Secure and Low Energy Consumption Range Query in Tiered Sensor Networks

http://dx.doi.org/10.3991/ijoe.v12i07.5510

Song Ling, Qi Dong-yang
Guangxi University, Nanning, China

***Abstract*—In order to achieve low energy consumption as well as high privacy preservation, a range query protocol called SPRQ was proposed that requires low energy consumption, is secure, and is verifiable. SPRQ uses a novel prime aggregation to protect the privacy of the query data. Furthermore, an and value chain is proposed, whereby data items collected by each sensor will be linked with each other just like a chain. The Sink verifies the integrity of query results by checking whether the data chain of each sensor is complete or not. The results of simulation experiments prove that prime aggregation can effectively reduce the amount of increased data in the prefix encoding process; therefore, the network energy consumption is lower as compared to other secure range query protocols.**

***Index Terms*—energy consumption, integrity, privacy, range query, tiered sensor networks**

I. INTRODUCTION

With the wide application and development of wireless sensor networks (WSNs) [1], the application deployment process is subject to serious privacy data leakages or tampering [2]. Therefore, it is very important to study and solve the problem of large-scale application in wireless sensor networks.

In the actual application process, the wireless sensor is often placed with no supervision in order to increase the service life of the network. A key issue to consider is a reduction in the energy consumption of the whole network. From the previous experiment, it takes about 0.84nJ energy to execute an instruction on the nodes, and the energy consumption of a data transmission is 0.685mJ. This demonstrates that communication energy consumption occupies most of the energy consumption in wireless sensor networks. Therefore, a key inquiry in the application of wireless sensor networks is determining how to reduce the total amount of communication.

To solve these problems, we propose a range query WSNs security protocol SPRQ. SPRQ uses a novel prime aggregation to protect the privacy of the query data and uses AVC (and value chain) to achieve integrity verification of the query results. The results of simulation experiments prove that prime aggregation can effectively reduce the amount of increased data in the prefix encoding process, and as a result, the network power consumption is lower as compared with other secure range query protocols.

II. RELATED WORK

Secure range query contains query process data privacy and integrity verification of the query results. Effective privacy protection protocol requires that:

An attacker cannot record or speculate sensitive information in sensors and cannot destroy the privacy of range query in the sensor network.

If an attacker inserts, tampers, or deletes the query response result to destroy the range query integrity in the sensor network, then the sink node can detect or reject an incorrect or incomplete query response.

In sensor networks, achieving a safe and efficient range query protocol is an important, yet challenging, issue. Li Rui *et al.* [3] proposed order-preserving functions to protect data privacy, and proposed an integrity verification method that included a watermark chain. In 2010, Chen Fei proposed a range query protocol known as SafeQ [4]. SafeQ uses prefix member verification technology to achieve the privacy of the data, through the introduction of the idea of encryption data chain to achieve integrity verification for the results of query. The proposed protocol overcomes the limitations of barrel pattern. In 2013, Dai Hua *et al.* [5] introduced Z-O and HMAC coding technology in range query because storage nodes do not obtain the original data to complete the inquiry process; however, complete query integrity verification is achieved by allowing the sensor node to send a verification code.

One important inquiry in sensor networks is rang queries. Past research in this area has focused on query optimization results to reduce power consumption targets. However, safety and effectiveness have not been focused on in the literature. At the same time, due to the severe limitation of sensor node resources and to increase the service life of wireless sensor networks, sensor nodes should reduce power consumption, the two-tiered network model should be used, and full use of storage nodes should be made in order to obtain query data.

III. NETWORK MODEL

This paper presents a range of privacy preserving protocols based on the two-layer sensor network [6]. As shown in Figure 1, the whole network is divided into multiple units, with each containing a storage node and a large number of sensor nodes. The low level nodes are the sensor nodes that have limited resources, communication, and computing power. As the important connection in the network [7,8], the storage node does not possess such defects. It has a huge storage capacity and powerful computing and communication ability. It also has a steady stream of energy supply. The storage node is responsible for storing the sensor nodes to transmit data and communicating with the base station (the Sink). It receives the Sink query request and then returns the stored information results to the Sink.

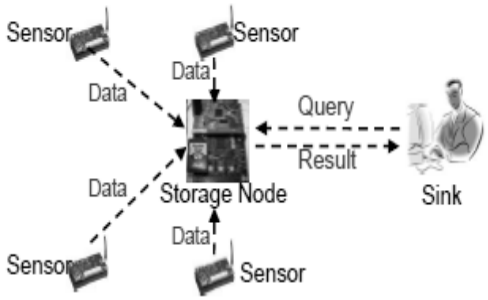

Figure 1. Architecture of two-tiered sensor networks

In two-tiered sensor networks, data gathering is focused on the storage node, which can effectively reduce the network traffic of sensor nodes. Therefore, the nodes of a two-layer sensor network have a long life, easy expansion, and other related advantages; thus, there are broad prospects for development.

IV. SECURE RANGE QUERY PROTOCOL-SPRQ

In the case of ensuring data privacy, reducing energy consumption and improving the WSN service life are key to protocol application. The SPRQ protocol proposed in this paper can reduce energy consumption by reducing the amount of additional data in the encoding process.

A. Key Technology

1) Prefix Encoding Process

Definition 1: For a random number x, binary coded to give $a_1a_2\ldots a_{n-1}a_n$, the prefix set is represented as T(x)={$a_1a_2\ldots a_{n-1}a_n$,$a_1a_2\ldots a_{n-1}$*,… ,a_1*…*,*…*}. Set x=13, T(13)={1101,110*,11**,1***,****}that represents a collection of its prefix.

Theorem 1: For a range [b1, b2], the range prefix of rang set that M [b1, b2]. For example: range prefix of [9, 13] can be expressed as M [9, 13]={100*,11**}. For a random number x and a range [b1, b2], the prefix code verification theory can be drawn□

if and only if T(x)∩M[b1,b2]≠∅→x□[b1,b2]

Prefix encoding verification technology is very clever to transform the comparison between the numerical and the range to the intersection of the set, which is a great help to ensure that the storage node to complete the query process on the privacy protection data. In order to make prefix coding technology has better development and application, and more convenient for solving the problem of privacy protection in data query, the following two aspects still need to be addressed:

Through the analysis of prefix encoding verification techniques , we learn that the simple binary encoding and the method of set comparison are to solve the problem of the relationship between the numerical range and the range, it can not solve the comparison between the two encrypted private data..

For a random number x and a range [b1, b2], assuming their binary coded bits are both n, then n+1 prefixes are required to represent the prefix set for x, and the number of range prefix [9] of [b1, b2] are at least 2n-2. Through the analysis, we know, in order to achieve privacy protection in data query, it needs to transfer a large number of prefix for each sensing data. Due to the huge amount of data in the wireless network, it is difficult to imagine the pressure caused by the huge data prefix to the sensor node which is limited by its own resources. For reducing the network energy consumption and improving the service life of the network, it is necessary to reduce the huge additional communication and storage overhead.

2) Prime Aggregation

In response to the limitations of the prefix encoding verification, this paper proposes a solution that uses prime aggregation for verification. This solution can successfully resolve the fact that a lot of data have the same prefix. It can also reduce network power consumption, and the storage node can query the data without knowing the original data. Additionally, it can ensure the accuracy of query results. The prime aggregation solution transforms all prefixes into prime numbers.

To better express the idea, we assume the prime quadrature of the member prefix to be PT, and assume the prime quadrature of the range prefix to be PM. Thus, the following is concluded:

Theorem 2: Given a random number x and a range [m, n], the conditions x∈[m,n] are established if and only if the following condition is true:

$$\gcd(PT(x),\ PM[m,n]) \neq 1$$

The prime of prefix fi and rj are denoted by pfi and prj, and then:

$$PT(x)=\prod_{i=1}^{P} pf_i \qquad PM([m,n])=\prod_{i=1}^{Q} pr_j$$

The verification of the prefix coding shows that if x∈[m,n], then:

$$T(x)\ \mathrm{I}\ M([m,n])=f_h=r_l\neq\varnothing, h\in[1,P], l\in[1,Q]$$

The following expression can be obtained:

$$pf_h = pr_l$$

The following conclusion is equal to the above expression:

$$x\in[m,n]\rightarrow \gcd\{PT(x)\ ,\ PM[m,n]\}=pf_h=pr_l\neq 1$$

The inverse proposition of the above conclusion can also be proven as the same. Thus, the following is obtained:

$$x\in[m,n]\Leftrightarrow \gcd\{PT(x)\ ,\ PM[m,n]\}\neq 1$$

Before the sensor nodes were deployed to the specific location, every node stored the generated prime inside. To improve the conversion efficiency, it is necessary to convert the prefix to a value. All prefixes are converted to a specific value by the numerical prefix; the detailed operations are as follows:

a) Numerical prefix: Suppose x has a prefix $a_1\ a_2\ldots a_k$*…*, and it has n digits. Here, a 1 is inserted between a_k and the * symbol, and by doing this, the $a_1\ a_2\ldots a_k$ and *…* are separated. Then, all * symbols are replaced with 0. If a prefix does not have a * symbol, then a 1 is inserted at the end of the prefix. For example, {11100}=28 is obtained from the prefix {11**} by numerical prefix. In contrast, if there is no * symbol in the prefix, such as {1010}, then {10101}=21 is obtained.

b) Construct pseudorandom function: In order to ensure that all prefixes can be converted to a specific prime number, a pseudorandom function is employed, and the function factor will change as time t changes. The function can be expressed as follows: $seed_t$=hash($seed_{t-1}$), and a hash() function exists inside. The pseudorandom function

can generate the pseudorandom number, and it is used as an intermediate variable when converting the prefix to a prime number.

The specific process is that in each data query cycle t, a random function will generate a series of random numbers for the sensor nodes. The prefix code is converted to the corresponding prime numbers, and the sensor nodes should first of all be numerical prefix codes. Then, its corresponding numerical value is obtained through a random number after conversion. For example, a numerical prefix code is 9, and the sensor nodes will obtain a random sequence corresponding to the nine random numbers in order to obtain the corresponding prime numbers.

c) Prime aggregation: After all prefixes have been converted to prime numbers, the sensor node multiplies the primes, which are obtained from each sensor data set, and finally, the result is obtained.

The following is an example that further explains the solutions. As in Figure 2, supposse that 12 is the data of the sensor node, and Q[10,15] is the query range of the condition sent by the Sink. We judge whether gcd(PT(12),PM([10,15]))≠1 is true or not. This judgment depends on the storage node's judgment of whether 12 meets the query condition Q[10,15], even if the sensor data is encrypted.

3) Data Submit

Suppose node S_i collects data denoted as $d_i,1,d_i,2$,…,d_i,n , when the Sink sends the query condition [x,y], to ensure that the sensor data does not leak, the sensor data that is converted from the prefix also needs to be encrypted. Additionally, each sensor node and sink node has different keys in different query periods. Suppose that in a query period Si, Sink has key k_i,t, then $k_i,t=hash(k_i,t-1)$ is obtained. Each k_i,t is generated by the hash() function, and at the beginning of each period, the new key replace the old key.

The sensor node collects the sensor data, and after a series process, it then sends it to the storage node, which is used to query:

The sensor node first sorts the collected data and outputs $d_i,1> d_i,2>…> d_i,n$. It is supposed that all sensor data are different from others.

For each sensor data d_i,j, the prefix set $T(d_i,j)$ and query range prefix $M([d_x,d_y]),j∈[1,n]$ are obtained by using prefix encoding, and then, these prefixes are converted to concrete values.

Prime aggregation is conducted for each prefix. Afterwards, prefix set $T(d_i,j)$ will be PTi,j, and the range prefix will be PMx,y.

The sensor node encrypts the sensor data by using the key which is shared by the Sink. After encryption, all sensor data is converted to $ki,t(d_i,1),ki,t(d_i,2),…,ki,t(d_i,n)$. That is to say, the original sensor data is made into privacy data.

Finally, the sensor node sends $\{[k_i,t(d_i,1),PTi,1], [k_i,t(d_i,2),PTi,2], …,[k_i,t(d_i,n), PTi,n]\}$, and the result PMx,y is achieved, which is obtained by prime aggregation of the query range to the storage node.

4) Query Process

When the network needs to carry out the range query, the Sink sends the request Q=(t,[x,y]) to all sensor nodes. The sensor nodes receive the request and use the encryption method, which was expressed in the last section, to encrypt the sensor data and then upload it to the storage node. The storage node receives the encryption data from the sensor node, and in period t, queries the sensor data from the encryption data, which meets the query condition [x,y]. By this, the storage node can finish the query without knowing the original sensor data. Suppose $k(d_i,n)$ is the encryption data of the sensor node, and PMx,y is the result of the query range's prime aggregation. By using theory 2, the storage node can conveniently and efficiently verify whether $k(d_i,n)$ can meet the query range; for verification:

$gcd\{PT(d_i,n),PM[x,y]\}≠1$

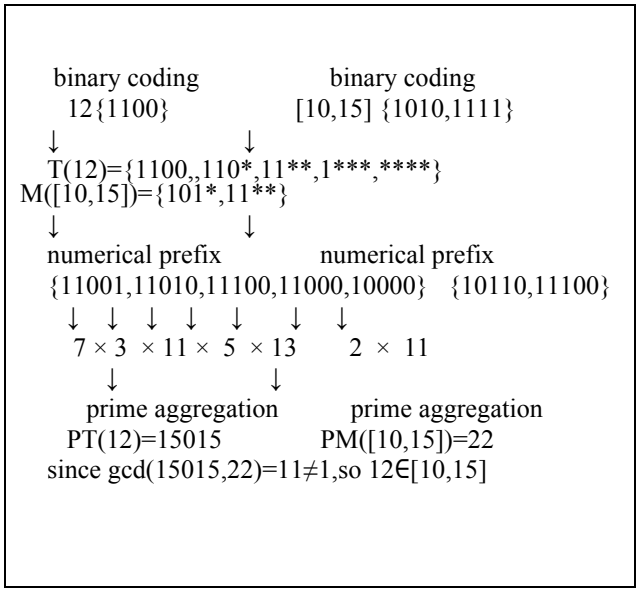

Figure 2. Prime Aggregation Process

If it is true, we know that:

$d_i,n∈[x,y]$

After verification, the storage node can conveniently and efficiently finish the query range without knowing the original sensor node.

5) Search Results Integrity Verification Program - And Value Chain (AVC)

Algorithm 1: AVC

```
Input: sensor nodes Si collect data d[], Maximum and minimum settings dmax, dmin
Output: embedded data link information
sort (d) by descending order; n=d. size
if n=1 then
D[1] ←(dmax+d[1]) | d[1] | (d[1]+ dmin); return
else if n=2 then
D[1] ←(dmax+d[1]) | d[1] | (d[1]+ d[2])
D[2] ←d[2] | d[2] | (d[2]+ dmin); return(D)
else
for i=2: n-1
D[i] ←d[i] | (d[i]+d[i+1])
end for
D[1] ←(dmax+d[1]) | d[1] | (d[1]+ d[2])
D[n] ←d[n] | (d[n]+ dmin); return(D)
end if
```

The basic idea of AVC is that the sensor sorts the collection data and sequentially adds the two adjacent data value. The resulting values are embedded in the previous data. By this means, can make connection between the front and back of the sensing data, and a hidden link is constructed.

Based on the above analysis, the sensor node receives n sensor data in the query period t, and depending on the AVC, data $d_i,1,d_i,2,\ldots,d_i,n$ can be converted to $D_i,1,D_i,2,\ldots,D_i,n$, which adds value information. Then, after the AVC data is encrypted by hash, the new AVC $\{ ki,t(D_i,1),ki,t(D_i,2),\ldots,ki,t(D_i,n) \}$ is obtained, and it is finally uploaded to the storage node for query.

Algorithm 1 describes the process in detail, which converts the sensor data to AVC, and the symbol " | " is used instead of the connect of the sensor data.

B. SPRQ Protocol Query Process

Algorithm 2: A Range Query Protocol-SPRQ

Input: time period t, the query [a,b], collect data from the sensor node $d_1,\ldots, d_n$
Output: search results
Sink→ Storage Node: t, PM([a,b])
Si→Storage Node: i,t,{$[K_{i,t}(D_{i,1}),PT_{i,1}]$, $[K_{i,t}(D_{i,2}),PT_{i,2}],\ldots[K_{i,t}(D_{i,n}),PT_{i,n}]$}
Storage Node: for j=1 to n do
if ($gcd(PT_{i,n},PM([a,b])) \neq 1$)
len++
end if
next j
Storage Node→Sink:i,t,
{$K_{i,t}(D_{i,j})$, $K_{i,t}(D_{i,j+1}),\ldots,K_{i,t}(D_{i,j+n-1})$}

The concrete query process of SPRQ is as follows:

First, the Sink obtains the PM([a,b]) from the query condition [a,b] by prime aggregation, and then, the query condition {t,PM([a,b])} is sent to the storage node to query data.

The sensor nodes sort the collection sensor data and output the d1,d2,…,dn. First, by the AVC, the sensors sequentially add the adjacent data value, the and values are embedded in the previous data. The sensing data is then prefix encoded, and prime integration is conducted, whereby the sensing data can be converted to $[D_{i,1},PT_{i,1}]$, $[D_{i,2},PT_{i,2}],\ldots[D_{i,n}, PT_{i,n}]$. Then, sensor nodes using a shared key k_t with the Sink encrypts the above information to $[K_{i,t}(D_{i,1}),PT_{i,1}]$, $[K_{i,t}(D_{i,2}),PT_{i,2}],\ldots[K_{i,t}(D_{i},n),PT_{i,n}]$, and the information is sent to the storage node.

The storage node receives query instructions, using the formula $gcd(PT_{i,n}, PM([a,b]))\neq 1$ to judge whether the data meets the query condition. If the formula is true, the sensor data is added to the query result; otherwise, the sensor data is dropped.

The Sink decrypts the query result, and thus, the AVC that is hidden in the query result can be obtained in order to verify data integrity in the decryption process. Depending on the results of the AVC integrity verification, the integrity of the query result can be judged.

V. PERFORMANCE ANALYSIS

A. Analysis of Data Privacy

The sensor node S_i uses the key ki,t that is shared with the Sink to encrypt all of the collection sensor data, and then, it sends this encryption data to the storage node. In two-tiered wireless sensor networks, attackers primarily target the storage node because the data in the storage node has been encrypted. Even if attackers target the storage node, they can only obtain the encryption data but not the original data since they do not have the key.

The attackers can estimate the data by embedding the product of the prime aggregation in the sensor data; however, the prefix encoding is randomly converted to a specific prime number in the query period t. The probability of the attackers converting the prefix encoding to the specific prime is 1/(2n+1). Therefore, it can be concluded that it is a low possibility event. Otherwise, the random function that is used to generate the prime is always changed in every period; thus, the prime in each period is changed too, so it will be difficult for attackers to estimate the relationship between them. Therefore, SPRQ can let the storage node accurately complete the query operation. Even though the storage node is compromised, a sensor data leak will not occur as a result.

B. Analysis of the Query Results' Integrity

SPRQ verifies the integrity by using AVC. The Sink and sensor node share a same key kt, which is used to decrypt the query result $\{K_{i,t}(D_{i,j}),K_{i,t}(D_{i,j+1}),...,K_{i,t}(D_{i,j+n-1})\}$. Therefore, the AVC can be obtained from the hidden information after decryption. By analyzing the relationship between the previous and later data, the Sink node can detect the attack means that might change the data or drop part of a query result.

The only attack method requires that the attackers replace all of the chains that meet the query condition. Then, the query result will not satisfy the query conditions. Assuming the sensor node S_i uploaded it satisfactorily, the query data $\{K_{i,t}(D_{i,j}),K_{i,t}(D_{i,j+1}),...,K_{i,t}(D_{i,j+n-1})\}$ is replaced in its entirety for not satisfying the query condition, and sensor node S_j uploads the data $\{K_{i,t}(D_{i,j}),K_{i,t}(D_{i,j+1}),\ldots,K_{i,t}(D_{i,j+n-1})\}$. In response to this attack, the Sink only compares the query results with the query. If the query results are replaced, then they did not satisfy the query conditions, and the Sink can judge this to be false query results. Through the above analysis, it can be found that the integrity verification scheme of SPRQ can ensure the authenticity or integrity of the final query results to be verified by Sink.

C. Analysis of Energy Consumption

The prefix code validation methods can preserve the privacy of the sensing data; however, the sensing data needs to be translated into a large number of prefix sets. This causes a lot of additional network data transmission, resulting in a communication burden to the WSNs. The SPRQ generates the final product of the encoded set corresponding to the values, and it effectively compensates for the lack of a prefix code, thereby reducing the amount of sensing data and thus effectively reducing energy consumption and increasing network life.

D. Simulations

To illustrate the performances of the proposed SPRQ, a range query is selected from a representative security query protocol named SafeQ[4] as contrast objects. SafeQ does not use a conventional barrel range query mode to ensure data privacy; instead, it uses prefixed member verification techniques and verifies data integrity through a data encryption chain. As the presented SPRQ, SafeQ also implements the protections of data privacy and integrity.

During the experiments, the following performances indicators are used to analyze and evaluate the design of the safety query protocol.

Compare the power consumption involved in finishing the query between SPRQ and SafeQ. In such a comparison, SPRQ has an energy advantage over SafeQ.

Compare the efficient query ratio. Such a comparison demonstrates that SPRQ can more efficiently use the energy of sensor network to complete the query.

Under a multi-dimension data condition, compare the power consumption trends of the sensor node and storage node. Such a comparison reveals that SPRQ has an advantage in multi-dimension data query.

The experiment uses MATLAB, and 100 sensor nodes are randomly placed in a 300m * 300m square. The Sink is located in the center, and four storage nodes are evenly placed in the whole sensor network. The classical route algorithm TAG was used to build a router, and the transmit range was set to 75m. The sensor can transmit data to a storage node at an average of 1.8hops.

From Figures 3 and 4, for a sensor node, the power consumption of SafeQ is 3.6 times than of the proposed SPRQ. For a storage node, SafeQ is 2.3 times than our protocol SPRQ.

From Figure 5, the query rate of SPRQ is 2.4 times that of SafeQ, which indicates that SPRQ can complete the query more efficiently because it consumes less power than SafeQ because it efficiently uses network resources.

Figures 6 and 7 illustrate that the node power consumption of SPRQ is showing linear growth even during the query of multidimensional data. It is obvious that with an increase in data dimension, the communication power will inevitably grow, and the normal linear increase of SPRQ proves it can also be applied to multidimensional data queries.

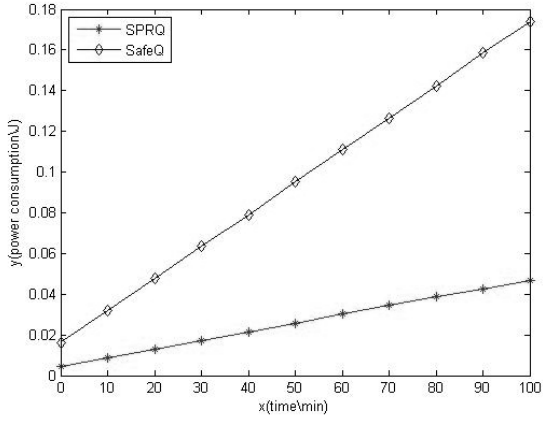

Figure 3. Average power consumption for a sensor

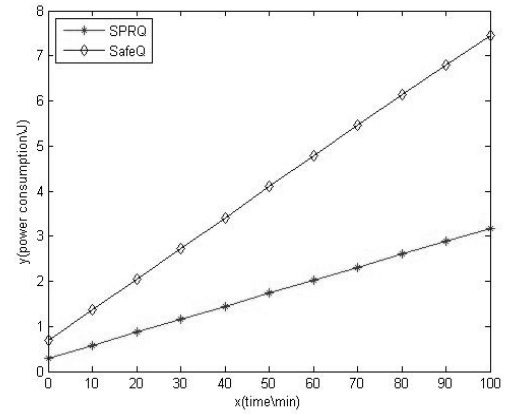

Figure 4. Average power consumption for a storage sensor

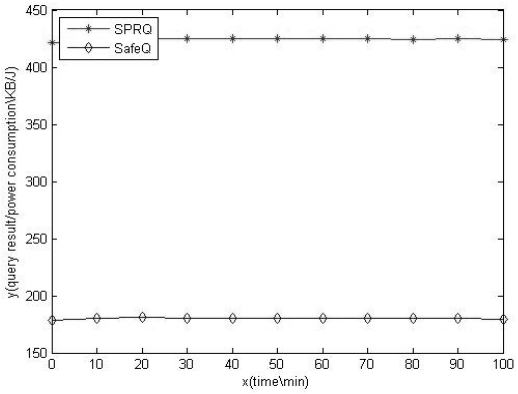

Figure 5. Effective query ratio

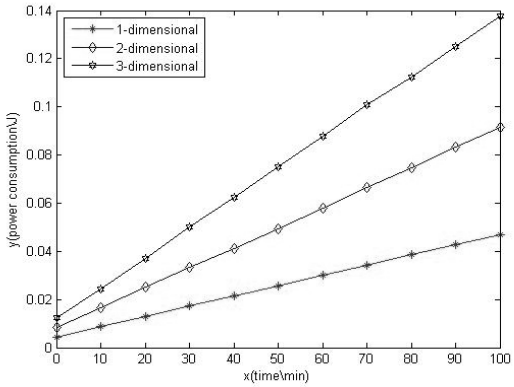

Figure 6. SPRQ's average power consumption for a sensor with different dimensional data

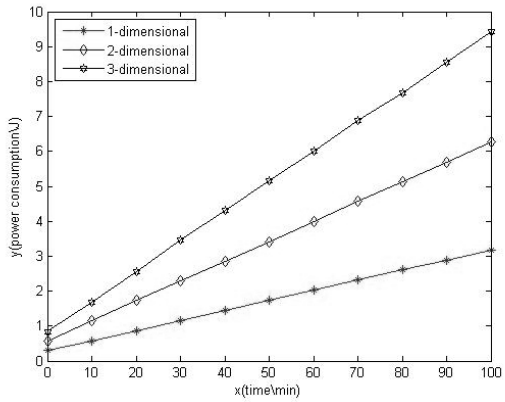

Figure 7. SPRQ's average power consumption for a storage sensor with different dimensional data

VI. CONCLUSION

In this paper, a secure range query protocol SPRQ is proposed. The SPRQ protocol uses prime fusion techniques so that the storage nodes complete the data query under the condition of not perceiving the original data. Thus, the privacy of the data is well protected. At the same time, using the technology of the and value chain, the sink will eventually be able to effectively verify the integrity of the query results. In order to specify the applicability of SPRQ, SafeQ was selected for the contrast experiments. The results reveal that under the premise of protecting query privacy and integrity, by using prime fusion, SPRQ can reduce additional data encoding, thereby reducing the energy consumption of nodes communication and achieving the expected effect.

REFERENCES

[1] Jiang N, Li F, Wan T, et al.Poisson dynamics in fitness evolution model for wireless sensor networks[J]. Journal of Ambient Intelligence & Humanized Computing, 2014,1(5):919-927. http://dx.doi.org/10.1007/s12652-014-0249-4

[2] Yi Y, Li R, Chen F, et al. A digital watermarking approach to secure and precise range query processing in sensor networks[C]. INFOCOM, 2013 Proceedings IEEE. IEEE, 2013:1950-1958.

[3] Li Rui,Lin Ya-pin,Yi Ye-qing,Hu Yu-peng.A Privacy and Integrity Preserving Range Query Protocol in Two-Tiered Sensor Networks[J].Chinese Journal of Computers,2013,36(6):1194-1208.

[4] Chen F, Liu A X. SafeQ: Secure and efficient query processing in sensor networks[C].INFOCOM, 2010 Proceedings IEEE. IEEE, 2010:1-9.

[5] Dou Tie,Huang Hai-ping,Wang Ru-chuan.Secure Range Query in Two-Tiered Wireless Sensor Networks[J].Journal of Computer Research and Development,2013,50(6):1253-1266.

[6] Yao Y, Liu J, Xiong N N. Privacy-preserving data aggregation in two-tiered wireless sensor networks with mobile nodes.[J]. Sensors, 2014, 14(11) :21174-21194 http://dx.doi.org/10.3390/s141121174

[7] Lei, XIE. Towards energy-efficient storage placement in large scale sensor networks[J]. Frontiers of Computer Science in China 2014, 8(3):409-425. http://dx.doi.org/10.1007/s11704-014-2278-8

[8] Bo Sheng, Qun Li, Weizhen Mao. Optimize Storage Placement in Sensor Networks[J]. IEEE Transactions on Mobile Computing, 2010, 9(10):1437-1450. http://dx.doi.org/10.1109/TMC.2010.98

[9] Li J, Lin Y, Wang G, et al. Privacy and integrity preserving skyline queries in tiered sensor networks[J]. Security & Communication Networks, 2014, 7(7):1177-1188. http://dx.doi.org/10.1002/sec.852

AUTHORS

SONG Ling and **QI Dong-yang** are with the School of Computer and Electronics Information, Guangxi University, Nanning, China.

This research was financially supported by the National Natural Science Foundation of China (61363067), the Guangxi Natural Science Foundation (2013GXNSFAA253003) and Key Laboratory of Multimedia Communication and Information Processing of University in Guangxi. Submitted 22 January 2016. Published as resubmitted by the authors 03 March 2016.