Compressed Bloom Filters for Secure Range Queries in Sensor Networks

[1] S.Rajya Lakshmi ,[2] J.Srinivas Rao

[1] Dept. of CSE, Nova College of Engineerng & Technology,Vijayawada,AP,India.

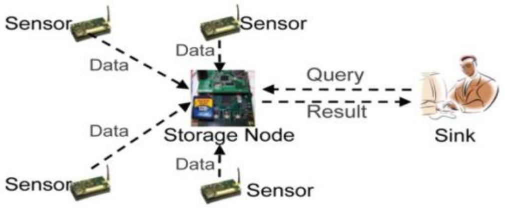
[2]Professor, Nova College of Engineerng & Technology,Vijayawada,AP,India.

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.

Key words: Wireless sensor networks, Privacy preserving, Bloom filters, Compressed Bloom Filter.

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.



Architecture of two-tired sensor networks.

Figure 1: Wireless sensor network architecture

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 eventdriven

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. BACKGROUND WORK

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

Several products of storage nodes, such as StarGate and RISE, are commercially available suggesting their importance. 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. 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.

Traditionally propose to develop SafeQ, a novel privacy- and integrity-preserving range query protocol for two-tiered sensor networks that prevents attackers from gaining information from both sensor collected data and sink issued queries, which typically can be modeled as range queries, and allows the sink to detect compromised storage nodes when they misbehave.

1. The ideas of SafeQ are fundamentally different from the S&L scheme.
2. To preserve privacy, 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 integrity, we propose two schemes—one using Merkle hash trees and another using a new data structure called neighborhood chains—to generate integrity verification information such that a sink can use this information to verify whether the result of a query contains exactly the data items that satisfy the query.

III. PROPOSED APPROACH

Presents an optimization technique based on Bloom filters to reduce the communication cost between sensors and storage nodes. This cost can be significant because of two reasons. First, in each submission, a sensor needs to convert each range query into two variable, where the two variables are two numbers of w bits, to prefix numbers in the worst case. Second, the sensor applies HMAC to each prefix number, which results in a 128-bit string if we choose HMAC-MD5 or a 160-bit string if we choose HMAC-SHA1. Reducing communication cost for sensors is important because of power consumption. Our basic idea is to use a Bloom filter to represent a large data with a small data. Thus, 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. For better performance in terms of speed and computations we suggest to use compressed Bloom filters than plain ones. By using compressed Bloom filters, sensor nodes can reduce the number of bits broadcast, the false positive rate, and/or the amount of computation per lookup. The cost is the processing time for compression and decompression, which can use simple arithmetic coding, and less memory use at the storage nodes, that utilizes the larger uncompressed form of the Bloom filter.

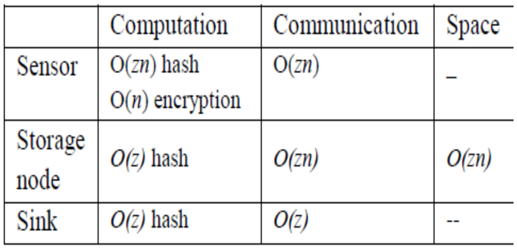
IV. PRIVACY PRESERVING APPLICATION DEVELOPMENT

As in the single dimensional privacy technique, each dimension in multi-dimensional is applied. Sensor si 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. The dashed arrows illustrate the chain along the Y dimension and solid arrows illustrate the chain along the X dimension.

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 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.



	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)*	--

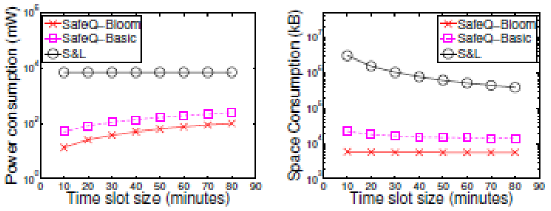
Table 1: Complexity analysis of the sensor networks with privacy preserving applications.

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 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.

V. EXPERIMENTAL RESULTS

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, SafeQBloom 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 2: 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.2 the average space consumption of storage nodes for each data item versus the number of dimensions of the data item.

VI. CONCLUSION

Presents an optimization technique based on Bloom filters to reduce the communication cost between sensors and storage nodes. Our basic idea is to use a Bloom filter to represent a large data with a small data. Thus, 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. For better performance in terms of speed and computations we suggest to use compressed Bloom filters than plain ones. By using compressed Bloom filters, sensor nodes can reduce the number of bits broadcast, the false positive rate, and/or the amount of computation per lookup. The cost is the processing time for compression and decompression, which can use simple arithmetic coding, and less memory use at the storage nodes, that utilizes the larger uncompressed form of the Bloom filter.

VII.REFERENCES

[1] "Privacy- and Integrity-Preserving Range Queries in Sensor Networks", by Fei Chen and Alex X. Liu, IEEE/ACM TRANSACTIONS ON NETWORKING,2012.

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

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

[5] M. Narasimha and G. Tsudik, "Authentication of outsourced databases using signature aggregation and chaining," in *Proc. DASFAA*, 2006, pp. 420–436.

[6] W. Cheng, H. Pang, and K.-L. Tan, "Authenticating multi-dimensional query results in data publishing," in *Proc. DBSec*, 2006, pp. 60–73.

[7] H.Chen, X.Man,W.Hsu, N. Li, and Q.Wang, "Access control friendly query verification for outsourced data publishing," in *Proc. ESORICS*, 2008, pp. 177–191.

[8] R. Merkle, "Protocols for public key cryptosystems," in *Proc. IEEE S&P*, 1980, pp. 122–134.

[9] E.-J. Goh, H. Shacham, N. Modadugu, and D. Boneh, "Sirius: Securing remote untrusted storage," in *Proc. NDSS*, 2003, pp. 131–145.

About Author:

I am S.Rajya Lakshmi completed MCA. Now working as software Engineer in KL University. My Interests are research in data mining,cloud computing etc.

Dr. J.SRINIVAS RAO M.Tech, P.Hd. Received his M.Tech in comuter science & engineering from KL University in 2008, Ph D from CMJ University Meghalaya, INDIA .He is an Outstanding Administrator & Coordinator. He is having 16 years of experience and handled both UG and PG classes. Currently he is working as a Director & Professor in NOVA College of Engineering Technology, Vijayawada, A.P, INDIA .He has Published 30 research Papers in various international Journals and workshops with his incredible work to gain the knowledge for feature errands.