Private And Secure Storage of Wireless Sensor data in potentially insecure Storage Node

Ishwarya S, Dr Arbind Kumar Gupta

Student, Professor
Dept. Of information science and Engg.
Dayananda Sagar College of Engineering, Bangalore, India

Abstract : Privacy and Security playing crucial role in the applications of **two-tiered** sensor networks. The storage nodes, which act as a intermediate node between the source and the sink, which could be compromised so attackers can easily steal sensitive data and calculate query results. Existing schemes on secure query processing are weak, because they reveal **non-negligible** information, and therefore, attackers can statistically measure the data values with the help of domain knowledge and the history of query results. It is challenging to process range query while still protecting sensitive data from dis-closure. Existing work mainly focuses on privacy-preserving range query, but neglects the damage of collusion attacks, probability attacks and differential attacks**.** In this paper, we propose the first **top-*k*** query processing scheme that protects the privacy of sensor data and also provides the security for query results. To preserve privacy, it is necessary to construct an index for each sensor collected data item using **pseudo-random** hash function and Bloom filters and transform **top-*k*** queries into top- range queries. Initially we have to apply mapping technique to sensor data before sending the information to storage node.

***Index Terms* - Privacy, Security, Top-k Query, Wireless Sensor Network, Bloom Filter.**

I. INTRODUCTION

TWO-TIERED sensor networks have been widely adopted for their scalability and energy efficiency. Recently WSN gained much attention from research area due to its advantages in various field such as military, medical field, vehicle surveillance etc.. Generally these network contains several sensor nodes and those nodes are deployed over network region which has to deliver the collected information to the sink node. But here energy efficiency is major problem because sensor nodes are equipped with limited battery power. In order to overcome this issue an intermediate node called storage node has been introduced. Some storage nodes are usually embedded with large storage and powerful computating capacity, are deployed among sensors for storing measurement data from the neighbouring sensors, as shown in Figure 1.

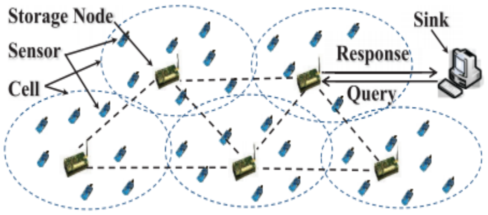

FIG 1. Storage Node Network

Wireless Sensor Networks (WSN) consists of spatially distributed autonomous sensor nodes which are organized into a three different topology such as single hop, multiple hop and two tier hierarchal cluster. wireless sensor network is composed of wireless sensor nodes and a sink node. Nodes are wirelessly interconnected to one another and to the sink. These networks are characterized as Low-power and Lossy Networks (LLNs), as individual nodes possess limited power and operate in harsh environments. If a node is not in direct communication range with the sink, the data it captures is reported in a multi-hop manner. Wireless sensor network are highly distributed network of all small and light weighted nodes, which are spread over the system in large numbers by the measurement of physical parameters like temperature, pressure, relative humidity. *etc.* The storage nodes offer two major benefits compared to an unstructured sensor network model. First, the storage nodes are responsible for the collection, storage and transmission of the sensory data from the sensors to the sink. Second, the storage nodes have more computing power and storage capacity than the sensors. Therefore, the sink can issue complex queries, such as the range or top-k queries, to retrieve several data items in a single query. This saves the sensor nodes' energy and network bandwidth required for answering the sink queries.

II. PROBLEM STATEMENT

Let's see why preserving the privacy of individual sensor readings while transferring the sensor data to sink node through storage node. However, due to storage node importance in network operations, the storage nodes are more vulnerable to attack and compromise. Attackers can not only steal the sensitive information on the storage node, but also leverage the query processing functionality of the storage node to feed false information to the sink. Because protection of privacy gives us add-on benefits including enhanced security and helps to avoid the leakage of secured data.

III. RELATED WORK.

Existing schemes to our approach can be found similar in cloud computing and database domains. Privacy of data can be maintained in several ways and those works can be categorized into three classes: bucketing schemes, order preserving schemes, and public-key schemes. Hacigumus et al. proposed the first bucket partition scheme to query encrypted data items without involving the server to know the exact data values. Hore et al. investigated the problem of optimal bucket partitioning and proposed two secure query schemes, one for one-dimensional data and the other for multi-dimensional data Agrawal et al. adopted the bucketing scheme's idea and proposed an order preserving scheme to further protect data privacy Boldyreva et al. proposed two order preserving schemes.

However, end-to-end encryption or link level encryption alone is not a good apporach for private data preserving. This is because: 1) If end-to-end communications are encrypted, the intermediate nodes or storage nodes could not easily perform in-network processing to get aggregated results. 2) Even when data are encrypted at the link level, the other end of the communication is still able to decrypt it and get the private data. Hence privacy using encryption technique is violated.. In [30], Li et al., propose a privacy preserving range query processing scheme for outsourced data items in cloud computing, which is proved to be secure under IND-CKA security model. But the major disadvantage is this scheme cannot perform top-k querying with integrity verification for sensor networks.

In the public key cryptography based schemes, Boneh Waters proposed a database privacy-preserving scheme to support conjunctive, subset and range queries . Shi et al. proposed a range query scheme using identity based encryption primitives. However, public-key cryptography is generally unaffordable in two-tiered sensor networks for its computational complexity (software implementation) and system setup/maintenance cost (hardware implementation).may leads to system complexity.
A significant amount of work has been proposed to pre-serve privacy and security for query results in two-tiered sensor networks, All of these works require abundant information and an additional verification mechanism along with the query results. We show that our privacy verification mechanism is less expensive than these approaches.

IV.PROPOSED SYSTEM

We propose the first top-k query processing scheme that protects the privacy of sensor data and the integrity of query results. To preserve privacy, we build an index for each sensor collected data item using pseudo-random hash function and Bloom filters and transform top-k queries into top range queries. To preserve security and integrity, we propose a data partition algorithm to partition each data item into an interval and attach the partition information with the data. The attached information ensures that the sink can verify the integrity of query results.

IV.A. Sensor Data Pre-Processing Mapping And Partitioning

Our data processing involves four important steps: approximating uniform data distribution, data partitioning, interval embedding and index selection. First, we describe a method to transform the arbitrary sensor data into an approximate uniform data distribution. Second, we describe a data partitioning algorithm, for security as well as integrity. Third, we show the method to embed the interval information required for security verification of the query results. At last, we show the index value selection for building the secure index for top-*k* querying. We address the problem of privacy preserving top-*k* queries in two-tiered sensor networks to protect against storage node compromise. Our aim is to design scheme to enable storage nodes to process top-k queries correctly without knowing the actual value of data stored in them and allow the sink to detect misbehavior of storage nodes. Top-k query processing, *i.e.*, finding the *k* smallest or largest data items collected from a specified sensed area, is a fundamental operation in sensor networks [8]–[10]. Such top-k queries enable users to get their most desired environmental information such as pollution index, temperature, humidity and so on.

IV.B. The System architecture is shown below.

System architecture is the conceptual design that defines the structure and behavior of a system. An architecture description is a formal description of a system, organized in a way that supports reasoning about the structural properties of the system.

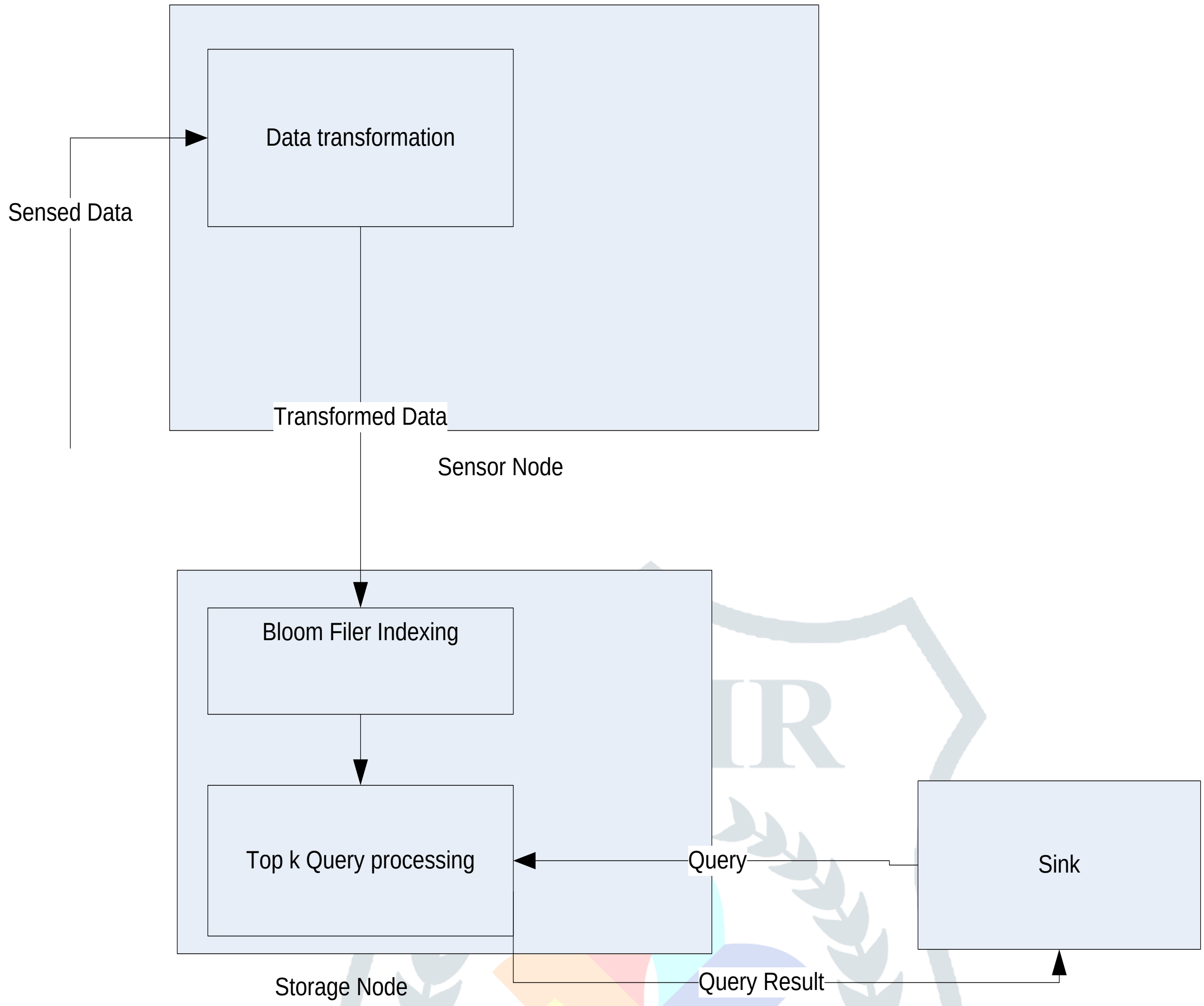

Fig 2: System Architecture

There are two important components in the architecture
1) Sink
2) Sensor Node

Sink: Sink creates Ring clustering of network based on the position information of nodes. After clusters are established, energy efficient cluster heads are selected for each cluster and sensor nodes advertised about the cluster head node.

Sensor Node: Sensor node generated packets at periodic intervals and forward via energy based routing module to the storage node. Where the data will be transferred in the form of mapped data. Let's consider an scenario where temperature and humidity sensors senses the temperature and humidity .then collected data should be mapped with mapping technique .later the storage node has to answer for the query sent by sink node without unmapping the stored mapped data.

IV. B. Data Partitioning for Integrity Verification and Index Selection

In this section, we describe our approach to verify the security of the results returned due to the top-*k* query by the sink node. Intuitively, integrity and security verification approach is to partition the data into distinct intervals and securely embed the interval information along with the data item. Note that, due to the transformation of the sensor data into uniform distribution, the indexing is performed on the new index values derived from the uniform distribution. However, even though the top-*k* query will be run over the new data items, the storage node will send back the corresponding original data items to the sink node. This implies that the sink node needs to be able to verify that received data items correspond to the mapped data items, on which the top-*k* query was executed.

Towards this, we describe an efficient data partition algorithm that partitions the new data items, in the uniform distribution, into different intervals depending on the data item values. At the time of index generation, we embed this interval information inside the original sensor data items returned to the sink node and use this information to verify the security as well as integrity. since sink node is aware about the uniform distribution it will be able to derive the uniform data respectively from the storage node. We next give the details of partitioning. query result is included in the result, we need to include the interval information of all data items covered by the top-k query. Our idea is to select the lower bound value of the interval containing the index data item.
The rationale is that, a suitably chosen prefix respectively to the lower bound value of the interval can be used to represent the entire range of the interval. Consider Figure 4 where six data items [2, 5, 6, 7, 10, 11]are partitioned into six intervals [0, 3], [4, 5], [6,

6], [7, 7], [8, 10], [11, 15], respectively, using our partitioning algorithm. For instance, to index 2, we consider the interval [0, 3] and choose 0 as the indexing point. This is because the prefix 00 ∗ ∗, generated from the 4-bit binary representation of 0, can represent the entire interval [0, 3]. We use $d^m{}_i$ to denote the index of data d_i. Although the method in can be extended in some straightforward way to the method with data confidentiality guarantee, such extension actually implies some of the other severe weaknesses, which are unacceptable in the design of a verifiable query scheme.

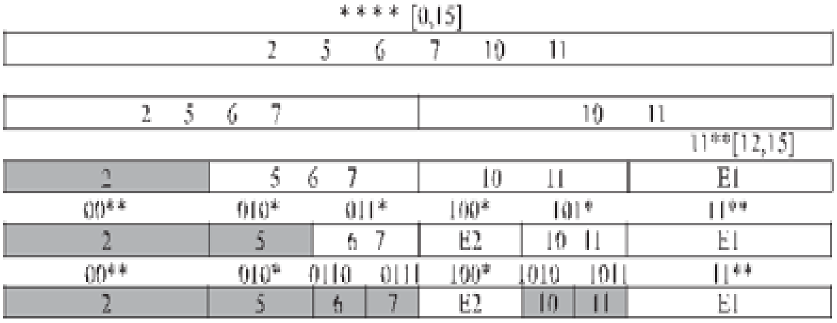

FIG.3. DATA PARTITIONING ALGORITHM

```
Input [do,dn] //DataSet
Output[do',dn'] // Generated new data after partitioning
low =do'; High=dn' ;
While p={*};
RandRow P*rad()
If po=0****, p1=1****€ p
{
po=0**0 or po=0**1 ;
P1=1**0 or p1=1**1;
}
P(i,all column)=DS(single row/single Column);
End if
Next i
Output//Generated data
```

Consider the case that the sensor readings are encrypted by popular encryption functions, like DES and AES. In this case, the storage node will not be able to answer the top-*k* query issued by the authority due to the lack of the numeric order of sensor readings. On the other hand, consider the case that order-preserving encryption (OPE) is used to encrypt sensor readings. In this case, the numeric order of sensor readings is preserved. Nevertheless, this is achieved by all of the sensors sharing a common OPE key. A consequence of doing so is that once a sensor is compromised, the OPE key is exposed to the effect and the data confidentiality is completely breached. Above two methods offers data confidentiality but fails to answering for the data query or resilience against the compromised sensor.

Our approach to executing top-*k* queries on the sensor data is to convert the top-*k* query into a *top-range* query. The aim behind this is that, directly performing top-*k* queries on a set of sensor data items requires comparisons among them. But, given the prefix membership scheme from the previous section, we can check if a particular range of query prefixes are matched by any of the stored prefixes in the Bloom filter. This forms the basis of our approach t we transform the top-*k* query into a suitably crafted range query, which will obtain the same results as the top-*k* query. We next describe the details of this transformation. Finally, upon receiving a query result of the query trapdoor $Q_{[d0,d\ ]}$ from a storage node, the sink verifies the integrity for the query result as the follows. The sink divides the received data items into groups according to their source sensors. Next, the sink recovers the interval information for these data items The experimental results show that *our proposed scheme has practical accepted bandwidth cost*. The experiment results indicate that our scheme consumes 0.55 times more bandwidth than QuerySec and spends 1.22 times less bandwidth of SafeQ. The experimental results show that *our proposed scheme has practical accepted bandwidth cost.* Figure 5 shows the average bandwidth cost between a sensor and its nearby storage node in a time slot. The size of time slot varies from 10 minutes to 80 minutes. The experiment results indicate that our scheme consumes 0.55 times more bandwidth than QuerySec and spends 1.22 times less bandwidth of SafeQ. Figure 5 shows the average bandwidth cost of a top-*k* query between the sink and 4 storage nodes, using three different schemes. Figure 4 shows the experiment result that we fix *k* = 100 and time slot size ranging from 10 minutes to 80 minutes, and figure shows the experiment result that we fix time slot *t* = 40 minutes and *k*

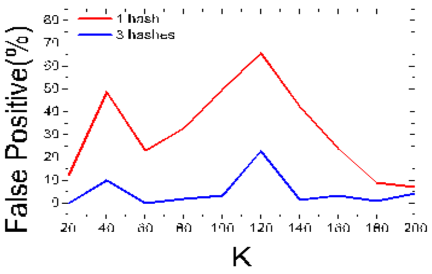

Fig 4 False rate from bloom filter Configuration

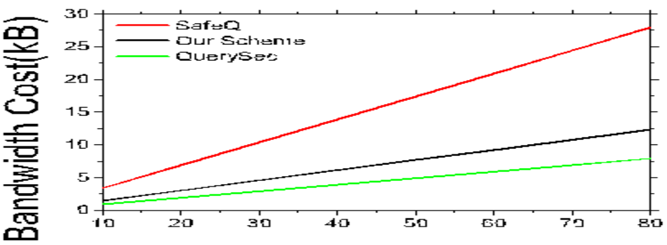

Fig 5 Comparison of average bandwidth cost of various Methods

V .CONCLUSION

In this paper, we propose the first secure top-*k* query processing scheme that is secure under the IND-CKA security model. The data privacy is guaranteed by encryption as well as a careful generation of data indexes. We make two key contributions in this paper. The first contribution is to transform a top-*k* query to a top-range query and adopt membership testing to test whether a data item should be Included in the query or not. The second contribution is the data partition, index selection, and interval information embedding technique. This technique guarantees that at least one data item of each sensor collected data will be included in a query result and allows the sink to verify the integrity of query result without extra verification objects.

VI. REFERENCES

[1] Z. Zhou, Y. Wang, Q. M. J. Wu, C.-N. Yang, and X. Sun, "Effective and efficient global context verification for image copy detection," IEEE Trans. Inf. Forensics Security, vol. 12, no. 1, pp. 48–63, Jan. 2017.
[2] Privacy and Integrity Preserving Top-k Query Processing for Two-Tiered Sensor Networks Rui Li, Alex X. Liu, Sheng Xiao, Hongyue Xu, Bezawada Bruhadeshwar, and Ann L. Wang IEEE AUG-2017
[3] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multikeyword ranked search scheme over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol 27, no. 2, pp. 340–352, Feb. 2016.
[4**]** C.-M. Yu, G.-K. Ni, I.-Y. Chen, E. Gelenbe, and S.-Y. Kuo, "Top-k query result completeness verification in tiered sensor networks," IEEE Trans. Inf. Forensics Security, vol. 9, no. 1, pp. 109–124, Jan. 2014.
[5]X. Liao and J. Li, "Privacy-preserving and secure top-k query in two tier wireless sensor network," in Proc. 55th GLOBECOM, Dec. 2012, pp. 335–341.
[6]J. Chen, G. Wu, L. Shen, and Z. Ji, "Differentiated security levels for personal identifiable information in identity management system," Expert Syst. Appl., vol. 38, no. 11, pp. 14156–14162, Oct. 2011.
[7] R. Zhang, J. Shi, and Y. Zhang, "Secure multidimensional range queries in sensor networks," in Proc. 10th ACM MobiHoc, May 2009, pp. 197–206.
[8] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two tiered sensor networks," in Proc. 27th INFOCOM, Apr. 2008, pp. 46–50.