ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)

SafeQ Protocol to preserve Privacy and Integrity Sensor Networks

[1] MR KURAPATI VENKATA PURNACHANDRA PRASAD, [2] RAJU CHEKKA

[1]PG Scholar, Dept. of CSE, VRS & YRN College of Engg. & Technology, Chirala, AP

[2]Assistant Professor, Dept. of CSE, VRS & YRN College of Engg. & Technology, Chirala, AP

Abstract: Two-tier sensor network architecture has been widely adopted because of the benefits of power and storage saving for sensors as well as the efficiency of query processing. In two-tier sensor network storage node serves as the intermediate tier between sensors and a sink for storing the data and for query processing. We propose SafeQ protocol that prevents attackers from gaining information from both sensor-collected data and sink issued queries. When the storage node misbehaves then the SafeQ allows to detecting the compressed storage node. For preserving the privacy, the SafeQ uses a novel technique to encode both data and queries such that a storage node can correctly process encoded queries over encoded data without knowing their values. We propose two schemes to preserve integrity. They are a) one using Merkle hash trees b) new data structure called neighborhood chains. The proposed schemes are to generate integrity verification information so that a sink can use this information to verify whether the result of a query contains exactly the data items that satisfy the query. Moreover, we propose an optimized technique to improve the performance using Bloom filters to reduce the communication cost between sensors and storage nodes.
Keywords: Sensor network, Range query, Integrity, Privacy.

I. INTRODUCTION

A wireless sensor network (WSN) consists of spatially distributed autonomous sensors to *monitor* environmental or physical conditions like pressure, sound, temperature and so on. It has been widely deployed for varied applications, like setting sensing, building safety monitoring, and earthquake prediction and so on. We consider a two-tiered sensor network architecture in which storage nodes gather data from nearby sensors and answer queries from the sink of the network. An intermediate tier between the sensors and the sink serves as the storage node for processing query and the storing data. Storage nodes bring three main benefits to sensor networks.

a. Sensors save power by sending all collected data to their closest storage node instead of sending them to the sink through long routes.
b. Sensors can be memory-limited because data are mainly stored on storage nodes.
c. Query processing becomes more efficient because the sink only communicates with storage nodes for queries.

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. The storage node imposes the significant threats to a sensor network.

- The attackers may obtain sensitive data that has been stored in the storage node.
- The storage node may return the forged data for the query.
- This storage node may not include all data items that satisfy the query.

We want to design a protocol that prevents attackers from gaining information from both sensor-collected data, sink issued queries, and 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 stored in the node. As well as the queries that the storage node has received, and will receive. For *Integrity*, *t*he 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. For solving the privacy and integrity, there are two key challenges.

- A storage node needs to correctly process encoded queries over encoded data without knowing their actual values.
- 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.

SafeQ uses a novel technique to encode both data and queries such that a storage node can correctly process encoded queries over encoded data without knowing their actual values to preserve the integrity. We propose two schemes

- One using Merkle hash trees
- A new data structure called neighborhood chains

We propose a solution to adapt SafeQ for event-driven sensor networks then a sensor submits data to its nearby storage node only when a certain event happens and the event may occur infrequently. Our results show that the power and space savings of SafeQ over prior art grow exponentially with the number of dimensions. SafeQ consumes 184.9 times less power for sensors and 76.8 times less power for storage nodes for three-dimensional data.

II. RELATED WORK

A. Preserving the Integrity and Privacy in WSN's

In our scheme, we use bucket-partitioning idea for database privacy. Privacy- and integrity-preserving range queries in WSNs have drawn people's attention recently. The size of which is computed based on the distribution of data values and the location of sensors. A sensor collects data items from the environment encrypts them together in each bucket and then sends each encrypted bucket along with its bucket ID to a nearby storage node. The bucket is empty, when sensor sends an encoding number to a nearby storage node that can be used by the sink to verify that the bucket is empty. When the sink wants to perform, a range query it finds the smallest set of bucket IDs that contains the range in the query. The S&L scheme has two main drawbacks inherited from the bucket-partitioning technique.

- The bucket-partitioning technique allows compromised storage nodes to obtain a reasonable estimation on the actual value of both data items and queries.
- For multidimensional data as well as the space consumption of storage nodes, the power consumption of both sensors and storage nodes increases exponentially with the number of dimensions due to the exponential increase of the number of buckets.

The basic idea of their optimization is that each sensor uses a bit map to represent which buckets have data and broadcasts its bit map to the nearby sensors. An optimized version of S&L's integrity preserving scheme aiming to reduce the communication cost between sensors and storage nodes. The sink verifies query result integrity for a sensor by examining the bit maps from its nearby sensors.

B. Privacy preservence in the databases

We have observed the proposed bucket partitioning idea for querying encrypted data in the database-as-service model (DAS) where sensitive data are outsourced to an untrusted server. It is used the bucket-partitioning idea to investigate range queries on numerical data. It cannot be used to solve our privacy problem because it is too expensive for sensor networks. Hence, sensor to perform O (zD) encryption for each data submission that is the number of dimensions and is the domain size of each dimension.

C. Preserving integrity in the database

The Merkle hash trees have been used for the authentication of data elements and they were used for verifying the integrity of database queries. In a database signature of the each tuple by signing the concatenation of the digests of the tuple itself as well

ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)

as the tuple's left and right neighbors. Moreover, our neighborhood chaining technique seems similar to the above signature aggregation and chaining technique. For storing the counting information for multidimensional data such that this counting information can be used for integrity verification without leaking boundary information. The result requires each sensor to compute and send an encrypted multidimensional CRT with approximately $n(logD)^z$ overhead to a storage node.

D. Untrusted servers for Secure File System

Our aims to design a system where users can store their files on an untrusted server and the server cannot read the content of the files. The files are not able to process the query in the untrusted server. Our main design goal for SafeQ processing queries in a privacy-preserving manner at storage nodes.

III. PROBLEM STATEMENT

As shown in the fig.1 two-tiered sensor networks as illustrated.

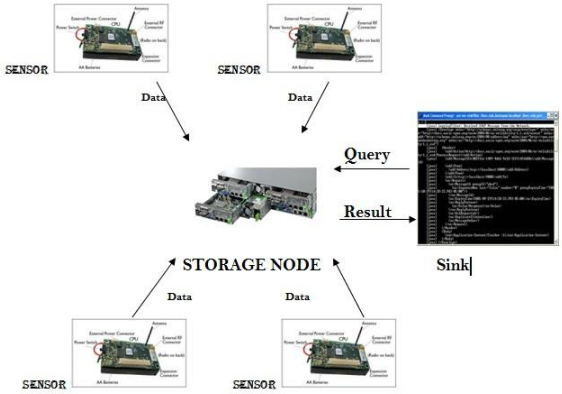


Figure 1. Architecture of two-tiered sensor networks.

There are three types of nodes considered in the two tiered sensor network. They are

a. Sensor
b. Storage node
c. A sink (Base Station)

Limited storage and the computation power for the sensors, which are the inexpensive sensing device. Storage nodes are powerful wireless devices that are equipped with much more storage capacity and computing power than sensors. The contact of user is done by the sink or the base station. First translates the question into multiple queries and then disseminates the queries to the corresponding storage nodes. The query result is unified from the multiple storage nodes into the final answer and sends it back to the user. With loose synchronization and every sensor collects data once per *time interval* we divide time into fixed duration intervals. A *Time-slot* is formed, from the starting time that all sensors and the sink agree upon every time intervals. After a sensor collects data for n times it sends the message that contains a 3-tuple (i, t, {d1----d_n}). A range query "finding all the data items collected at time-slot in the range [a, b]" is denoted as {t, [a, b]}. We address privacy and integrity-preserving ranges queries for event-driven sensor networks.

Symbol	*Description*
s_i	A sensor with ID i
k_i	The secret key of sensor s_i
t	The sequence number of a time-slot
d_1----d_n	n 1-dimensional data items
D_1----D_n	n z-dimensional data items
$\mathcal{H}, \mathcal{G}, \mathcal{E}$	Three "magic" functions
$\mathcal{F}(x)$	The prefix family of x
$S(\|d_1,d_2\|)$	The minimum set of prefixes converted from [d_i, d_2]
$\mathcal{N}$	A prefix numericalization function
$HMAC_g$	An HMAC function with key g
QR	A query result
VO	An integrity verification object

TABLE I: SUMMARY OF NOTATION

We assume that the sensors and the sink are trusted for a two-tiered sensor network. The sensors and the storage nodes are compressed in the hostile environment. 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. Compromising a storage node can cause much greater damage to the sensor network than compromising a sensor. The data from one sensor constitute a small fraction of the collected data of the whole sensor network. The large quantity of data stored on the node will be known to the attacker.

ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)

A falsified result can be returned by the compressed storage node formed by including forged data or excluding legitimate data. In the two-tier architecture the fundamental problem statement is: "*How can we 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. In additional, query privacy means that a storage node cannot know the actual value of sink issued queries.

2. *Data integrity*

If a query result that a storage node sends to the sink includes forged data or excludes legitimate data. Still, the query result is guaranteed to be detected by the sink as invalid.

IV. PRIVACY & INTEGRITY FOR ONE-DIMENSIONAL DATA

Each sensor s_i encrypts data items d1,---,dn using its secret key k_i and it is denoted as $(d_1)k_i$,---,$(d_n)k_i$ to preserve privacy. k_i is a shared secret key with the sink. The key challenge is how a storage node processes encrypted queries over encrypted data without knowing their values. The values of data items and range queries to prevent a storage node from knowing the values. Consider sensor collected data {1, 4, 5, 7, 9} and a sink issued query [3,6] in Fig. 2.

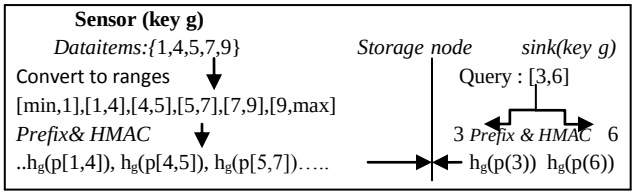


Figure 2: SafeQ Privacy preserving scheme

The sensor first converts the collected data to ranges [min,1], [1,4], ----, [9,max], where min and max denote the lower and upper bound for all possible data items. The sensor converts each range [dj , dj+1] to prefixes and is denoted as p([dj , dj+1]), and then apply HMAC to each prefix in p([dj , dj+1]). The sink performs query first converts 3 and 6 to prefixes. The query denoted as p(3) and p(6) and then apply HMAC to each prefix in p(3) and p(6) and denoted as hg(p(3)) and hg(p(6)). On consequent receiving query hg(p(3)) and hg(p(6)) from the sink the storage node checks which hg(p([dj , dj+1])) has common elements with hg(p(3)) or hg(p(6)). The query result of [3,6] includes two data items 4 and 5 to find in the storage node. The storage node sends (4)ki and (5)ki to the sink.

The query response from a storage node to the sink consists of two parts:

a. *The query result QR*

It includes all the encrypted data items that satisfy the query

b. *The verification object V O*

It includes information for the sink to verify the integrity of QR

Neighborhood chaining technique is presented to preserve integrity of a query result. If a storage node excludes any data item that satisfies the query the sink can detect it, as sensor encrypts each item with its left neighbor.

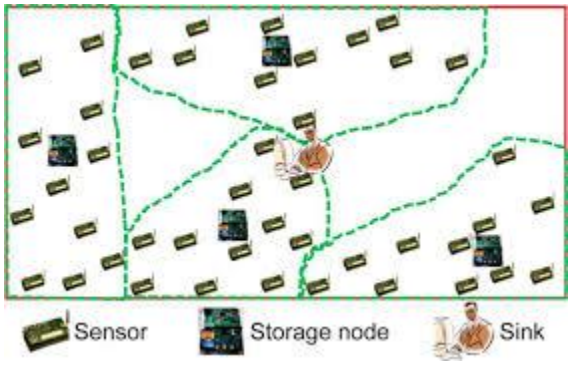


Figure 3: Example neighborhood chain in WSN

As shown in the figure 3 the storage node, sink and sensor shows the how the neighborhood chain is formed. Figure 4 shows the neighborhood chain with the considered example.

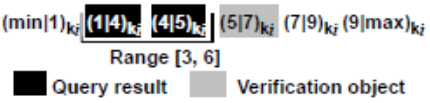


Figure 4: An example neighborhood chain

For the range query [3,6] and the query result QR is {(1|4)ki , (4|5)ki } and the verification object V O is {(5|7)ki }. The items in QR and V O do not form a neighborhood chain when the sink can detect this error.

ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)

V. PRIVACY AND INTEGRITY FOR MULTI-DIMENSIONAL DATA

As in the single dimensional privacy technique, each dimension in multi-dimensional is applied. Sensor s_i collects 5 two-dimensional data items (1,11), (3,5), (6,8), (7,1) and (9,4), it will apply the 1-dimensional privacy preserving techniques to the first dimensional values {1, 3, 6, 7, 9} and the second dimensional values {1, 4, 5, 8, 11}. To preserve the integrity of multi-dimensional data we build a multi-dimensional neighborhood chain. Fig.5 illustrates this chain the two grey points denote the lower and upper bounds.

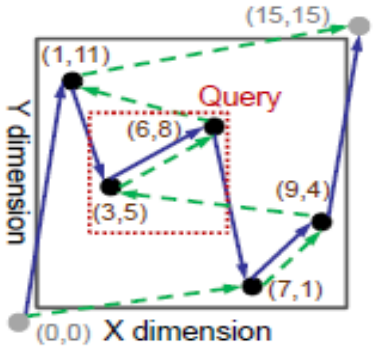


Figure 5: A 2-dimensional neighborhood chain

The dashed arrows illustrate the chain along the Y dimension and solid arrows illustrate the chain along the X dimension.

VI. SAFEQ OPTIMIZATION

To reduce the communication cost between sensors and storage nodes we present an optimization technique based on Bloom filters. This cost can be significant because of two reasons.

- A sensor needs to convert each range [dj , dj+1]
- The sensor applies HMAC to each prefix number that results in 128-bit string

Our basic idea is to use a Bloom filter to represent HMACg (N(S[d0, d1])), ….., HMACg (N(S[dn, dn+1])).A sensor only needs to send the Bloom filter instead of the hashes to a storage node. Consider the data items d_1,….. d_n use a Bloom filter to represent hg(p([min, d1])), hg(p([d1, d2])), ….., hg(p([dn−1, dn])), hg(p([dn,max])). Bloom filter is send instead of the sensor of the hashes to a storage node. Number of bits needed to represent the Bloom filter is much smaller than that needed to represent the hashes. Consider hg(p([4, 5]))={v1} and hg(p([5, 7]))={v2, v3}. We can control the false positive rate by adjusting Bloom filter parameters.

VII. EVENT-DRIVEN NETWORKS QURIES

We have assumed that at each time-slot a sensor sends to a storage node the data that it collected at that time-slot. This assumption does not hold for event-driven networks that a sensor only reports data to a storage node when a certain event happens. The sink cannot verify whether a sensor collected data at a time-slot when if we directly apply our solution. We address the above challenge by sensors reporting their idle period to storage node each time when they submit data after an idle period or when the idle period is longer than a threshold. Hence, storage nodes can use such idle period reported by sensors to prove to the sink that a sensor did not submit any data at any time-slot in that idle period.

Sensors: An *idle period* for a sensor is a time-slot interval [t1,t2] that indicates that the sensor has no data to submit from t1 and t2. Let γ be the threshold of a sensor being idle without reporting to a storage node.

Storage Nodes: When a storage node receives a query from the sink then first it checks weather si has submitted data at time-slot.

Sink: Changes on the sink side are minimal.

VIII. SECURITY&COMPLEXITY ANALYSIS

A protected 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 in the SafeQ. A storage node only receives encrypted data items and the secure hash values of prefixes converted from the data items only in the submission on the protocol. It is computationally infeasible to compute the actual values of sensor collected data, without knowing the keys used the corresponding prefixes in the encryption and secure

ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)

hashing. The key used in the secure hashing is without knowing the computationally infeasible to compute the actual values of sink issued queries. The result of query can be detected by the sink, which contains all the data items that satisfy the query and whether it contains forged data.

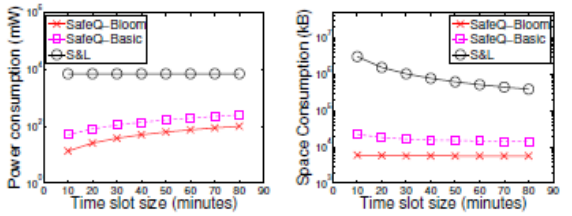
	Computation	Communication	Space
Sensor	O(*zn*) hash O(*n*) encryption	O(*zn*)	_
Storage node	*O(z)* hash	*O(zn)*	*O(zn)*
Sink	*O(z)* hash	*O(z)*	--

Table 2: Complexity analysis of SafeQ

Excluding any item in the middle or changing any item violates the chaining property. Based on the three properties QR and V O, the correctness of this claim satisfy for a query. As shown in the table2 n z-dimensional data items that a sensor collects in a time slot and storage space of SafeQ are described. The communication cost denotes the number of bytes sent for each submission or query.

IX. EXPERIMENTAL ANALAYSIS

Our experimental results shows the SafeQ-Bloom consumes 184.9 times less power for sensors and 182.4 times less space for storage nodes. We implemented both SafeQ and the state-of-the-art on a large real data set. For 2-dimensional data, SafeQ-Bloom consumes 10.3 times less power for sensors and 10.2 times less space for storage nodes. As shown in the fig.6 the average power and space consumption for 3-dimensional.



(a) Sensor: power consumption (b) Storage node: space consumption

Figure 6: Ave. power and space consumption for 3-dimensional data

The three-dimensional shows the safeQ-NC+ consumes 182.4 times less space and SafeQ-MHT+ consumes 169.1 times less space. As shown in the fig.6 the average space consumption of storage nodes for each data item versus the number of dimensions of the data item.

X. CONCLUSION

In our scheme, we mainly contribute three factors. They are SafeQ, Optimized technology using bloom filters and Adaption of safeQ. It is a novel and efficient protocol for handling range queries in two-tiered sensor networks in a privacy- and integrity-preserving fashion. Prefix membership verification, Merkle hash trees, and neighborhood chaining are used in the safeQ. It significantly strengthens the security of two-tiered sensor networks. Our results show that SafeQ significantly outperforms prior art for multidimensional data in terms of both power consumption and storage space. Using the bloom filters, an optimized technology is significantly used to reduce the communication cost between sensors and storage nodes.

XI. REFERENCE

[1] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in Proc. IEEE INFOCOM, 2008, pp. 46–50.

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

[3] J. Shi, R. Zhang, and Y. Zhang, "Secure range queries in tiered sensor networks," in Proc. IEEE INFOCOM, 2009.

[4] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proc. IEEE INFOCOM*, 2008, pp. 46–50.

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

[6] W. A. Najjar, A. Banerjee, and A. Mitra, "RISE:More powerful, energy efficient, gigabyte scale storage high performance sensors," 2005 [Online]. Available: http://www.cs.ucr.edu/~rise

[7] P. Desnoyers, D. Ganesan, H. Li, and P. Shenoy, "Presto:A predictive storage architecture for sensor networks," in *Proc. HotOS*, 2005, p. 23.

ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)