Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 9754973, 10 pages
https://doi.org/10.1155/2020/9754973

Research Article

Query Privacy Preserving for Data Aggregation in Wireless Sensor Networks

X. Liu,[1] **X. Zhang**,[1] **J. Yu**,[2] **and C. Fu**[1]

[1]*School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China*
[2]*School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong 250353, China*

Correspondence should be addressed to X. Liu; liuxw@qfnu.edu.cn

Received 8 August 2019; Revised 24 December 2019; Accepted 21 January 2020; Published 24 February 2020

Academic Editor: Hui Cheng

Wireless Sensor Networks (WSNs) are increasingly involved in many applications. However, communication overhead and energy efficiency of sensor nodes are the major concerns in WSNs. In addition, the broadcast communication mode of WSNs makes the network vulnerable to privacy disclosure when the sensor nodes are subject to malicious behaviours. Based on the abovementioned issues, we present a Queries Privacy Preserving mechanism for Data Aggregation (QPPDA) which may reduce energy consumption by allowing multiple queries to be aggregated into a single packet and preserve data privacy effectively by employing a privacy homomorphic encryption scheme. The performance evaluations obtained from the theoretical analysis and the experimental simulation show that our mechanism can reduce the communication overhead of the network and protect the private data from being compromised.

1. Introduction

As a novel and modern technique, Wireless Sensor Networks (WSNs) have been introduced into a variety of scenarios such as medical applications [1], smart homes [2, 3] autonomous vehicles [4], traffic administration [5] and military battlefields [6]. A WSN is composed of hundreds or thousands of tiny resource-constrained sensor nodes which are generally deployed in an unattended even hostile area. These nodes are difficult to be replaced or recharged. This prevents WSNs from being applied into more critical applications, especially in scenarios where the long lifetime and the high quality services are needed. It is important that traffic and computation overhead should be kept as low as possible to extend the lifetime of WSNs. The Data Aggregation (DA) [7–13] technique is one of the most effective ways for the network to save energy and improve efficiency. It can reduce the quantity of information transmission through aggregating the data from different nodes, decreasing redundancy, and achieving the goal of prolonging the lifetime of the network. Unfortunately, DA is vulnerable to some attacks. Taking the aggregation node as an instance, it is an intermediate tier between sensor nodes and Base Station (BS). The main roles of aggregation nodes are to store the sensing data and reply the queries received from BS. If most of the aggregation nodes have been compromised successfully, the data of whole network may be revealed and tampered with easily. This may result in serious threat or economic loss, even the damage to the safety of state property. Therefore, the Security Data Aggregation (SDA) plays an important role in the critical application of WSNs.

Privacy Preserving (PP) has attracted much attention in many fields, such as smart grid [14], Internet of Things [15, 16], edge computing [17], social network [18] and other application scenarios [19–21]. PP can also protect the privacy of sensing data when DA is adopted in a WSN, and some interesting schemes have been proposed in recent years [22–25]. However, these solutions cannot guarantee the data integrity. Although the schemes discussed in [26–28] exploited the issue of data integrity, they may cause the leakage of concealed data due to the decryption at the aggregation nodes. A proposed scheme in [29] attempted to

bridge the gap between PP and data integrity through integrating an encryption algorithm with an MAC authentication mechanism, but it has the risk of putting a heavy computation burden on sensor nodes.

In general, BS has two ways to collect information in a WSN. One is that BS sends a query and the nodes reply accordingly. The other is that the nodes periodically report information to the BS. We focus on the former one in this paper for the reason that the latter one consumes more resources in transmission replies which is inconsistent with our intention of saving energy. The data query has been widely exploited in the current studies. For example, the maximum/minimum query was used to monitor a patient and identify the maximum or minimum value of an indicator which could be regarded as a symbol to determine whether the patient is in a good state or not [30]. Up to now, the single query with PP, such as range query [31], verifiable top-k query [32], and location query [33], has been well addressed. However, the single query method cannot meet the requirements of application when it is introduced into a large-scale network. Therefore, how to enrich the function of query becomes an urgent research challenge. As one of the reasonable solutions, the multiple queries mechanism has been proposed in which many queries can be executed simultaneously [34]. However, the multiple query mechanism with PP is an emerging direction, and many valuable issues need to be solved in the future.

To address the abovementioned issues, we propose a Queries Privacy Preserving mechanism for Data Aggregation (QPPDA) in this paper. The goal of our work is to bridge the gap between PP and energy consumption, and the following techniques are adopted. Firstly, the multiple queries are aggregated into a single packet in order to reduce energy consumption. Then, a homomorphic encryption scheme is carried out, and the confidentiality of private data is ensured. Next, the data for different queries in a single aggregated packet can be distinguished from each other in the decryption of the aggregated data at BS. Compared with the single query, QPPDA may greatly decrease the communication and computation overhead. The main contributions of this paper are as follows.

(i) *Improvement of Gridding Technology*. The high computation complexity of cell limits the application of the grid technique. We break this restriction through improving the relative location algorithm in grid topology. As a result, the computation complexity is decreased, and the relative location provides an efficient way to maintain a dynamic WSN.

(ii) *Effective Privacy Preserving*. Privacy is easily destroyed by an attacker for a WSN usually deployed in an unattended even hostile environment. The elliptic curve encryption combined with the homomorphic algorithm is adopted to effectively protect the private data from being compromised.

(iii) *Efficient Reply*. Sending multiple replies individually leads to the wastage of network resources. Through aggregating the multiple queries into a single packet, the performances of WSN are promoted in terms of energy consumption and lifetime.

The rest of the paper is organized as follows. Section 2 introduces related work. Section 3 discusses the topology construction of the network. Section 4 elaborates our scheme in detail. Section 5 evaluates the performance of QPPDA. We conclude this paper in Section 6.

2. Related Work

2.1. Grid Topology. The connectivity is one of the key issues in WSNs, and many valuable solutions have been proposed to deal with this challenge. A grid-based SDA scheme was proposed in [35]. The whole network was divided into some nonoverlapping virtual cells which were small enough to ensure that the radio coverage of a node can cover its surrounding cells, namely, each node in a cell can directly communicate with the nodes in the neighbouring cells. In [36], the nodes were divided into groups according to their geographic locations with only one node reserved in each group which can connect to the backbone network. In this way, the proposed scheme in [36] not only ensures the connectivity of nodes, but also speeds up the convergence rate of the network. Although the connectivity of the network is guaranteed, the grid topology causes a higher computational complexity than tree or cluster topology.

2.2. Privacy Preserving. As to PP, some cryptographic schemes have been adopted to carry out the hop-by-hop encryption [37]. He et al. presented an Integrity-protecting Private Data Aggregation scheme (IPDA) [38], which is an improvement on the Cluster-based Private Data Aggregation (CPDA) [22]. Both IPDA and CPDA achieve privacy preserving through the technique of data slicing and assembling which ensures integrity by constructing two disjointed aggregation trees. However, the disjointed aggregation trees are computation- and communication-consuming and inapplicable to resource-constrained WSNs. As far as the hop-by-hop scheme is concerned, data privacy cannot be guaranteed because the ciphertext must be decrypted in the intermediate nodes when DA technique is applied. Therefore, the end-to-end scheme is a desirable choice in a network with DA. In [30, 31], the nodes directly sent the encrypted data to the BS without the decryption operation involved in the intermediate nodes. Castelluccia et al. [39] proposed a simple and provable secure additive homomorphic stream which permitted the efficient aggregation of encrypted data. Girao et al. [40] discussed a mechanism which can conceal the sensing data and the aggregation data in an end-to-end manner. Though these schemes are efficient in preserving data privacy of DA, they cannot prevent the private data from being eavesdropped by their neighbours. Compared with [40], the Integrity Protecting Hierarchical Concealed Data Aggregation (IPHCDA) for WSNs ensured that no private data of a sensor node were released to any other nodes under the support of asymmetric cryptography [41]. It employed the elliptic curve-based Privacy Homomorphic (PH) and

allowed the concealed aggregation data to be encrypted with different keys. The scheme includes the following steps.

Step 1: generate key pairs according to the point on the elliptic curve (p_u, p_r).

Step 2: encrypt m using $c = g^m + h^r$, where + is the addition operation of elliptic curve points, r is a random number, and g^m and h^r are the scalar multiplication of elliptic curve points.

Step 3: perform the DA. Two ciphertexts $(c_1 = g^{m_1} + h^{r_1}, c_2 = g^{m_2} + h^{r_2})$ are fused to a single ciphertext $c\prime = c_1 + c_2 = g^{m_1+m_2} + h^{r_1+r_2}$.

Step 4: decrypt a ciphertext using the private key p_r at BS.

2.3. Query Privacy Preserving. The contributions presented in [42–44] investigated the privacy schemes of single query when attackers attempted to tamper with or eavesdrop on the private information of nodes. Papadopoulos et al. proposed a privacy-preserving scheme of range query [42] based on the bucketing technique [45], in which the domain of data values was divided into multiple buckets, and the time was divided into slots as well. In each time slot, data items collected by a sensor node were classified into different buckets with different IDs. If BS wants to perform a range query, it does not send the range directly. Instead, the bucket with various IDs that covers the required range is sent to the storage nodes. However, the bucket partitioning technique cannot prevent a compromised storage node from carrying out malicious activities in a WSN. Faced with this challenge, the proposed scheme in [46] discussed the privacy of query by encoding the sensing data. However, it needs high computation overhead and communication cost. To the best of our knowledge, rare contribution is found in investigating the privacy preserving of multiple queries with DA.

Different from the abovementioned approaches, QPPDA has the advantages of decreasing the resource consumption and protecting the private data from being compromised simultaneously.

3. Network Model

3.1. Sensor Networks and Data Aggregation Model. A sensor network is modelled to a grid which is divided into many cells with each one containing a number of sensor nodes. There are three types of nodes in a network: BS, Aggregation Node (AN), and Member Node (MN). It is assumed that BS is trusted and has unlimited energy, computing resource, and storage capacity. MN collects the sensing data and sends them to AN. And AN is responsible for forwarding the query sent by BS and aggregating the data of MNs. The network size is N which means that there are N nodes in a WSN. The sensor nodes are organized into a grid structure as shown in Figure 1. Notice that we adopt the three-dimensional model rather than other two dimensional models in most of the related works with grid topology. This model may expand the application scenario of QPPDA, and it can be used in many complex natural environments.

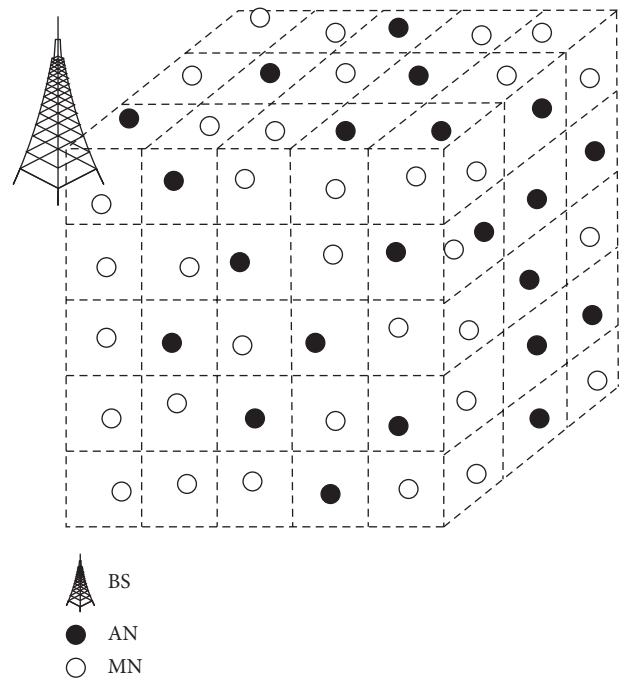

Figure 1: Grid network model.

Let $D = \{d_1, d_2, \ldots, d_N\}$ be raw data gathered at MNs. The set of sensing data, D, can be transmitted to BS hop-by-hop. However, transmitting all the raw data to BS may result in a huge burden on the bandwidth and high energy consumption. Therefore, DA is a favorite technique to decrease the occupancy of resources.

A data aggregation function is defined as $y(t) = f(d_1(t), d_2(t), \ldots, d_N(t))$ at time t, where f represents the aggregation function which may be addition, average, min, max, and count. We focus on addition aggregation functions in our model $f(t) = \sum_{i=1}^{N} d_i(t)$. It should be noticed that the addition aggregation function is not too restrictive because many other functions such as average and count which can be deduced from the addition function.

3.2. Threat Model. When queries are initialized, BS broadcasts them to the whole network. The nodes which meet the requirements of the queries send their reply data to AN, and the data are sensitive to the malicious activates if the security mechanism is absent. We adopt the well-known "honest but curious" threat model [47], in which the adversaries attempt to break the privacy but faithfully follow the protocol specification during the process of DA. Meanwhile, adversaries can overhear the original data of sensors through eavesdropping on the wireless link. In addition, a few nodes may collude with each other to violate the data privacy of the overall network.

4. Privacy Data Aggregation Protocol

We present a privacy data aggregation protocol called Queries Privacy Protection for Data Aggregation (QPPDA) which involves three phases: the grid division, the key generation, and the query processing. Firstly, a network is

divided into adjacent virtual cells, and the nodes within neighbouring cells can directly communicate with each other. Secondly, the corresponding key for each type of query is generated in order to guarantee the data privacy. Finally, the nodes aggregate multiple replies into a single packet which is transmitted to BS hop-by-hop.

4.1. Grid Division. The grid division phase is responsible for the construction of the network structure. In the Geographical Adaptive Fidelity (GAF) algorithm, a network area was divided into grid topology which consisted of many contiguous cells according to the geographic information and the radio coverage of nodes [46]. In order to make GAF suitable for the WSNs in practice, some improvements of GAF were proposed, and a relative position was adopted to obtain the grid information [48]. However, some valuable issues, such as data privacy and accuracy, are left for future study.

We define all the cells that have a common edge as the neighboring cells. In the division process, it should be determined that all the nodes of a cell can directly communicate with the nodes in the neighboring cells. Thus, equation (1) needs to be satisfied.

$$r^2 + (2r)^2 + (2r)^2 \le R^2 \longrightarrow 9r^2 \le R^2 \longrightarrow r \le \frac{1}{3}R, \quad (1)$$

where r denotes the side length of the cell and R is the communication radius of the node.

The relationship between r and R can be shown as Figure 2. We take the following steps to divide a grid into adjacent cells. Firstly, BS broadcasts its location, $L_{bs}(x_{bs}, y_{bs}, z_{bs})$, and the side length r of each cell to all the nodes in a WSN. Node $i(x_i, y_i, z_i)$ can calculate the coordinate of cell $G_i(g(x_i), g(y_i), g(z_i))$ and determine which cell node i belongs to using the following equation:

$$\begin{cases} g(x_i) = \left\lceil \dfrac{(x_i - x_{bs})}{r} \right\rceil, \\ g(y_i) = \left\lceil \dfrac{(y_i - y_{bs})}{r} \right\rceil, \\ g(z_i) = \left\lceil \dfrac{(z_i - z_{bs})}{r} \right\rceil. \end{cases} \quad (2)$$

Now, we use a simple example to explain why we use the top integral instead of the bottom integral when the cell coordinate is determined in equation (2). Assume that the coordinate of node i is (9, 11, 10) and that of BS is (0, 0, 0), respectively. The side length r is 4, as shown in Figure 3. We firstly calculate the cell where node i stays using the top integral according to equation (2) $g(x) = (9 - 0)/4 = 3$, $g(y) = (11 - 0)/4 = 3$, $g(z) = (10 - 0)/4 = 3$. Therefore, node i is inside $G_i(3, 3, 3)$. On the contrary, we can obtain G_i is (2, 2, 2) using the bottom integral. It can be seen from Figure 3 that $G_i(3, 3, 3)$ is the desired one.

The pseudocode of the grid division is listed in Algorithm 1. N nodes compute their coordinates from Line 1 to Line 10, and the computation complexity of grid division is $O(N)$.

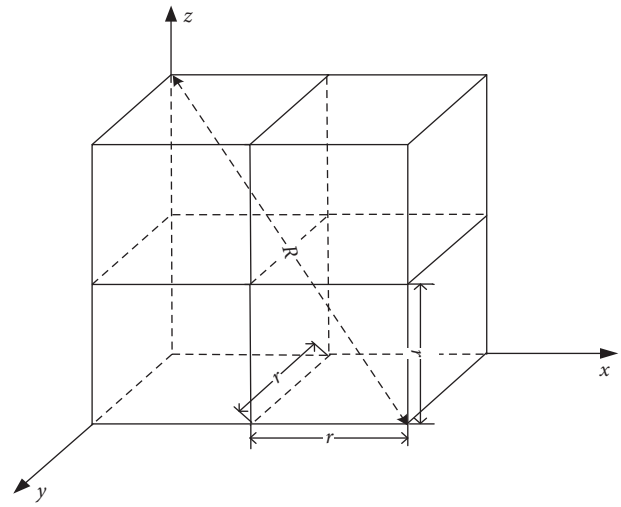

Figure 2: The relation between (r) and (R).

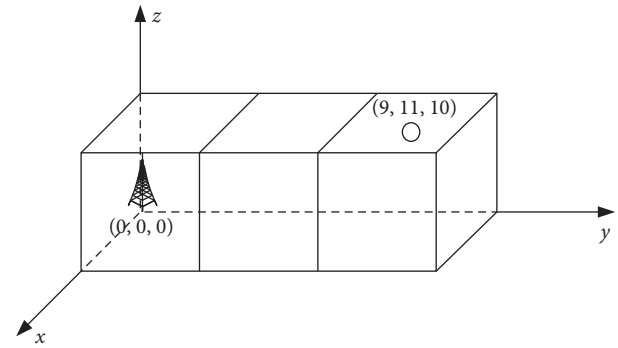

Figure 3: The illustration of the cell coordinate.

After the grid is divided into cells, some sensor nodes (ANs) in different cells are selected and organized into an aggregation tree rooted at BS. In a cell, the member nodes send the data to AN, and AN sends the aggregation results to BS hop-by-hop along the aggregation tree. Figure 4 demonstrates the aggregation tree and the data aggregation in a cell, respectively.

4.2. Key Generation. We introduce the homomorphic encryption scheme based on the elliptic curve [14] into QPPDA, which can protect private data from being revealed. The encryption method assigns different keys to the concealed data acquired from different nodes, and BS can correctly distinguish them in the aggregation process [41].

Assume that k types of query are supported by a network. Therefore, k public and private key pairs are required. We take the following steps to generate the key pairs.

Given a parameter τ, we define an algorithm $\Psi(\tau)$ to output a tuple $(q_1, q_2, \ldots, q_{k+1}, E)$, where E is a set of elliptic curve points that form a cyclic group. The order of E is n where $n = q_1 \times q_2 \times \cdots \times q_{k+1}$. $\Psi(\tau)$ works as follows.

Input: The side length of cell, r; The communication radius, R; The position of BS, $L_{bs}(x_{bs}, y_{bs}, z_{bs})$; The position of a node, $i(x_i, y_i, z_i)$.
Output: the nodes belong to the cells, $G = \{G_1, G_2, \ldots, G_N\}$.
(1) **Begin**
(2) BS broadcasts L_{bs}
(3) **For** $(i = 1; i \le N; i++)$
(4) **If** node i receives L_{bs} **then**
(5) Calculate $G_i(g(x_i), g(y_i), g(z_i))$ according to equation (2)
(6) $\begin{cases} g(x) = \lceil (x_i - x_{bs})/r \rceil, \\ g(y) = \lceil (y_i - y_{bs})/r \rceil, \\ g(z) = \lceil (z_i - z_{bs})/r \rceil. \end{cases}$
(7) Obtain the position $G(g(x_i), g(y_i), g(z_i))$;
(8) **End If**
(9) **End For**
(10) **Return** Cell information of all nodes, $G = \{G_1, G_2, \ldots, G_N\}$.
(11) **End**

Algorithm 1: The grid division.

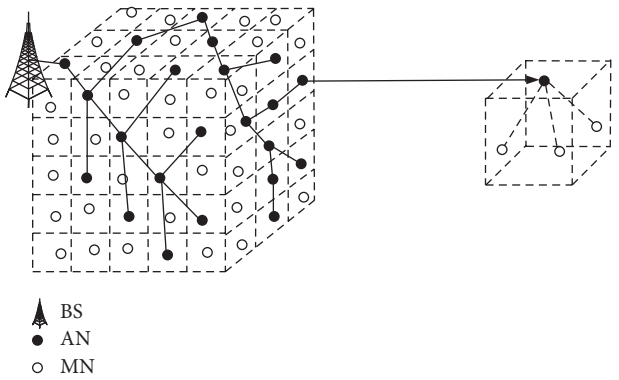

Figure 4: The aggregation tree and the data aggregation in a cell.

(i) Generate $k+1$ random τ-bit primes $(q_1, q_2, \ldots, q_{k+1}, E)$ and set $n = q_1 \times q_2 \times \cdots \times q_{k+1}$
(ii) Generate a set of elliptic curve points E
(iii) Output the security elements ()

The points $(\alpha_1, \alpha_2, \ldots, \alpha_{k+1}, \beta)$ are randomly chosen, and the order of these points are n. Then, we calculate γ as

$$\gamma = \alpha_{k+1}^{\prod_{r=1}^{k} q_r}. \tag{3}$$

The order of γ is q_{k+1}. Next, k public keys are computed for k queries according to the following equation:

$$h_j = \beta_j^{\delta}, \quad \delta = \prod_{r=1, r \ne j}^{k+1} q_r, \quad j = 1, 2, \ldots, k, \tag{4}$$

where the order of h_j is q_j. Finally, we establish the public key set $P_{uk} = \{n, E, h_1, h_2, \ldots, h_k, \beta, \gamma\}$ and the private key set $P_{rk} = \{q_1, q_2, \ldots, q_{k+1}\}$. Therefore, the jth key pair, (P_{uj}, P_{rj}) is generated for the jth query.

The key generation is illustrated in Algorithm 2 where Lines 4 to 6 obtain a tuple by the elliptic curve algorithm, and Lines 7 to 15 display the process of producing keys. N nodes execute the elliptic curve algorithm to generate key pairs for k queries with the complexity of $O(N^2)$.

4.3. Query Processing. After the aggregator receives a request of query from BS, it broadcasts $Q = (\text{type} = p, \text{epoch} = t, \text{time} = T)$ to node i $(i = 1, 2, \ldots, N)$, where p and t represent the types of queries and the query epoch, respectively. T denotes the time that AN spends on replying to BS. Four steps should be taken to process the query: the data collection, the data encryption, the data aggregation, and the data decryption.

4.3.1. Data Collection. After receiving the queries, the nodes collect the sensing data $d_i \in \{0, 1, \ldots, D\}$ where D is the maximum value of d_i according to the query types.

4.3.2. Data Encryption. The nodes encrypt data using the public key and the encryption process is as follows.

Step 1: a node chooses a random number r from $\{0, 1, \ldots, N-1\}$.

Step 2: node i selects a key according to the type of query. If the type of query is x, that is $p = x$, the public key is $P_{uk} = P_{ux} = (n, E, g, h_x)$, where $h_x = u^{q^{j+1}}$.

Step 3: Node i computes the ciphertext $C_i^x = f(d_i) = (g^{d_i} + h_x^{r_x})$.

4.3.3. Data Aggregation. Let d_x denote the reply message of the xth query. Consequently, k ciphertexts in node i (aggregator) are aggregated into a ciphertext of C_i' using the following equation:

$$C_i' = f\left(\sum_{x=1}^{k} d_x\right) = \sum_{x=1}^{k} g^{d_x} + h_x^{r_x} = g^{\sum_{x=1}^{k} d_x} + h_x^{\sum_{x=1}^{k} r_x}. \tag{5}$$

Input: security parameter, τ; query types, k.
Output: key sets (P_{uk}, P_{rk}).
(1) **Begin**
(2) Generate a tuple $(\cdot)$
(3) Find $k+1$ random τ-bit primes $\{q_1, q_2, \ldots, q_{k+1}\}$
(4) Let $n = q_1 \times q_2 \times \cdots \times q_{k+1}$
(5) Find a set of elliptic curve points E
(6) Generate a tuple $(q_1, q_2, \ldots, q_{k+1}, E)$
(7) **For** node $(i = 1; i <= N; i++)$
(8) Choose points $(\alpha_1, \alpha_2, \ldots, \alpha_{k+1}, \beta)$ from E
(9) $\gamma = \alpha_{k+1} \prod_{r=1}^{k} q_r$
(10) **For** query $(j = 1; j <= k; j++)$
(11) $h_j = \beta_j^{\delta}$, $\delta = \prod_{r=1, r\neq j}^{k+1} q_r$, $j = 1, 2, \ldots, k$
(12) $pk_j \in P_{uk} = \{n, E, h_1, h_2, \ldots, h_k, \beta, \gamma\}$
(13) $pr_j \in P_{rk} = \{q_1, q_2, \ldots, q_{k+1}\}$
(14) **End For**
(15) **End For**
(16) **Return** (P_{uk}, P_{rk}).
(17) **End**

Algorithm 2: The key generation.

4.3.4. Data Decryption. During the decryption, BS is able to decrypt the data of each query separately from the aggregated ciphertext C_i'. To decrypt a ciphertext C_i^x, BS needs to obtain the plaintext from equation (6) using the private key q_x.

$$d_x = \log_{g^{q_x}} (C_i^x)^{q_x}. \tag{6}$$

The pseudocode of query processing is shown in Algorithm 3. Lines 1 to 10 describe the data collection, the data encryption, and the data aggregation in detail. Lines 11 to 16 delineate the process of separating the decryption data at BS with the complexity of $O(N)$.

5. Performance Analysis and Simulation Experiment

We evaluated the performance of QPPDA in terms of privacy preservation, communication efficiency, and computation overhead through theoretical analysis and simulation experiment. QPPDA was implemented using MATLAB. A WSN with 600 nodes was considered, and these nodes were randomly deployed in a 400 m $*$ 400 m area. The transmission range of the sensor node was 50 m.

5.1. Privacy-Preservation Analysis. We analyze the privacy preservation performance of QPPDA when a node is compromised by physical attack. If an adversary compromises an AN, it can perform an unauthorized aggregation and send false aggregation results to BS. However, due to the asymmetry of public key, an adversary cannot gain any additional information related to the data aggregation. Hence, the compromised node may affect the data integrity but not the data confidentiality in QPPDA.

Through the analysis, we can conclude that the privacy can be revealed because of the leakage of keys. Assume that p_{ovk} and p_{ovr} are the probabilities of the key and the random number which are broken, respectively. Therefore, the probability of information leakage is $p = (p_{\text{ovk}} \times p_{\text{ovr}})^k$. Figure 5 demonstrates the privacy performance of different types of the queries. We may find that the exposure probability of privacy is less than 0.45% even if the reveal probability of key is 0.4. Besides, the more frequently the queries are sent, the better the data confidentiality will be. This proves that QPPDA can effectively preserve the data privacy.

5.2. Energy Consumption. In our experiments, we considered the efficiency of communication and computation and adopted the typical data query schemes, the single query [42] and the Slice-Mix-AggRegaTe (SMART) [22], as the benchmarks to verify the energy consumption of QPPDA.

5.2.1. Communication Overhead. Communication overhead is mainly derived from data transmission, e.g., a node transmits its sensing data to AN or BS in a WSN. For node i, the length of data is l_{ci} bits and the average number of hops between any two nodes is L. Thus, the communication overhead of a cell with k queries can be computed as

$$C_{TK} = \sum_{i=1}^{N} l_{ci} L. \tag{7}$$

The overhead of a single query comes from sending the encrypted data and HMACs. The HAMCs of node i is l_{hi} bits and the data is l_{ci} bits. Then, the overhead of a cell with k single queries are formalized as

$$C_{Ts} = \sum_{i=1}^{N} [(3l_{ci} - 1) l_{hi} + l_{ci}] Lk. \tag{8}$$

In SMART, each node divides its sensing data into three slices, two of which are sent to the neighboring nodes and

Input: query $Q = (\text{type} = p, \text{epoch} = t, \text{time} = T)$.
Output: decrypted data of each query x, d_x.

(1) **Begin**
(2) **For** node $(i = 1; i \le N; i++)$
(3) **If** $(p == x)$ **Then**
(4) Collect the data $d_i \in \{0, 1 \ldots, M\}$;
(5) Choose a random number r from $\{0, 1, \ldots, N-1\}$;
(6) Select a key $P_{ux} = (n, E, g, h_x)$;
(7) Encrypt the data, $C_i^x = f(d) = (g^d + h_x^{r_x})$;
(8) Aggregate the data, $C_i' = f(\sum_{x=1}^{k} d_x) = g^{\sum_{x=1}^{k} d_x} + h_x^{\sum_{x=1}^{k} r_x}$;
(9) **End If**
(10) **End For**
(11) **For** $(i = 1; i \le x; i++)$ at BS
(12) Decrypt the data, $d_x = \log_{g^{q_x}} (C_i^x)^{q_x}$;
(13) **End If**
(14) **Return** $\{d_1, d_2, \ldots, d_x\}$
(15) **End**

Algorithm 3: The query processing.

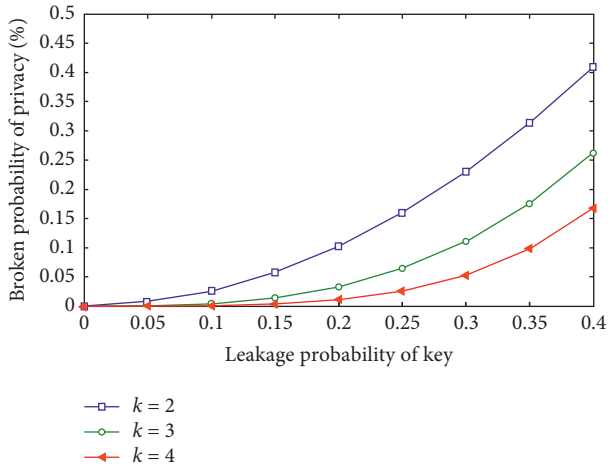

Figure 5: The exposure probability of privacy.

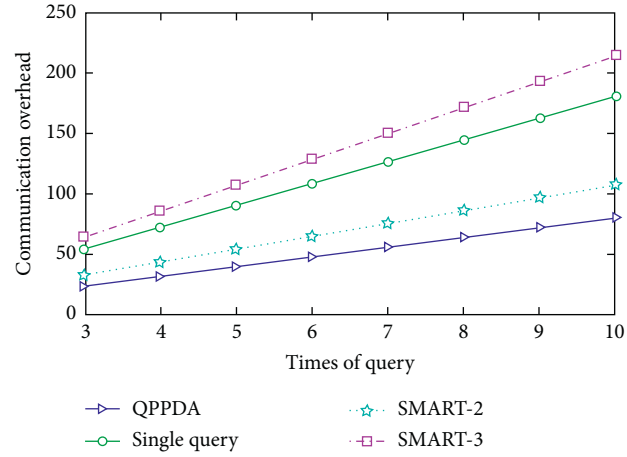

Figure 6: The communication overhead.

one is preserved by itself. Assume that a node receives n slices from other nodes. Therefore, the overhead with k queries can be described as

$$C_{T\text{smart}} = \sum_{i=1}^{N} \left(\frac{2}{3} + \frac{n}{3}\right) l_{ci} LK. \quad (9)$$

Figure 6 shows the communication consumption of QPPDA, Single query, and SMART, where SMART-n denotes that a node receives n slices. A conclusion can be drawn from Figure 6 that the communication overhead of QPPDA is much less than that of SMART-n or Single query. QPPDA is one of the most efficient schemes in decreasing the communication overhead.

5.2.2. Computation Cost. Single query converts the query scope to a prefix format before the data are transmitted. The number of binary prefix is nearly $(2l_{ci} - 2)$, and there are exactly $(l_{ci} + 1)$ prefixes. Therefore, the node needs to perform about $(2l_{ci} - 2) * (l_{ci} + 1)$ comparisons, and the computation complexity of query is $O(w^2 Nk)$ in the worst case. The computation overhead of QPPDA comes from data encryption, and its computation complexity is $O(l_{ci} N^2)$ according to Algorithm 2. Data mixing is the prime computational consumption in SMART, which is $O(N)$. Consequently, it is observed that the computation consumption of single query is higher than that of QPPDA and SMART when the number of nodes is fixed in a WSN according to Figure 7. It should be noticed that the computation overhead of SMART is less than that of QPPDA when one slice mechanism (SMART-1) is adopted. However, one slice SMART may result in a lower security level compared with QPPDA. Therefore, our scheme is a better tradeoff between security and computation complexity.

5.3. Aggregation Accuracy. The accuracy is defined as the ratio between the collected summation by the data aggregation and the real summation of all individual sensor nodes in [22]. Figure 8 illustrates the accuracy of QPPDA, Single query, and

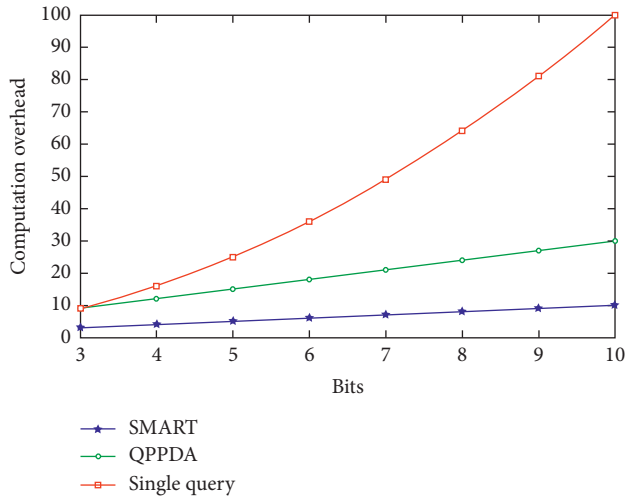

Figure 7: The computation cost.

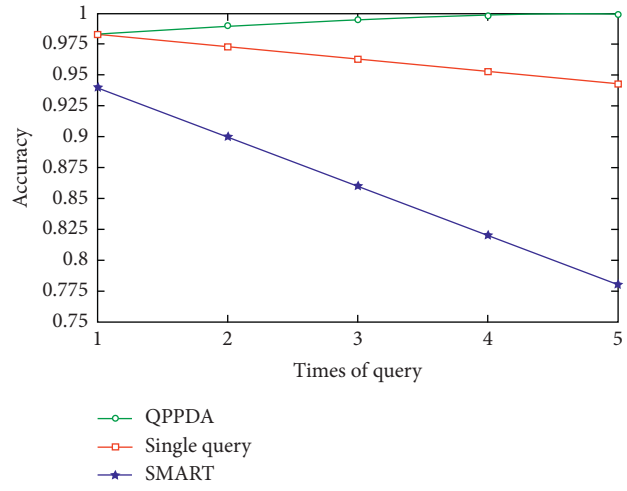

Figure 8: Accuracy of three schemes.

SMART with respect to different query times in our simulation. From Figure 8, we can observe that the accuracy of QPPDA improves as the times of query increase. Two reasons contribute to this which have already been analyzed in [8]: (i) with a longer time interval, the data messages to be sent within this duration will have less chance to collide; (ii) with a longer time interval, the data messages will have a better chance of being delivered before the deadline.

Besides, we can observe that QPPDA has a better accuracy than single query and SMART. It has been demonstrated that the communication overhead of QPPDA is reduced significantly, and the amount of transmission of QPPDA is much less than that of single query and SMART in Section 5.2.1. Therefore, the chance of collision and packet loss are also decreased, which leads to an improvement in aggregation accuracy.

6. Conclusion

The energy consumption and data privacy are two important concerns in WSNs. The limited energy of sensor nodes may shorten the lifetime of network, and the nodes are often deployed in dangerous areas where the data privacy may be more likely to be destroyed easier than in the cable network. Faced with these challenges, we present a query privacy protection mechanism for data aggregation which can reduce energy consumption and preserve the data privacy as well. Experimental results show that our scheme can guarantee the data privacy, decrease the system overhead, and improve the accuracy of data aggregation. For the future work, we will focus on other aggregation functions, such as mean, max, and counter except the additive aggregation. The privacy of QPPDA is closely related to the number of keys, and it is a challenging work to promote the security of QPPDA without the complex key distribution so as to save energy and decrease the requirement of storage. In addition, tree or cluster topology will be discussed in our subsequent study in order to expand the application scenarios of our scheme.

Data Availability

The datasets generated or analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 61672321, 61832012, 61771289, and 61373027 and the Shandong Graduate Education Quality Improvement Plan SDYY17138.

References

[1] A. B. García-Hernando, J. F. Martínez-Ortega, J. M. López-Navarro, A. Prayati, and L. Redondo-López, *WSN Application Scenarios: Problem Solving for Wireless Sensor Networks*, Springer, Berlin, Germany, 2008.

[2] W. Huiyong, W. Jingyang, and H. Min, "Building a smart home system with WSN and service robot," in *Proceedings of the 5th International Conference on Measuring Technology and Mechatronics Automation*, pp. 353–356, Hong Kong, China, January 2013.

[3] T. Song, R. Li, B. Mei, J. Yu, X. Xing, and X. Cheng, "A privacy preserving communication protocol for IoT applications in smart homes," *IEEE Internet of Things Journal*, vol. 4, no. 12, pp. 1844–1852, 2017.

[4] J. Wang, Z. Cai, and J. Yu, "Achieving personalized k-anonymity based content privacy for autonomous vehicles in CPS," *IEEE Transactions on Industrial Informatics*, 2019.

[5] Z. Cai, X. Zheng, and J. Yu, "A differential-private framework for urban traffic flows estimation via taxi companies," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6492–6499, 2019.

[6] M. Hussain, P. Khan, and K. K. Sup, "WSN research activities for military application," in *Proceedings of the 11th International*

Conference on Advanced Communication Technology, pp. 271–274, Phoenix Park, Korea, February 2009.

[7] S. Ozdemir and Y. Xiao, "Secure data aggregation in wireless sensor networks: a comprehensive overview," *Computer Networks*, vol. 53, no. 12, pp. 2022–2037, 2009.

[8] B. Sujatha, C. T. Jilo, and C. S. Rao, "Energy-efficient data route-in-network aggregation with secure eedrina," in *Proceedings of the International Conference on Computational Intelligence and Data Engineering*, pp. 1–9, Guntur, India, September 2018.

[9] E. Choudhari, K. D. Bodhe, and S. M. Mundada, "Secure data aggregation in wsn using iterative filtering algorithm," in *Proceedings of the International Conference on Innovative Mechanisms for Industry Applications*, pp. 1–5, Bangalore, India, February 2017.

[10] H. Wang, Z. Wang, and J. Domingo-Ferrer, "Anonymous and secure aggregation scheme in fog-based public cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 712–719, 2018.

[11] H. Wang, D. He, J. Yu, and Z. Wang, "Incentive and unconditionally anonymous identity-based public provable data possession," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 824–835, 2019.

[12] H. Wang, Q. Wang, and D. He, "Blockchain-based private provable data possession," *IEEE Transactions on Dependable and Secure Computing*, 2019.

[13] H. Wang, D. He, A. Fu, Q. Li, and Q. Wang, "Provable data possession with outsourced data transfer," *IEEE Transactions on Services Computing*, 2019.

[14] D. He, S. Zeadally, H. Wang, and Q. Liu, "Lightweight data aggregation scheme against internal attackers in smart grid using elliptic curve cryptography," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 3194845, 11 pages, 2017.

[15] H. Yan, C. Yong, and Y. Lu, "Re-ADP: real-time data aggregation with adaptive-event differential privacy for fog computing," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 6285719, 13 pages, 2018.

[16] Y. Pu, J. Luo, C. Hu et al., "Two secure privacy-preserving data aggregation schemes for IoT," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 3985232, 11 pages, 2019.

[17] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, "Edge computing security: state of the art and challenges," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1608–1631, 2019.

[18] Z. He, Z. Cai, and J. Yu, "Latent-data privacy preserving with customized data utility for social network data," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 665–673, 2018.

[19] X. Liu, J. Yu, F. Li, W. Lv, Y. Wang, and X. Cheng, "Data aggregation in wireless sensor networks: from the perspective of security," *IEEE Internet of Things Journal*, 2019.

[20] L. Ni, C. Li, X. Wang, H. Jiang, and J. Yu, "DP-MCDBSCAN: differential privacy preserving multi-core DBSCAN clustering for network user data," *IEEE Access*, vol. 6, pp. 21053–21063, 2018.

[21] W. Li, C. Hu, T. Song, J. Yu, X. Xing, and Z. Cai, "Privacy-preserving data collection in context-aware applications," in *Proceedings of IEEE Symposium on Privacy-Aware Computing (PAC)*, pp. 75–85, Washington, DC, USA, September 2018.

[22] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "PDA: privacy-preserving data aggregation in wireless sensor networks," in *Proceedings of the 26th International Conference on Computer Communications*, pp. 2045–2053, Barcelona, Spain, May 2007.

[23] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacy-preserving data aggregation without secure channel: multivariate polynomial evaluation," in *Proceedings of the International Conference on Computer Communications (INFOCOM)*, pp. 2734–2742, Turin, Italy, April 2013.

[24] K. Xing, Z. Wan, P. Hu et al., "Mutual privacy-preserving regression modeling in participatory sensing," in *Proceedings of the International Conference on Computer Communications (INFOCOM)*, pp. 3039–3047, 2013.

[25] G. Yang, S. Li, X. Xu, H. Dai, and Z. Yang, "Precision-enhanced and encryption-mixed privacy-preserving data aggregation in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 4, Article ID 427275, 2013.

[26] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: a secure hop-by-hop data aggregation protocol for sensor network," in *Proceedings of 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 356–367, Catania, Italy, July 2006.

[27] K. B. Frikken, K. Kauffman, and K. Steele, "General secure sensor aggregation in the presence of malicious nodes," in *Proceedings of the 27th International Proceedings on Data Engineering*, pp. 506–516, Hannover, Germany, April 2011.

[28] L. Zhu, Z. Yang, M. Li, and D. Liu, "An efficient data aggregation protocol concentrated on data integrity in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 5, Article ID 256852, 2013.

[29] Q. Zhou, G. Yang, and L. He, "An efficient secure data aggregation based on homomorphic primitives in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 10, no. 1, Article ID 962925, 2014.

[30] H. Dai, X. Qin, L. Liu, Y.-M. Ji, X. Fu, and Y. Sun, "Z-O encoding based privacy-preserving max/min query protocol in two-tiered wireless sensor networks," *Journal of Electronics Information Technology*, vol. 35, no. 4, pp. 970–976, 2013.

[31] T. Wang, L. Qin, and L. Liang, "Privacy-preserving range query in two-tiered wireless sensor networks," *Journal of Beijing University of Posts Telecommunications*, vol. 37, no. 2, pp. 104–108, 2014.

[32] S. Shekhar and H. Xiong, *Top-K Query*, Springer, Berlin, Germany, 2008.

[33] D. Song, M. Song, and K. Park, "A privacy-preserving spatial index for spatial query processing," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2067047, 9 pages, 2018.

[34] E. G. Prathima, T. S. Prakash, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "SDAMQ: secure data aggregation for multiple queries in wireless sensor networks," *Procedia Computer Science*, vol. 89, pp. 283–292, 2016.

[35] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad-hoc routing," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, pp. 70–84, London, UK, September 2001.

[36] X. Luo, Y. Yan, S. Li, and X. Guan, "Topology control based on optimally rigid graph in wireless sensor networks," *Computer Networks*, vol. 57, no. 4, pp. 1037–1047, 2013.

[37] H. Li, K. Lin, and K. Li, *Energy-efficient and High-Accuracy Secure Data Aggregation in Wireless Sensor Networks*, Elsevier, Amsterdam, Netherlands, 2011.

[38] W. He, H. Nguyen, X. Liuy, K. Nahrstedt, and T. Abdelzaher, "iPDA: an integrity-protecting private data aggregation scheme for wireless sensor networks," in *Proceedings of the*

International Conference on Military Communications, pp. 1–7, San Diego, CA, USA, November 2009.

[39] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Proceedings of International Conference on Mobile and Ubiquitous Systems*, pp. 109–117, San Diego, CA, USA, July 2005.

[40] J. Girao, D. Westhoff, and M. Schneider, "CDA: concealed data aggregation for reverse multicast traffic in wireless sensor networks," in *Proceedings on the International Conference on Communications*, pp. 3044–3049, Seoul, Republic of Korea, March 2005.

[41] S. Ozdemir and Y. Xiao, "Integrity protecting hierarchical concealed data aggregation for wireless sensor networks," *Computer Networks*, vol. 55, no. 8, pp. 1735–1746, 2011.

[42] S. Papadopoulos, A. Kiayias, and D. Papadias, "Exact in-network aggregation with integrity and confidentiality," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 10, pp. 1760–1773, 2012.

[43] C. Castelluccia, A. C.-F. Chan, E. Mykletun, and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 3, p. 20, 2009.

[44] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proceedings of International Conference on Computer Communications*, pp. 46–50, Phoenix, AZ, USA, April 2008.

[45] J. Shi, R. Zhang, and Y. Zhang, "Secure range queries in tiered sensor networks," in *Proceedings of the International Conference on Information Communications*, pp. 945–953, Rio de Janeiro, Brazil, April 2009.

[46] F. Chen and A. Liu, "SafeQ: secure and efficient query processing in sensor networks," in *Proceedings of the International Conference on Information Communications*, pp. 2642–2650, San Diego, CA, USA, March 2010.

[47] O. Goldreich, *Foundations of Cryptography: Basic Applications*, Cambridge University Press, Cambridge, UK, 2004.

[48] X. Dong, "Secure data aggregation approach based on monitoring in wireless sensor networks," *China Communications*, vol. 3, no. 3, pp. 101–148, 2012.