Preserving Privacy and Integrity in Wireless Sensor Network using Range Queries

Bhavya B.R
P.G.Scholar
Computer science and Engineering
APS College of Engineering
Bangalore, India
Bhavyabr28@gmail.com

Shobha. N
Asst. Professor Dept of ISE
APS College of Engineering
Bangalore, India
shobha.venk20@gmail.com

***Abstract*—The architecture of two tiered sensor networks where storage nodes acts like intermediate tier between sensors and a sink for storing data and processing queries. The benefits of storage node are saving power and storage space for sensors as well as the efficiency of query processing. The main thing of storage nodes also makes them attractive to attackers. A SafeQ protocol that prevents attackers from gaining information from sensor collected data and sinks issued queries is proposed. It also allows a sink to detect compromised storage nodes when they misbehave. To preserve *privacy*, SafeQ uses a technique to encode data and queries so that a storage node can process a encoded queries over encoded data without knowing their values. To preserve *integrity*, SafeQ uses some techniques to generate integrity verification information. A sink can use this information to check whether the query result contains correct data values tha satisfy the query. To improve *performance*, Bloom filters technique is used to reduce the communication cost between storage nodes and sensors.**

Keywords—Integrity; privacy; range queries; sensor networks.

I. INTRODUCTION

The architecture of two-tiered sensor networks where storage nodes serve as an intermediate tier between sensors and a sink for storing data and processing queries. Here the storage nodes gather data from nearby sensors and answer queries from the sink of the network.

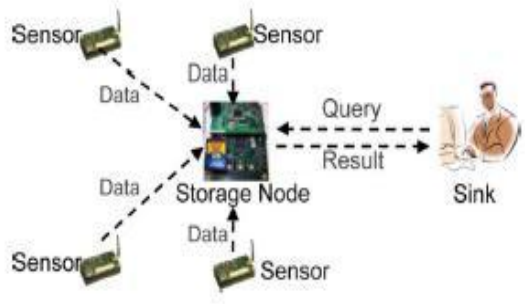

Fig. 1. Architecture of two-tiered sensor networks

Storage nodes bring three main benefits to sensor networks. First, sensors save power by sending all collected data to their closest storage node instead of sending them to the sink through long routes. Second, sensors can be memory limited because data are mainly stored on storage nodes. Third, query processing becomes more efficient because the sink only communicates with storage nodes for queries. The inclusion of storage nodes also brings significant security challenges. As storage nodes store data received from sensors and serve as an important role for answering queries they are more vulnerable to be compromised especially in a hostile environment. A compromised storage node imposes significant threats to a sensor network. First, the attacker may obtain sensitive data that has been or will be stored in the storage node. Second, the compromised storage node may return forged data for a query. Third, this storage node may not include all data items that satisfy the query. To overcome the above mentioned problems a protocol is designed which prevents attackers from gaining information from both sensor collected data and sink issued queries which typically can be modeled as range queries and it also allows the sink to detect compromised storage nodes when they misbehave.

For privacy, compromising a storage node should not allow the attacker to obtain the sensitive information that has been and will be stored in the node as well as the queries that the storage node has received and will receive. For integrity, the sink needs to detect whether a query result from a storage node includes forged data items or does not include all the data that satisfy the query.

There are two key challenges in solving the privacy and integrity preserving range query problem.

First, a storage node needs to correctly process encoded queries over encoded data without knowing their actual values. Second, a sink needs to verify that the result of a query contains all the data items that satisfy the query and does not contain any forged data.

A. *Characteristics and Application of WSN*

There is following characteristics of WSN which are Power consumption constrains for nodes using batteries or energy harvesting, Communication failures, Ability to cope with node failures, Mobility of nodes, Dynamic network topology, Heterogeneity of nodes, Scalability to large scale of deployment, Ability to withstand harsh environmental conditions, Easy of use Unattended operation.

Applications of WSN are:

- Industrial automation
- Automated and smart homes
- Video surveillance
- Traffic monitoring
- Medical device monitoring
- Monitoring of weather conditions
- Military surveillance
- Inventory tracking

II. MODELS AND PROBLEMS STATEMENT

A. *System Model*

A Two tired sensor network consists of three types of nodes: sensors, storage nodes, and a sink. Sensors are inexpensive sensing devices with limited storage and computing power. They are often massively distributed in a field for collecting physical or environmental data example temperature. Storage nodes are powerful wireless devices that are equipped with much more storage capacity and computing power than sensors.

Each sensor periodically sends collected data to its nearby storage node. The sink is the point of contact for users of the sensor network. Each time the sink receives a question from a user it first translates the question into multiple queries and then distributes the queries to the corresponding storage nodes which process the queries based on their data and return the query results to the sink. The sink unifies the query results from multiple storage nodes into the final answer and sends it back to the user.

B. *Threat Model*

A two-tiered sensor network, the sensors and the sink are trusted but the storage nodes are not. In a hostile environment both sensors and storage nodes can be compromised. If a sensor is compromised the subsequent collected data of the sensor will be known to the attacker and the compromised sensor may send forged data to its closest storage node. It is extremely difficult to prevent such attacks without the use of tamperproof hardware.

However, the data from one sensor constitute a small fraction of the collected data of the whole sensor network. Where, a storage node is compromised. Compromising a storage node can cause much greater damage to the sensor network than compromising a sensor. After a storage node is compromised the large quantity of data stored on the node will be known to the attacker and upon receiving a query from the sink the compromised storage node may return a falsified result formed by including forged data or excluding legitimate data. Therefore attackers are more motivated to compromise storage nodes.

C. *Problem Statement*

The fundamental problem for a two tired sensor network is the following: How can design the storage scheme and the query protocol in a privacy and integrity preserving manner? A satisfactory solution to this problem should meet the following two requirements.

1) Data and query privacy: Data privacy means that a storage node cannot know the actual values of sensor collected data. This ensures that an attacker cannot understand the data stored on a compromised storage node. Query privacy means that a storage node cannot know the actual value of sink issued queries. This ensures that an attacker cannot understand or deduce useful information from the queries that a compromised storage node receives.

2) Data integrity: If a query result that a storage node sends to the sink includes forged data or excludes legitimate data the query result is guaranteed to be detected by the sink as invalid. Besides these two hard requirements a desirable solution should have low power and space consumption because these wireless devices have limited resources.

III. MODULES DESCRIPTION

A. *Sensor Module*

Sensor nodes are responsible to collect the data from environment. Each sensor node in a network shares a secret key with the sink. The sensor node encodes the collected data items using key. The sensor node forwards the encoded data to its closest storage node. The collected data are stored into the storage node. Sensor node has limited storage capacity. All the sensor nodes should have capability to collect and store the data at the same time.

B. *Storage Node Module*

Storage nodes are powerful wireless devices that are equipped with much more storage capacity and computing power than sensors. The storage node collects all data from the sensor nodes. The storage node allows only the Authorized user to view the actual value of sensor node data. If any unauthorized user trying to view the sensor node data sink detect misbehave of storage node and the unauthorized user can able to view the encoded data only.

C. Sink Module

The sink is the point of contact for users of the sensor network. Each time the sink receives a question from a user it first translates the question into multiple queries and then disseminates the queries to the corresponding storage nodes which process the queries based on their data and return the query results to the sink. The sink unifies the query results from multiple storage nodes into the final answer. Sink can uses the integrity verification information to verify whether the result of a query contains exactly the data items that satisfy the query and sends it back to the user. Sink can detect compromised storage nodes when they misbehave.

IV. Privacy For 1-Dementional Data

To preserve privacy each sensor S_i encrypts data items d_1 to d_n using its secret key k_i denoted as $(d_1)_{ki}$ to $(d_n)_{ki}$. k_i is a shared secret key with the sink. However, the key challenge is how a storage node processes encrypted queries over encrypted data without knowing their values. The solution is to convert sensor collected data and sink issued queries to prefixes, and then use prefix membership verification to check whether a data item satisfies a range query. To prevent a storage node from knowing the values of data items and range queries, sensors and the sink apply Hash Message Authentication Code (HMAC) to each prefix converted from the data items and range queries.

For example consider sensor collected data {1, 4, 5, 7, 9} and a sink issued query [3,6] in Figure 2. The sensor first converts the collected data to ranges [min, 1], [1, 4] to [9, max] where *min* and *max* denote the lower and upper bound for all possible data items respectively. Second, the sensor converts each range $[d_j , d_{j+1}]$ to prefixes denoted as $p([d_j , d_{j+1}])$ and then apply HMAC to each prefix in $p([d_j , d_{j+1}])$ denoted as $h_g (p([d_j , d_{j+1}]))$. Third, the sensor sends the result to a storage node. When the sink performs query [3,6] it first converts 3 and 6 to prefixes denoted as p(3) and p(6) and then apply HMAC to each prefix in p(3) and p(6) denoted as $h_g (p(3))$ and $h_g (p(6))$. Upon receiving query $h_g (p(3))$ and $h_g (p(6))$ from the sink the storage node checks which $h_g (p([d_j , d_{j+1}]))$ has common elements with $h_g (p(3))$ or $h_g (p(6))$. Based on prefix membership verification, if $h_g (p(a)) \cap h_g (p([d_j , d_{j+1}])) \neq \emptyset$, $a \in [d_j , d_{j+1}]$. Therefore, $h_g (p(3)) \cap h_g (p([1, 4])) \neq \emptyset$ and $h_g (p(6)) \cap h_g (p([5, 7])) \neq \emptyset$. Finally the storage node finds that the query result of [3, 6] includes two data items 4 and 5 then sends $(4)_{ki}$ and $(5)_{ki}$ to the sink.

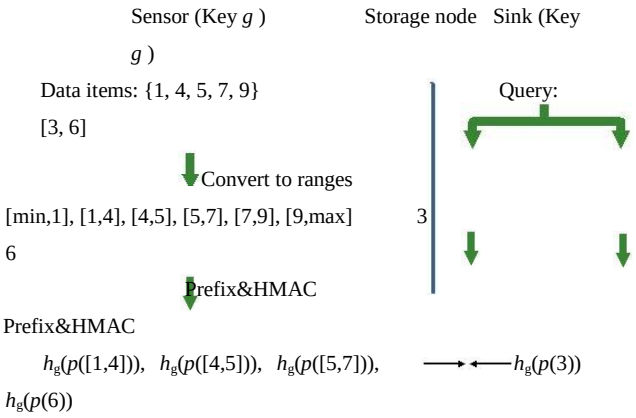

Fig. 2. Privacy preserving scheme of SafeQ

V. Integrity For 1-Dementional Data

To allow the sink to verify the integrity of a query result the query response from a storage node to the sink consists of two parts: (1) the query result QR, which includes all the encrypted data items that satisfy the query. (2) the verification object VO, which includes information for the sink to verify the integrity of QR. To present neighborhood chaining technique to preserve integrity of a query result. The idea of this technique is that instead of encrypting each data item individually a sensor encrypts each item with its left neighbor such that if a storage node excludes any data item that satisfies the query the sink can detect it. Figure 3 shows the neighborhood chain for the sensor collected data in Figure 2. Here "|" denotes concatenation. For the range query [3, 6] the query result QR is $\{(1|4)_{ki} , (4|5)_{ki}\}$ and the verification object V O is $\{(5|7)_{ki}\}$. If a storage node excludes $(4|5)_{kI}$ in QR, the sink can detect this error because the items in QR and V O do not form a neighborhood chain.

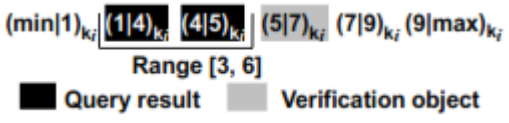

Fig. 3. An example neighborhood chain

VI. Privacy And Integrity for Multi Dimensional Data

To preserve the privacy of multidimensional data to apply a one dimensional privacy preserving techniques to each dimension of multidimensional data. For example sensor S_i collects 5 two-dimensional data items (1,11), (3,5), (6,8), (7,1) and (9,4), it will apply the one dimensional privacy preserving techniques to the first dimensional values {1, 3, 6, 7, 9} and the second dimensional values {1, 4, 5, 8, 11}. Given a range query ([2,6],[3,8]), the query result QR^1 for the sub query [2,6] is the encrypted data items of (3,5),(6,8) and the query result QR^2 for the sub query [3,8] is the encrypted

data items of (9,4),(3,5),(6,8). Therefore the query result QR is the encrypted data items of (3, 5), (6, 8).

To preserve the integrity of multidimensional data to build a multidimensional neighborhood chain. The ideas is that for the value of each dimension in a data item to find its left neighbor along each dimension and embed this information when encrypt the item. Such neighborhood information is used by the sink for integrity verification.

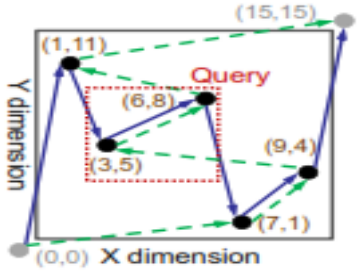

Fig. 4. A two dimensional neighborhood chain

Considering a example of two dimensional data items (1,11),(3,5),(6,8),(7,1),(9,4) with lower bound (0, 0) and upper bound (15,15) the corresponding multidimensional neighborhood chain encrypted with key k_i is $(0|1, 9|11)_{ki}$, $(1|3, 4|5)k_i$, $(3|6, 5|8)_{ki}$, $(6|7, 0|1)_{ki}$, $(7|9, 1|4)k_i$ and $(9|15, 11|15)_{ki}$. Figure 4 illustrates this chain where each black point denotes an item the two grey points denote the lower and upper bounds the solid arrows illustrate the chain along the X dimension and the dashed arrows illustrate the chain along the Y dimension.

VII. SAFEQ OPTIMIZATION

To reduce the communication cost between sensors and storage nodes. For n data items $d_{1....}$ d_n to use a Bloom filter to represent $h_g(p([min, d_1]))$, $h_g(p([d_1, d_2]))$ to $h_g(p([d_{n-1}, d_n]))$, $h_g(p([d_n, max]))$. A sensor only needs to send the Bloom filter instead of the hashes to a storage node. The number of bits needed to represent the Bloom filter is much smaller than that needed to represent the hashes.

Taking $h_g(p([4, 5]))$ and $h_g(p([5, 7]))$ in Figure 2 here we assume that $h_g(p([4,5]))=\{v_1\}$ and $h_g(p([5, 7]))=\{v_2, v_3\}$. $h_g(p([4, 5]))$ and h_g $(p([5, 7]))$ can be represented as the two arrays in Figure 5, where A is a bit array representing the Bloom filter and B is an array of pointers. Each pointer points to a list of indexes of ranges, example 2 is the index of [4,5] and 3 is the index of [5,7].

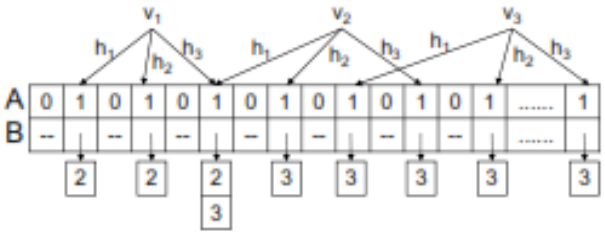

Fig. 5. An example Bloom filter

Although using Bloom filters may introduce false positives in the query result i.e. the data items that do not satisfy the query. in order to control the false positive rate by adjusting Bloom filter parameters.

VIII. SECURITY AND COMPLEXITY ANALYSIS

A. *Privacy Analysis*

In a two tiered sensor network, compromising a storage node does not allow the attacker to obtain the values of sensor collected data and sink issued queries. The correctness of this claim is based on the fact that the hash functions and encryption algorithm. In the submission protocol, a storage node only receives encrypted data items and the secure hash values of prefixes converted from the data items. Without knowing the keys used in the encryption and secure hashing, it is computationally infeasible to compute the actual values of sensor collected data and the corresponding prefixes.

In the query protocol, a storage node only receives the secure hash values of prefixes converted from a range query. Without knowing the key used in the secure hashing it is computationally infeasible to compute the actual values of sink issued queries.

B. *Integrity Analysis*

A two-tiered sensor network, the sink can detect whether the result of a query contains all the data items that satisfy the query and whether it contains forged data. The correctness of this claim is based on the following three properties that QR and VO should satisfy for a query.

First, items in QR ∪ VO form a chain. Excluding any item in the middle or changing any item violates the chaining property. Second, the first item in QR ∪ V O contains the value of its left neighbor which should be out of the range query on the smaller end. Third, the last item in QR ∪ VO contains the value of its right neighbor which should be out of the range query on the larger end.

IX. EXPERIMENTAL RESULTS

Experiments on both S&L and SafeQ scheme are conducted and make used of SafeQ-Basic and SafeQ-Bloom filter to denote SafeQ scheme without and with Bloom filters In terms of power consumption for multi dimensional data SafeQ-Bloom is 184.9 times less power for sensors and 76.8 times less power for

storage nodes SafeQ-Basic is 59.2 times less power for sensors and 76.8 times less power for storage nodes. In terms of space consumption, for multidimensional data SafeQ-Bloom is 182.4 times less space for storage nodes SafeQ-Basic is 58.5 times less space for storage nodes.

The data set was chosen from a large real data set and it consists of the temperature, humidity, and voltage data collected by number nodes.. Each data attribute follows Gaussian distribution. In implementing SafeQ HMAC-MD5 with 128-bit keys is used as the hash function for hashing prefix numbers. DES encryption algorithm is used in implementing both SafeQ and the S&L scheme.

In implementing Bloom filter optimization technique, choose the number of hash functions humidity and three-dimensional data of temperature, humidity, and voltage.

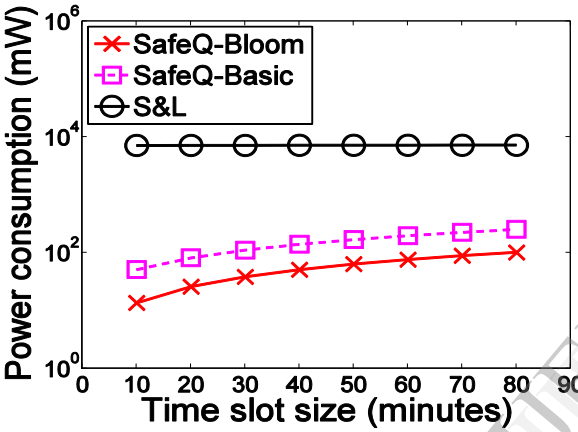

Fig. 6. Power consumption for sensor node

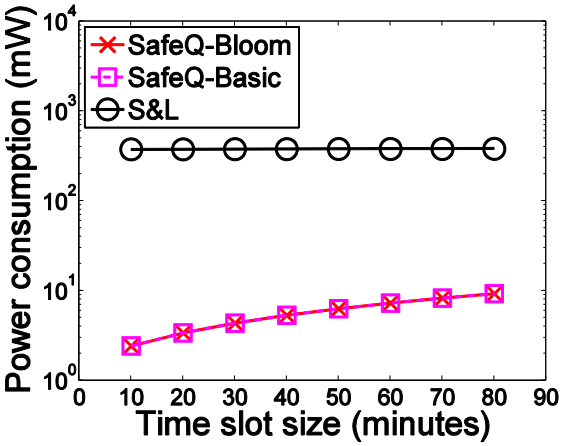

Fig. 7. Power consumption for storage nodes

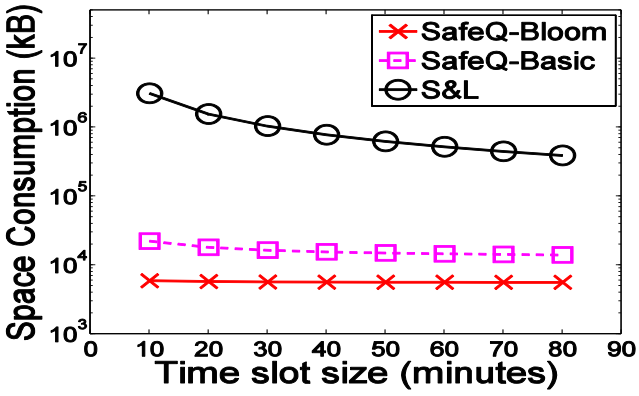

Fig. 8. Space consumption for storage nodes

X. CONCLUSION

A SafeQ, novel and efficient protocol for handling range queries in two-tiered sensor networks in a privacy and integrity preserving fashion is proposed. SafeQ uses the techniques of prefix membership verification and neighborhood chaining. In terms of security, SafeQ significantly strengthens the security of two-tiered sensor networks. SafeQ prevents a compromised storage node from obtaining a reasonable estimation on the actual values of sensor collected data items and sink issued queries. In terms of efficiency results show that SafeQ significantly outperforms for multidimensional data in terms of both power consumption and storage space. An optimization technique using Bloom filters is proposed to significantly reduce the communication cost between sensors and storage nodes.

REFERENCES

[1] F. Chen and A. X. Liu, "SafeQ: Secure and efficient query processing in sensor networks," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[2] S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, R. Govindan, L. Yin, and F. Yu, "Data-centric storage in sensornets with GHT, a geographic hash table," *Mobile Netw. Appl.*, vol. 8, no. 4, pp. 427–442, 2003.

[3] P.Desnoyers,D.Ganesan,H.Li,andP.Shenoy,"Presto:Apredicti ve storage architecture for sensor networks," in *Proc. HotOS*, 2005, p. 23.

[4] D. Zeinalipour-Yazti, S. Lin, V. Kalogeraki, D. Gunopulos, and W. A. Najjar, "Microhash: An efficient index structure for flash-based sensor devices," in *Proc. FAST*, 2005, pp. 31–44.

[5] B. Sheng, Q. Li, and W. Mao, "Data storage placement in sensor networks," in *Proc. ACM MobiHoc*, 2006, pp. 344–355.

[6] B. Sheng, C. C. Tan, Q. Li, and W. Mao, "An approximation algorithm for data storage placement in sensor networks," in *Proc. WASA*, 2007, pp. 71–78.

[7] B.ShengandQ.Li,"Verifiable privacy-preserving range query in twotiered sensor networks," in *Proc. IEEE INFOCOM*, 2008, pp. 46–50.

[8] Xbow, "Stargate gateway (spb400)," 2011 [Online]. Available: http://www.xbow.com

[9] W.A.Najjar,A.Banerjee,andA.Mitra,"RISE:Morepowerful,en ergy efficient, gigabyte scale storage high performance sensors," 2005 [Online]. Available: http://www.cs.ucr.edu/~rise

[10] S. Madden, "Intel lab data," 2004 [Online]. Available: http://berkeley. intel-research.net/labdata

[11] J. Shi, R. Zhang, and Y. Zhang, "Secure range queries in tiered sensor networks," in *Proc. IEEE INFOCOM*, 2009, pp. 945–953.

[12] R. Zhang, J. Shi, and Y. Zhang, "Secure multidimensional range queries in sensor networks," in *Proc. ACM MobiHoc*, 2009, pp. 197–206.

[13] H. Hacigümüs, B. Iyer, C. Li, and S. Mehrotra, "Executing SQL over encrypted data in the database-service-provider model," in *Proc. ACM SIGMOD*, 2002, pp. 216–227.

[14] B. Hore, S. Mehrotra, and G. Tsudik, "A privacy-preserving index for range queries," in *Proc. VLDB*, 2004, pp. 720–731.

[15] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in *Proc. ACM SIGMOD*, 2004, pp. 563–574.