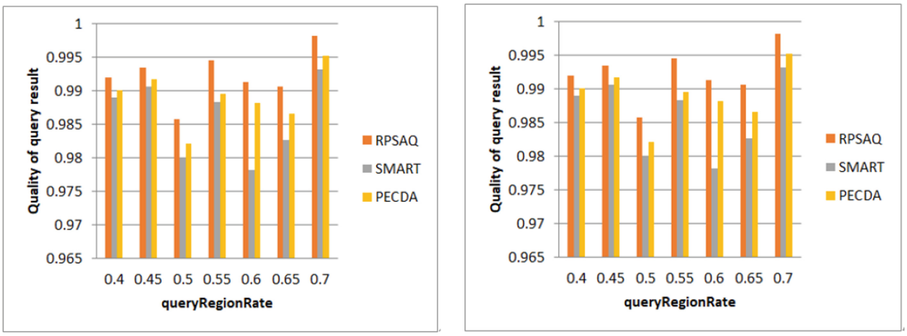
(a) (b)

Fig. 6. Influence of sensing data size on the number of packets sent & influence of query message size on the number of packets sent

5.2 Number of Packets

This set of experiments analyzes the impact of sensing data size and query message size on the number of transmitted packets. Figure 6(a) shows the effects of different sensing data sizes on the number of packets that need to be sent. We can observe that as the size of sensing data increases, the number of packets to be transmitted by SMART, PECDA and RPSAQ increases. This is because the larger the size of sensing data, the more packets the sensing data is divided into, and the more packets are transmitted in the network. Since the RPSAQ dynamically selects the travel route according to the link quality, the nodes along the query route remain a low packet loss rate, which reduces the number of packet retransmissions. However, SMART and PECDA do not take the link quality into consideration, which increases the number of packet retransmission. Also, SMART and PECDA are based on data slicing to protect data privacy and data slice transmission also increases the number of packets transmitted in the network.

Figure 6(b) shows the effect of different query message sizes on the number of packets that need to be sent. It can be seen from Fig. 6(b) that SMART and PECDA need to send more packets than RPSAQ. On the one hand, because the link quality is not taken into account in SMART and PECDA, the retransmission of the packets increases the number of packets sent. On the other hand, as SMART and PECDA include three stages: slicing, mixing and aggregation, the transmission of data fragments also increase the number of packets that need to be sent.



(a) (b)

Fig. 7. Influence of query area size on the quality of query result & influence of node number on the quality of query result

5.3 Query Result Quality

This set of experiments analyzes the query result quality of SMART, PECDA and RPSAQ under different query area size. The query result quality is defined as $sr = \frac{cn}{tn}$, where cn represents the number of nodes traversed in the query area, and tn represents the total number of nodes in the query area. Figure 7(a) shows the impact of query area size on query result quality. It can be seen that when the size of query area is small, the query results of SMART, PECDA and RPSAQ are of high quality. With the increase of the size of the query area, the query result quality of SMART and PECDA decreases significantly, while the query result quality of RPSAQ remains high in all cases. This is because when the size of query area is large, Fig. 7(a) shows the impact of query area size on query result quality. It can be seen that when the query area is small, the number of nodes traversed is small, and the query results of the three algorithms are of high quality. With the increase of query area, the route with high link quality selected by RPSAQ performs query processing, while the link loss rate of SMART and PECDA is larger, so the quality of query results decreases.

6 Conclusion

The existing privacy-preserving aggregation query processing methods in sensor networks rely on pre-established network topology. Maintaining the topology results in a large amount of energy overhead. In addition, the existing privacy protection algorithms assume that the communication model between nodes is ideal, that is, the nodes within the communication radius can perform perfect communication, which is impractical. Aiming to solve these problems, this paper proposes RPSAQ, a link aware aggregation query algorithm with privacy-preserving capability. RPSAQ does not depend on the pre-constructed topology structure, and dynamically divides the query area into several grids according to the link quality, and sequentially traverses and collects the partial aggregation result of the nodes in the grid. Our proposed algorithm not only reduces the packet loss rate of nodes, but also ensures the data privacy of the sensor nodes. The experimental results show that RPSAQ outperforms the existing privacy protection algorithms in terms of energy consumption, packet transmission and query result quality.

Acknowledgments. This work was supported by the Open Fund of Key Laboratory of Civil Aviation Intelligent Airport Theory and System, Civil Aviation University of China under no. SATS202206.

References

1. Boudia, O.R.M., Senouci, S.M., Feham, M.: A novel secure aggregation scheme for wireless sensor networks using stateful public key cryptography. Ad Hoc Netw. **32**, 98–113 (2015)
2. Coman, A., Nascimento, M.A., Sander, J.: A framework for spatio-temporal query processing over wireless sensor networks. In: Proceedings of the 1st International Workshop on Data Management for Sensor Networks: in Conjunction with VLDB 2004, pp. 104–110 (2004)
3. Coman, A., Sander, J., Nascimento, M.A.: Adaptive processing of historical spatial range queries in peer-to-peer sensor networks. Distrib. Parallel Databases **22**(2), 133–163 (2007). https://doi.org/10.1007/s10619-007-7018-8
4. Demirbas, M., Ferhatosmanoglu, H.: Peer-to-peer spatial queries in sensor networks. In: Proceedings Third International Conference on Peer-to-Peer Computing (P2P2003), pp. 32–39. IEEE (2003)
5. Deshpande, A., Guestrin, C., Hong, W., Madden, S.: Exploiting correlated attributes in acquisitional query processing. In: 21st International Conference on Data Engineering (ICDE 2005), pp. 143–154. IEEE (2005)
6. Eschenauer, L., Gligor, V.D.: A key-management scheme for distributed sensor networks. In: Proceedings of the 9th ACM Conference on Computer and Communications Security, pp. 41–47 (2002)
7. Fu, T.Y., Peng, W.C., Lee, W.C.: Parallelizing itinerary-based KNN query processing in wireless sensor networks. IEEE Trans. Knowl. Data Eng. **22**(5), 711–729 (2009)
8. Goldin, D., Song, M., Kutlu, A., Gao, H., Dave, H.: Georouting and delta-gathering: efficient data propagation techniques for geosensor networks, pp. 73–95. GeoSensor Networks, Boca Raton (2005)
9. He, W., Liu, X., Nguyen, H., Nahrstedt, K., Abdelzaher, T.: PDA: privacy-preserving data aggregation in wireless sensor networks. In: IEEE INFOCOM 2007–26th IEEE International Conference on Computer Communications, pp. 2045–2053. IEEE (2007)
10. Hu, S., Liu, L., Fang, L., Zhou, F., Ye, R.: A novel energy-efficient and privacy-preserving data aggregation for WSNs. IEEE Access **8**, 802–813 (2019)
11. Karp, B., Kung, H.T.: GPSR: Greedy perimeter stateless routing for wireless networks. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 243–254 (2000)
12. Lin, Y.H., Chang, S.Y., Sun, H.M.: CDAMA: concealed data aggregation scheme for multiple applications in wireless sensor networks. IEEE Trans. Knowl. Data Eng. **25**(7), 1471–1483 (2012)

13. Ozdemir, S., Çam, H.: Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks. IEEE/ACM Trans. Netw. **18**(3), 736–749 (2009)
14. Ozdemir, S., Xiao, Y.: Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. Comput. Netw. **55**(8), 1735–1746 (2011)
15. Prathima, E., Prakash, T.S., Venugopal, K., Iyengar, S., Patnaik, L.: SDAMQ: secure data aggregation for multiple queries in wireless sensor networks. Procedia Comput. Sci. **89**, 283–292 (2016)
16. Roy, S., Conti, M., Setia, S., Jajodia, S.: Secure data aggregation in wireless sensor networks: filtering out the attacker's impact. IEEE Trans. Inf. Forensics Secur. **9**(4), 681–694 (2014)
17. Seada, K., Zuniga, M., Helmy, A., Krishnamachari, B.: Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks. In: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, pp. 108–121 (2004)
18. Shim, K.A., Park, C.M.: A secure data aggregation scheme based on appropriate cryptographic primitives in heterogeneous wireless sensor networks. IEEE Trans. Parallel Distrib. Syst. **26**(8), 2128–2139 (2014)
19. Singh, V.K., Verma, S., Kumar, M.: Privacy preserving in-network aggregation in wireless sensor networks. Procedia Comput. Sci. **94**, 216–223 (2016)
20. Viejo, A., Wu, Q., Domingo-Ferrer, J.: Asymmetric homomorphisms for secure aggregation in heterogeneous scenarios. Inf. Fusion **13**(4), 285–295 (2012)
21. Wang, T., Qin, X., Ding, Y., Liu, L., Luo, Y.: Privacy-preserving and energy-efficient continuous data aggregation algorithm in wireless sensor networks. Wirel. Pers. Commun. **98**(1), 665–684 (2017). https://doi.org/10.1007/s11277-017-4889-5
22. Wu, S.H., Chuang, K.T., Chen, C.M., Chen, M.S.: DIKNN: an itinerary-based KNN query processing algorithm for mobile sensor networks. In: 2007 IEEE 23rd International Conference on Data Engineering, pp. 456–465. IEEE (2007)
23. Wu, S.H., Chuang, K.T., Chen, C.M., Chen, M.S.: Toward the optimal itinerary-based KNN query processing in mobile sensor networks. IEEE Trans. Knowl. Data Eng. **20**(12), 1655–1668 (2008)
24. Xie, K., et al.: An efficient privacy-preserving compressive data gathering scheme in WSNs. Inf. Sci. **390**, 82–94 (2017)
25. Xu, Y., Lee, W.C., Xu, J., Mitchell, G.: ProcessingWindow queries in wireless sensor networks. In: 22nd International Conference on Data Engineering (ICDE 2006), p. 70. IEEE (2006)
26. Zhang, K., Han, Q., Cai, Z., Yin, G.: RiPPAS: a ring-based privacy-preserving aggregation scheme in wireless sensor networks. Sensors **17**(2), 300 (2017)
27. Zhao, X., Zhu, J., Liang, X., Jiang, S., Chen, Q.: Lightweight and integrity-protecting oriented data aggregation scheme for wireless sensor networks. IET Inf. Secur. **11**(2), 82–88 (2017)
28. Zhou, L., Ge, C., Hu, S., Su, C.: Energy-efficient and privacy-preserving data aggregation algorithm for wireless sensor networks. IEEE Internet Things J. **7**(5), 3948–3957 (2019)
29. Zuniga, M., Krishnamachari, B.: Analyzing the transitional region in low power wireless links. In: 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON 2004), pp. 517–526. IEEE (2004)