SafeQ: A Privacy and Integrity Preserving Queries in Sensor Networks

Premlatha Khiyani, Dr. Rishi Sayal

Abstract— 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, has been widely adopted because of the benefits of power and storage saving for sensors as well as the efficiency of query processing. However, the importance of storage nodes also makes them attractive to attackers. In this project, we propose Safe Q, a protocol that prevents attackers from gaining information from both sensor collected data and sink issued queries. SafeQ also allows a sink to detect compromised storage nodes when they misbehave. 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 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 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. To improve performance, we propose an optimization technique using Bloom filters to reduce the communication cost between sensors and storage nodes.

Index Terms— SafeQ, Privacy, Sensors, Sink, Query, Data

1 INTRODUCTION

WIRELESS sensor networks (WSNs) have been widely deployed for various applications, such as environment sensing, building safety monitoring, earthquake prediction, etc. In this paper, 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. 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. The inclusion of storage nodes in sensor networks was first introduced and has been widely adopted. Several products of storage nodes, such as Star Gate and RISE, are commercially available. However, 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.

Therefore, we want to design a protocol that prevents attackers from gaining information from both sensor collected data and sink issued queries, which typically can be model eras range queries, and allows the sink to detect compromised storage nodes when they misbehave. For privacy, compromising 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. Note that we treat the queries from the sink as confidential because such queries may leak critical information about query issuers' interests, which need to be protected especially in military applications. 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. Theme of the project

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. Although important, the privacy- and integrity-preserving range query problem has been under investigated. The prior art solution to this problem was proposed by Sheng and Li in their recent seminal work. We call it the "S&L scheme."This scheme has two main drawbacks: 1) it allows attackers to obtain a reasonable estimation on both sensor collected data and sink issued queries; and 2) the power consumption and storage space for both sensors and storage nodes grow exponentially with the number of dimensions of collected data. In this project, we propose SafeQ, a novel privacy- and integrity-preserving range query protocol for two-tiered sensor networks. The ideas of SafeQ are fundamentally different from the S&L scheme. 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 anew 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. We also propose an optimization technique using Bloom filters to significantly reduce the communication cost between sensors and

storage nodes. Furthermore, we propose a solution to adapt SafeQ for event-driven sensor networks, where a sensor submits data to its nearby storage node only when a certain event happens and the event may occur infrequently. SafeQ excels state-of-the-art S&L scheme in two aspects. First, SafeQ provides significantly better security and privacy. While prior art allows a compromised storage node to obtain a reasonable estimation on the value of sensor collected data and sink issued queries, SafeQ makes such estimation very difficult. Second, SafeQ delivers orders of magnitude better performance on both power consumption and storage space for multidimensional data, which are most common in practice as most sensors are equipped with multiple sensing modules such as temperature, humidity, pressure, etc. We performed side-by-side comparison with prior art over a large real-world data set from Intel Lab. Our results show that the power and space savings of SafeQ over prior art grow exponentially with the number of dimensions. For power consumption, for three-dimensional data, SafeQ consumes184.9 times less power for sensors and 76.8 times less power and space consumption in the S&L scheme grow exponentially with the number of dimensions, whereas those in SafeQ grow linearly with the number of dimensions times the number of data items.

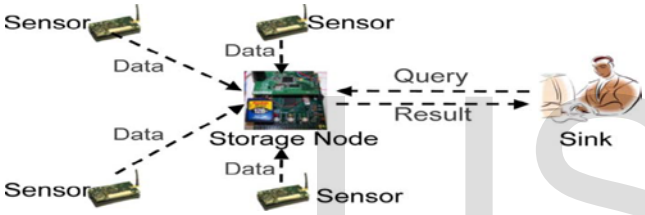


Fig. 1. Architecture of two-tiered sensor networks

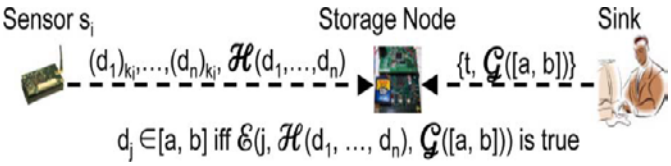
For the above network architecture, we assume that all sensor nodes and storage nodes are loosely synchronized with the sink. With loosely synchronization in place, we divide time into fixed duration intervals and every sensor collects data once per time interval. From a starting time that all sensors and the sink agree upon, every n time intervals form a time slot. From the same starting time, after a sensor collects data for n times, it sends a message that contains a 3-tuple (i, t, {d1, ••• , dn}), where i is the sensor ID and t is the sequence number of the time slot in which the n data items {d1, ••• , dn} are collected by sensor si. We address privacy and integrity preserving ranges queries for event-driven sensor networks, where a sensor only submits data to a nearby storage node when a certain event happens, in Section VII. We further assume that the queries from the sink are range queries. A range query "finding all the data items, which are collected at time slot t and whose value is in the range [a, b]" is denoted as {t, [a, b]}.Note that the queries in most sensor network applications can be easily modeled as range queries

DESCRIPTION

Problem Statement

The fundamental problem for a two-tired sensor network is the following: 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. 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. 3) Given h(d1,d2,…dn)and (Dj)k, it is computationally infeasible for the storage node to compute Dj . This condition guarantees data privacy. 4) Given G [a, b] , it is computationally infeasible for the storage node to compute [a, b] . This condition



guarantees query privacy.

Fig.2. Basic idea of SafeQ for preserving privacy

II. LITERATURE.SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration r taken into account for developing the proposed system.

Overview

Although many randomized asynchronous protocols have been designed throughout the years , only recently one implementation of a stack of randomized multicast and agreement protocols has been reported, SINTRA. These protocols are built on top of a binary consensus protocol that follows a Rabin-style approach, and in practice terminates in one or two communication steps. The protocols, however, depend heavily on public-key cryptography primitives like digital and threshold signatures. The implementation of the stack is in Java and uses several threads. RITAS uses a different approach, Ben-Or-style, and resorts only to fast cryptographic operations such as hash functions.

Randomization is only one of the techniques that can be used to circumvent the FLP impossibility result. Other techniques include failure detectors, partial synchrony and distributed wormholes. Some of these techniques have been employed in the past to build other intrusion-tolerant protocol suites.

III.REVIEW OF REALTED LITERATURE FROM RELATED RESEARCH PAPERS

Data-Centric Storage in Sensor nets with GHT, A Geographic Hash Table.Making effective use of the vast amounts of data gathered by large scale sensor networks (sensor nets) will require scalable, self-organizing, and energy-effcient data dissemination algorithms. For sensor nets, where the content of the data is more important than the identity of the node that gathers them, researchers have found it useful to move away from the Internet's point-to-point communication abstraction and instead adopt abstractions that are more data-centric. This approach entails naming the data and using communication abstractions that refer to those names rather than to node network addresses . Previous work on data-centric routing has shown it to be an energy-efficient data dissemination method for sensor nets . Herein, we argue that a companion method, data-centric storage (DCS), is also a useful approach. Under DCS, sensed data are stored at a node determined by the nameassociated with the sensed data.

In this paper, DCS and predict analytically where it outperforms other data dissemination approaches. We then describe GHT, a Geographic Hash Table system for DCS on sensor nets. GHT hashes keys into geographic coordinates, and stores a key-value pair at the sensor node geographically nearest the hash of its key. The system replicates stored data locally to ensure persistence when nodes fail. It uses an efficient consistency protocol to ensure that key-value pairs are stored at the appropriate nodes after topological changes. And it distributes load throughout the network using a geographic hierarchy. We evaluate the performance of GHT as a DCS system in simulation against two other dissemination approaches. Our results demonstrate that GHT is the preferable approach for the application workloads we analytically predict, offers high data availability, and scales to large sensor net deployments, even when nodes fail or are mobile.

PROBLEM WITH DCS

We have argued for the utility of a DCS service for sensor nets. Now we will define the data-centric storage problem in more detail: the storage abstraction DCS provides, the design goals a robust, scalable DCS system must meet, and our geographic hashing approach to DCS architecture that meets these design goals.

Sensor net Architecture

We organize this discussion in .layers. Ordered from bottom to top. These layers are used only to clarify the presentation and convey a sense of the role of data dissemination in a complete sensor net system; we don't mean to imply that sensor net architecture is organized into clean, well-separated layers. We begin our review at layer three (packet routing),as we are concerned with data dissemination, which interacts directly with layer three and above. Layers one(physical and OS) and two (low-level communication and self-configuration) are comparatively orthogonal to data dissemination.

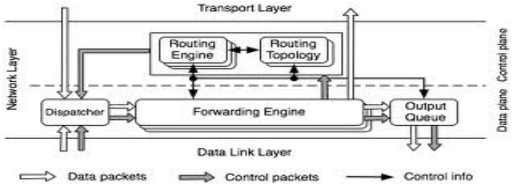


Fig.3.PRESTO: A Predictive Storage Architecture for Sensor Networks

PRESTO ARCHITECTURE

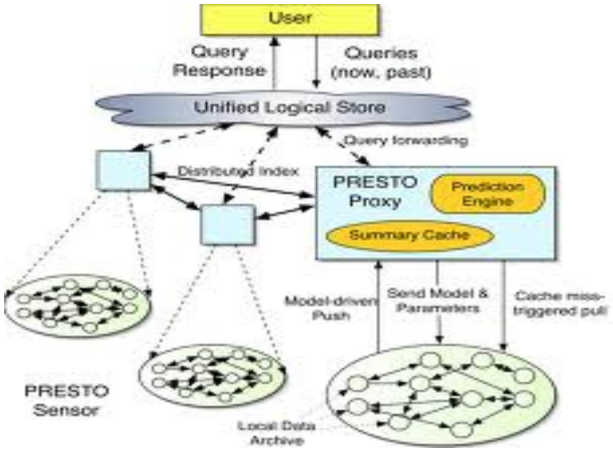


Fig.4 Presto Architecture

Data Storage Placement in Sensor Networks

- Data storage has become an important issue in sensor networks as a large amount of collected data need to be archived for future information retrieval.
- This paper introduces storage nodes to store the data collected from the sensors in their proximities.
- The storage nodes alleviate the heavy load of transmitting all the data to a central place for archiving and reduce the communication cost induced by the net-work query.
- This paper considers the storage node placement problem aiming to minimize the total energy cost for gathering data to the storage nodes and replying queries.
- We examine deterministic placement of storage nodes and present optimal algorithms based on dynamic programming.
- Further, we give stochastic analysis for random deployment and conduct simulation evaluation for both deterministic and random placements of storage nodes.

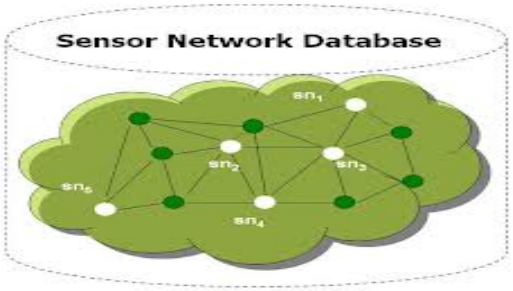


Fig.5. Data Storage Placement in Sensor Networks

IV. Secure Range Queries in Tiered Sensor Networks

We envision a two-tier sensor network which consists of resource-rich master nodes at the upper tier and resource poor sensor nodes at the lower tier.

- Master nodes collect data from sensor nodes and answer the queries from the network owner.
- The reliance on master nodes for data storage and query

processing raises concerns about both data confidentiality and query-result correctness in hostile environments. In particular, a compromised master node may leak hosted sensitive data to the adversary; it may also return juggled or incomplete data in response to a query.

- This paper presents a novel spatiotemporal cross check approach to ensure secure range queries in event driven two-tier sensor networks.
- It offers data confidentiality by preventing master nodes from reading hosted data and also enables efficient range-query processing.
- More importantly, it allows the network owner to verify with very high probability whether a query result is authentic and complete by examining the spatial and temporal relationships among the returned data.
- The high efficacy and efficiency of our approach are confirmed by detailed performance evaluations.

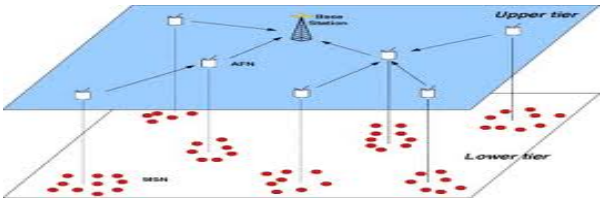
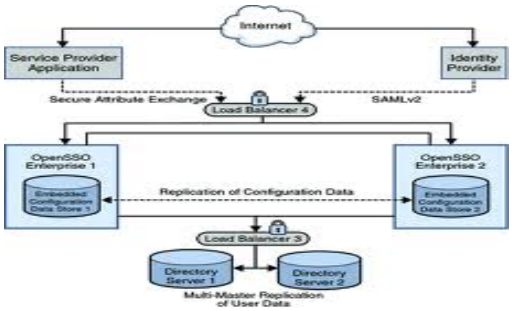


Fig.6 An abstract two-tier sensor network architecture.

Executing SQL over Encrypted Data in the Database Service Provider Model

- Rapid advances in networking and Internet technologies have fueled the emergence of the "software as a service" model for enterprise computing.
- Successful examples of commercially viable software services include rent-a-spreadsheet, electronic mail services, general storage services, disaster protection services.
- "Database as a Service" model provides users power to create, store, modify, and retrieve data from anywhere in the world, as long as they have access to the Internet. It introduces several challenges, an important issue being data privacy.
- It is in this context that we specifically address the issue of data privacy. There are two main privacy issues.
- First, the owner of the data needs to be assured that the data stored on the service-provider site is protected against data thefts from outsiders.
- Second, data needs to be protected even from the service providers, if the providers themselves cannot be trusted.
- In this paper, we focus on the second challenge. Specifically, we explore techniques to execute SQL queries over encrypted data.
- Our strategy is to process as much of the query as possible at the service providers' site, without having to decrypt the data.
- Decryption and the remainder of the query processing are performed at the client site. The paper explores an algebraic framework to split the query to minimize the computation at the client site. Results of experiments validating our approach are also presented

Database Service Provider



V. EXISTING ALGORITHMS

1.Merkle hash tree
2.Neighborhood Chain
3.Bloom filter

The purpose of this Software Requirement Specification (SRS) is to help the project. It is provided with some requirements which are used in Network Intrusion Detection System. All parts; design, coding and testing will be prepared with helping of SRS. The purpose of this document is to detail the requirements used in Network Intrusion Detection System and how the components of the system are to work with each other with external systems.

This document will be checked by group member's supervisor and it will corrected by members if supervisor orders.

Result Summary

SafeQ, a protocol that prevents attackers from gaining information from both sensor collected data and sink issued queries. SafeQ also allows a sink to detect compromised storage nodes when they misbehave. 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 values

Existing System:

Existing Wireless sensor networks once sensor nodes have been deployed, there will be minimal manual intervention and monitoring. But, when nodes are deployed in a hostile environment and there is no manual monitoring.

The prior art solution to this problem was proposed by Sheng and Li in their recent seminal work. It is "S&L scheme."

Proposed System:

The proposed a scheme to preserve the privacy and integrity of range queries in sensor networks .This scheme uses the bucket-partitioning for database privacy. The basic idea is to divide the domain of data values into multiple buckets, the size of which is computed based on the distribution of data values and the location of sensors. In each time-slot, a sensor collects data items from the environment, places them into buckets, encrypts them together in each bucket, and then sends each encrypted bucket along with its bucket ID to a nearby storage node. For each bucket that has no data items, the sensor sends an encoding number, which can be used by the sink to verify that the bucket is empty, to a nearby storage node. When the sink wants to perform a range query, it finds the smallest set of bucket IDs that contains the range in the query, then sends the set as the query to storage nodes. Upon receiving the bucket IDs, the storage node returns the corresponding encrypted data in all those buckets. SafeQ also allows a sink to detect compromised storage nodes when they misbehave. 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.

Proposed a SafeQ, novel privacy- and integrity-preserving range query protocol for two-tiered sensor networks.

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, two schemes has been proposed 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.

CONCLUSION

We make three key contributions in this project. First, we propose SafeQ, a novel and efficient protocol for handling range queries in two-tiered sensor networks in a privacy- and integrity-preserving fashion. SafeQ uses the techniques of prefix membership verification, Merkle hash trees, and neighborhood chaining. In terms of security, SafeQ significantly strengthens the security of two-tiered sensor networks. Unlike prior art, 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, our results show that SafeQ significantly outperforms prior art for multidimensional data in terms of both power consumption and storage space. Second, we propose an optimization technique using Bloom filters to significantly reduce the communication cost between sensors and storage nodes. Third, we propose a solution to adapt SafeQ for event-driven sensor networks.

Future Scope

Wireless sensor networks have been widely deployed for various applications, such as environment sensing, building safety monitoring, earthquake prediction, and also for hostile environments etc. The storage nodes serve as an intermediate tier between the sensors and the sink for storing data and processing queries. A compromised storage node imposes significant threats to a sensor network. Therefore, to avoid such types of attacks, a protocol is designed that prevents attackers from gaining information from both sensor collected data and sink issued queries.

ACKNOWLEDGMENT

This research was supported/partially supported by my HOD. I thank famous persons Rishi Sayal for sharing his pearls of wisdom with me during the course of this research, and although any errors are my own and should not tarnish the reputations of these esteemed persons.

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 sensor nets with GHT, a geographic hash table," Mobile Netw. Appl., vol. 8, no. 4, pp. 427–442, 2003.

[3] P. Desnoyers, D. Ganesan, H. Li, and P. Shenoy, "Presto: A predictive 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. Sheng and Q. Li, "Verifiable privacy-preserving range query in two tiered sensor networks," in Proc. IEEE INFOCOM, 2008, pp. 46–50.

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

[9] 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[10] S. Madden, "Intel lab data," 2004 [Online]. Available: http://berkeley.
intel-research.net/lab data

[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 rangequeries in sensor networks," in Proc. ACM MobiHoc, 2009, pp.