

# Three Tier Layered Approach in Wireless Sensor Networks with Secured K-Top Query

Mr. C.N.S.Vinoth kumar <sup>1</sup>  
Assistant Professor  
Arunai College of Engineering

Dr. A.Suhasini <sup>2</sup>  
Associate Professor  
Annamalai Univeristy

**Abstract** - Wireless Sensor Networks (WSN) are essentially difficult to secure with the pre-key distribution or certificate flags. Using pair wise key distribution and authentication between the sensor nodes, an attacker can easily get a large number of key matches and gain control by network arranging with some compromised keys. The context has two separate key pools, mobile sink for access the network, another for pairwise key distribution between the sensors. To reduce the damages caused by access node replication attacks with pair key distribution, we have reinforced the authentication mechanism with the secure K-top query processing performed on the time-slot sensing data set in three tier sensor network, and establish a set of privacy and correctness requirements for such a secure top-k query scheme to become a reality. We show that our security framework has higher network flexibility to a mobile sink replication attacks.

**Index Terms**—Distribution, Wireless Sensor networks

## I. INTRODUCTION

Wireless sensor networks become prevalent for pervasive computing and widely deployed for various applications such as intrusion detection, earthquake prediction and environment sensing, etc... It usually consists of resource limited nodes in terms of storage capacity and computing capacity. Various inherent limitations of it appear especially resource constraints which limit the storage capacity of sensing data and the computing capacity of processing query. These limitations deserve special attention especially in the remote and extreme environment where a high-speed and always-on connection is infeasible. Extensive research has been conducted to address these limitations by developing a three-tier wireless sensor network where the storage node is introduced. Several commercial storage nodes have appeared, such as Star Gate.

A three-tier wireless sensor network consists of large amount of resource-limited sensor nodes at the lower-tier which sense the environment information and plenty of resource rich relatively storage nodes at the upper-tier which gather data from the nearby sensor nodes and answer queries from the user. While storage nodes bring several benefits such

as prolonging network lifetime, saving the memory of sensor nodes and processing queries efficiently, the storage node faces serious security concerns in hostile environment. First, when the storage node is compromised, the sensing data from the sensor nodes, the history of query requests and the corresponding query results are exposed. Second, it may cause heavy loss when the compromised sensor nodes return fake, forged or incomplete data for a query especially in military and commercial application. Therefore, developing a privacy preserving and result-verifiable mechanism is of paramount importance such that the authenticity and completeness of the query results can be verified as well as the privacy of the sensitive data is protected.

In many of these applications, sensor nodes transmit critical information over the network; therefore, security services, such as, authentication and pairwise key establishment between sensor nodes and mobile sinks, are important. However, the resource constraints of the sensors and their nature of communication over a wireless medium make data confidentiality and integrity a nontrivial task. Traditional schemes in ad hoc networks using asymmetric keys are expensive due of their storage and computation cost. These limitations make key pre-distribution schemes the tools of choice to provide low cost, secure communication between sensor nodes and mobile sinks.

However, the problem of authentication and pairwise key establishment in sensor networks with MSs is still not solved in the face of mobile sink replication attacks. For the basic probabilistic and q-composite key pre-distribution schemes, an attacker can easily obtain a large number of keys by capturing a small fraction of the network sensor nodes, making it possible for the attacker to take control of the entire network by deploying a replicated mobile sink, preloaded with some compromised keys to authenticate and then initiate data communication with any sensor node.

To address the above-mentioned problem, we have developed a general framework that permits the use of any pairwise key pre-distribution scheme as its basic component, to provide authentication and pairwise key establishment between sensor nodes and mobile sinks. To facilitate the study of a new security technique, we first cultivated a general three-tier security framework for authentication and pairwise

key establishment, based on the polynomial pool-based key pre-distribution scheme. The proposed technique will substantially improve network resilience to mobile sink replication attacks compared to the single polynomial pool-based key pre-distribution approach

Recently various secure query schemes have been proposed in three-tier sensor network aiming at the range query and top-k query. However they didn't address the privacy problem of top-k query in three-tier wireless sensor network. In addition, all the secure query schemes are performed on the data set which consists of sensing data sampled from several successive time slots by all sensor nodes, referred to as a time interval. Sensor nodes submit to the nearest storage node all the data sensing during the time interval that the storage nodes process query in the time interval data set. As shown in the analysis, it is a general case in many real-world applications where the query is performed on the data set in which each data items are from the sensor nodes sampling and submitting at one single time slot, referred to as a time slot data set such as Intel Lab. The schemes they proposed are in client and introduce large amount of communication consumption when performed on the time slot data set.

In this paper, for the first time, we propose a privacy preserving and query-verifiable top-k query mechanism on the time slot data set in three-tier wireless sensor network while achieving energy efficiency. Among various top-k query semantics, we choose the single score function model that each data can be scored by the scoring function and ranked based on its score. In consideration of three levels of threat models, we propose the basic Pri-Sec Top k scheme by using order preserving encryption, and then improve it step by step to achieve various privacy requirements as well as the correctness requirements.

**II. RELATED WORK**

The key management problem is an active research area in wireless sensor networks. Eschenauer and Gilgor [12] proposed a probabilistic key pre-distribution scheme to bootstrap the initial trust between the sensor nodes. The main idea was to let each sensor node randomly pick a set of keys from a key pool before deployment, so that any two sensor nodes had a certain probability of sharing at least one common key. Chan et al. [13] further extended this idea and developed two key pre-distribution schemes:

The q-composite key pre-distribution scheme and the random pair-wise keys scheme. The q-composite key pre-distribution scheme also used a key pool, but required two sensor nodes to compute a pair-wise key from at least q pre-distributed keys that they shared. The random pair-wise keys scheme randomly picked pairs of sensor nodes and assigned each pair a unique random key. Both schemes improved the security over the basic probabilistic key pre-distribution scheme.

The pair-wise key establishment problem, however, is still not solved. For the basic probabilistic and the q-composite key pre-distribution schemes, as the number of compromised

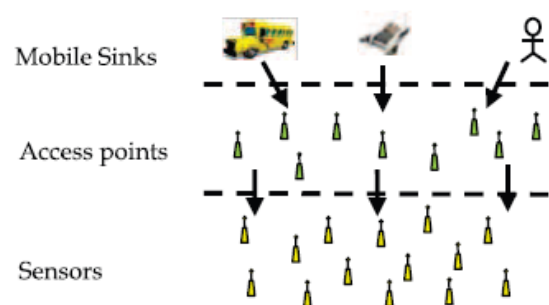
nodes increases, the fraction of affected pair-wise keys also increases quickly. As a result, a small number of compromised nodes may affect a large fraction of pair-wise keys. Although, the random pair-wise key does not suffer from the above mentioned problem, given a memory constraint, the network size is strictly limited by the desired probability that two sensor nodes share a pair-wise key, as also by the number of neighbor nodes with which a sensor can communicate. An enhanced scheme using the t-degree bi-variant key polynomial was proposed by Liu et al.

They developed a general framework for pair-wise key establishment using the polynomial-based key pre-distribution protocol and the probabilistic key distribution. Their scheme could tolerate no more than t compromised nodes, where the value of t was limited by the memory available in the sensor nodes.

**III. THE THREE-TIER SECURITY SCHEME**

In this study, we have chosen the Blundo scheme to construct our approach. As we shall see, the Blundo scheme provides a clear security guarantee. Use of the Blundo scheme, therefore, greatly eases the presentation of our study and enables us to provide a clearer security analysis. In the proposed scheme, we use two separate polynomial pools: the mobile polynomial pool and the static polynomial pool. Polynomials from the mobile polynomial pool are used to establish the authentication between mobile sinks and stationary access nodes, which will enable these mobile sinks to access the sensor network for data gathering. Thus, an attacker would need to compromise at least a single polynomial from the mobile pool to gain access to the network for the sensor's data gathering. Polynomials from the static polynomial pool are used to ascertain the authentication and keys setup between the sensor nodes

Prior to deployment, each mobile sink randomly picks a subset of polynomials from the mobile polynomial pool. In our scheme, to improve the network resilience to mobile sink replication attack as compared to the single polynomial pool-based approach, we intend to minimize the probability of a mobile polynomial being compromised if Rc sensor nodes are captured.



. The three-tier security scheme in WSN with mobile sinks.

As an adversary can use the captured mobile polynomial to launch a mobile sink replication attack, we achieve this by having a small fraction of randomly selected sensor nodes carry a polynomial from the mobile polynomial pool. These preselected sensor nodes are called the stationary access nodes. They act as authentication access points for the network and trigger sensor nodes to transmit their aggregated data to the mobile sinks. A mobile sink sends data request messages to the sensor nodes via a stationary access node. The mobile sink's data request messages will initiate the stationary access node to trigger sensor nodes to transmit their aggregated data to the requested sink. Each stationary access node may share a mobile polynomial with a mobile sink. All sensor nodes, including the stationary access nodes, randomly select a subset of polynomials from the static polynomial pool. The advantage of using separate pools is that mobile sink authentication is independent of the key distribution scheme used to connect the sensor network. We divide our scheme into two stages: static and mobile polynomial pre-distribution and key discovery between a mobile sink and a sensor node.

#### IV. ORDER-PRESERVING SYMMETRIC ENCRYPTION

The order-preserving symmetric encryption scheme (OPE) is a deterministic encryption scheme where the numerical ordering of the plaintexts gets preserved by the encryption function. OPE was proposed in the database community by Agrawal to support e-client range queries on encrypted data. Then Boldyreva proposed the first formal cryptographic treatment of OPE which was an e-client block cipher-based scheme provably meeting their security definition. Boldyreva proved that the order preserving function  $g(x)$  for a given point  $x \in \mathbb{F}_1$ ;  $M_g$  has a NHG distribution over a random choice of  $g$ . However, this encryption scheme leaks approximate value of any plaintext and approximate distance between any two plaintexts, each to an accuracy of about square root of the domain size. In the revised order-preserving symmetric encryption proposed by Boldyreva et al:[13], a more e-client and stronger range

#### V. THREAT MODEL

At first, the storage node is considered as "honest-but curious" in our model. Specifically, storage node acts in an "honest" fashion and correctly follows the designated protocol specification. However, it is "curious" to infer and analyze data in its storage and query messages received during the protocol so as to learn additional information. In particular, the storage nodes analyze the sensing data from the lower-tier sensor nodes to gain the environment information or mine the query request from the users to know the query preference. Then, we extend the threat model into a more general one where the compromised storage node do not follow the designated protocol but return fake and/or incomplete data in response to top-k queries from the user. Based on the intensity of the attack,

**Plaintext attack model** - In this model, storage nodes intend to gain the plaintext of the sensing data and the query request.

**Background attack model** - Storage node is supposed to possess some background from the collected data and the query result, such as the statistical information, in addition to what can be accessed in plaintext attack model.

**Compromised attack model** - The compromised storage nodes are instructed to return fake and/or incomplete data in addition to what can be accessed in background attack

#### FRAMEWORK AND SECURITY REQUIREMENT FOR PRISEC TOP K

In this section, we define the framework of top-k query over-encrypted sensing data and establish various strict security requirements for such a three-tier wireless sensor network.

##### A. PriSec Top K Framework

**Initialization**- Besides collecting sensing data, the sensor node generates the secret key used in the scheme and calculates the scores for sensing data.

**Msg Enc**-Performed on the time slot data set, sensor node encrypts the sensing score and the time stamp by the symmetric key and then outsources them to storage node along with Message Authentication Codes.

**Intit Query**- Before sending the query request, the user encrypts the query request.

**Query**- When the storage node receives a query request, it performs the top-k query on the set of encrypted sensing scores and then returns the top-k encrypted sensing records as query result.

**Msg Dec**- The user generates the secret key to decrypt the sensing records and then achieves the top-k sensing scores.

##### B. Security Requirement for PriSec Top K

1) **Privacy**: We explore and establish a set of strict privacy requirements specifically for the PriSec Top K Framework. As for the data privacy on the time-slot data set, the sensor nodes should utilize a stateless encryption scheme to encrypt the data before outsourcing. In this scheme his encryption algorithms is processed on the single plaintexts, and prevent the storage nodes from prying into outsourced data. With respect to the query privacy, the users prefer to keep their query request from being exposed to the third party i.e., the time stamp and the k value. Traditional symmetric key cryptography should be resorted to guarantee the query efficiency of the scheme on the encrypted data set. The query results returned by the storage nodes contain much information such as the order-relation and the distance relation in top k query, therefore we consider the result privacy to prevent from the background attack. Within the top-k query, the order relation exposes the value order of the sensing scores in the top-k query result while the distance-relation exposes the value distance of the sensing scores in the top-k query result.

2) **Authenticity/Completeness**: Extending the attack model from the "honest-but-curious" to a more general one, the

storage nodes may be compromised in hostile environments and then instructed to return fake and/or incomplete data in response to the top-k query. Therefore, the user should verify the authenticity and completeness of the query results which forces the compromised storage nodes to return both authentic and complete top-k query results to avoid being caught. In particular, authenticity guarantees the sensing records indeed generated by the sensor node not by the compromised storage nodes, and completeness means that the query results indeed contains the top k sensing data among all the candidates.

**VI. THE ENHANCED THREE-TIER SECURITY SCHEME WITH SECURE PRI SEC TOP K**

As described in the previous section, the three-tier security scheme provides better network resilience against mobile sink replication attack compared to the single polynomial pool approach. This scheme delivers the same security performance as the single polynomial pool approach when the network is under a stationary access node replication attack. In both schemes, for any sensor node u that needs to authenticate and establish a pair-wise key with a stationary access node A, the two nodes must share at least a common polynomial in their polynomial rings. To perform a stationary access node replication attack on a network, the adversary needs to compromise at least a single polynomial from the static pool. This can be obtained easily by capturing arbitrary sensor nodes in the network. Then, the adversary can make use of this compromised polynomial by a replicated stationary access node to enable insecure access to the network. When successful access to the network has been obtained through the compromised static polynomial, the replicated stationary access node transmits recorded mobile sink data request messages. Next, the sensor nodes that have the compromised polynomial in their rings will insecurely authenticate and establish a pair-wise key with the replicated node and thus deliver their data to the replicated node.

In this section, we remedy the security performance of the proposed scheme in the case of a stationary access node replication attack. We use a one-way hash chain algorithm in conjunction with the polynomial pool scheme. In addition to the static polynomial, a pool of randomly generated passwords is used to enhance the authentication between sensor nodes and stationary access nodes.

In the enhanced security scheme, each sensor node, such as u, is preloaded with a subset of Ks polynomials randomly chosen from the static pool jSj. In addition to the Ks preloaded static polynomials, node u randomly picks a subset of Gs passwords from the password pool jWj. Subsequently, for each of the Gs password Pwi that has been randomly chosen by node u, its rth hash value, is loaded into node u. Each password is blinded with the use of a collision-resistant hash function such as MD5. Due to the collision-resistant property, it is computationally infeasible for an attacker to find a value Pwx, such that. For stationary access nodes, each node is preloaded with Ks - 1 static polynomials and Ga hash values for the randomly chosen passwords from the pool jWj. To establish

an authentication between a sensor node and a stationary access node in the enhanced scheme, the two must share a common static polynomial. Also, they need to discover at least a single access node  $H(H^{r-1}(Pw_i)) = H^r(Pw_i)$  verification for which both the sensor node and the stationary access node have the same password Pwi, randomly chosen from the pool jWj. In the access node verification, to verify the authenticity of a stationary access node, the sensor node performs a single hash operation on the hash value that is sent from the stationary access node.

**Privacy-preserving and Secure PriSec Top K**

Before giving our intact result, we start with a straightforward yet ideal scheme where the storage nodes follow the designate protocol to better illustrate the privacy problem in the three-tier wireless sensor network. a more general attack model that the compromised storage nodes do not follow the designate protocol but return the fake/incomplete query result in addition to learn additional information over the sensing data and query request.

**A. PriSec Top K I: Basic Scheme**

With respect to the plaintext attack model, PriSec Top K I enables the privacy of sensing data and query request by encrypting the data items with revised order-preserving and traditional symmetric encryption respectively.

1) *PriSec Top K I*: To provide a privacy guarantee against the attack on the plaintext of sensing data and query request as well as the efficiency requirement, we utilize revised order preserving encryption scheme and traditional symmetric encryption to encrypt the sensing data and query request. The order-preserving symmetric encryption scheme (OPSE) is based on the observation that any order-preserving function g from  $f1; : : : ; Mg$  to  $f1; : : : ; Ng$  can be uniquely represented by a combination of M out of N order items. The whole scheme to achieve top-k query over encrypted data in three-tier wireless sensor network is as follows.

Before deployed, each sensor node u shares with the user a unique secret key which we call the node’s individual key. Also the user and the sensor nodes have the same pseudo random function f (), corresponding seed and collision resistant hash function. The sensor initiates the scheme by generating random keys x, y from the pseudo-random function f (). After gaining the sensing data due at the time-slot t, the sensor node u calculates the score su for the sensing data. Let E be a semantically secure symmetric encryption algorithm. Each sensor node u encrypts the score su by order-preserving symmetric encryption as shown in Algorithm 1 to gain the cipher-text of the score OPSE (su) and then encrypt node ID using key fx (t), compute the keyed-hash for time stamp t. Sensor node u sends the encrypted data along with Message Authentication Code to the closest storage node S .

## 2. PriSec TopK II

In this section, we consider a more general attack model instead of the "honest but curious" one which the compromised storage node may return fake and/or incomplete data in addition to what can be accessed in the background attack model. Now we further propose PriSec TopK II with stronger the security requirement of the top-k query in three-tier wireless sensor networks.

*PriSec TopK II:* To guarantee the authenticity and completeness of the top-k query in three-tier sensor networks, we require the storage node to return some computing commitment information during the query process. Through hypothesis testing method combined with computing commitment, the compromised storage nodes are forced to return both authentic and complete top-k query results to the network owner. Also these methods increase the detection rate and reduce the additional communication cost while enabling verifiable top-k queries.

Initialization() and MsgEnc() in PriSec Topk II are similar to those in PriSecTopk I. Considering the Intit Query step, the user encrypts the query request  $Q_q = \langle q; t; k; m \rangle$  where  $m$  is the sample times used in query process, where  $s$  is the number of chosen sensor node in query process. With the secret key  $x$  and the one-way hash function  $h$ , the user encrypts the query request and then sends it to the storage node.

1. *Privacy:* Similar to the PriSec TopK I, the data privacy, query privacy and result privacy are all well protected by the PriSec TopK III.

2. *Efficiency:* Except for the communication and computation overheads consumed in the PriSec TopK I, the strategy for authenticity/completeness verification introduces additional communication and computing consumption. As for the additional communication consumption, the storage node is asked to return computing commitment information along with the query result. Also, during the sampling verification process the user and storage nodes should communicate to exchange the verification information. The computing overhead of the verified step based on hypothesis testing method is  $O(N)$ .

3. *Authenticity/Completeness:* Using the individual node key  $k_u$  as the hash key of the MACs sent with the sensing records, the compromised storage nodes cannot manipulate or append the sensing record because the node key of every sensor node is only known by the user and the sensor node respectively. Therefore, the authenticity of the query result the storage node sent back guarantees. Before proving the completeness of the scheme, we illustrate  $k$  satisfaction rate follows the normal distribution which is related with the computing commitment information, and prove the correctness of the hypothesis testing method to verify the completeness.

## VII. PERFORMANCE EVALUATION

In this section, we conduct a thorough experimental evaluation of the PriSec TopK scheme on a real-world dataset: Intel Berkeley Research Lab Data which includes a log of about 2.3 million readings. We randomly select different data records to build the dataset. And the record schema we used is as follows.

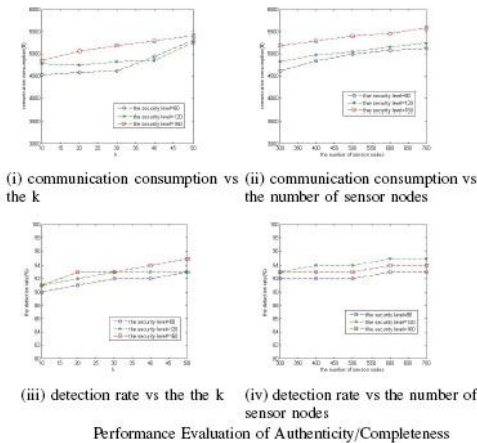
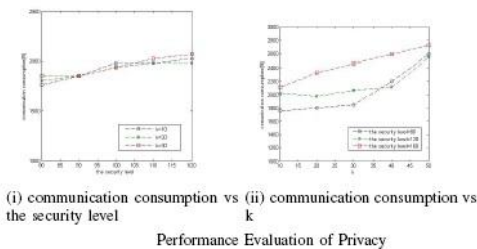
Date: 8B Time: 8B Mote Id: 4B Temperature: 4B

The whole experiment system is implemented by C language on a Linux machine with Intel CPU running at 3.0GHz. Algorithms use MATLAB libraries. The performance of the PriSec TopK scheme is evaluated regarding the trade off between privacy and efficiency as well as the authenticity/completeness.

Without loss of generality, suppose the size of MACs is 20B and the sampled times in the commitment information is 4B. In our experiment, we assume a cell with 400 sensor nodes and a storage node. The select times is 20,  $k = 30$  and the random function follows the normal distribution. The scheme we proposed does not introduce additional in network communication consumption between sensor nodes and storage node which is far more e-client and practical than the secure approaches. Therefore, we focus on the communication consumption between the storage node and the user as well as the computing consumption of the sensor nodes and the user.

### A. Privacy

As shown in Section V-B2, the communication overheads between the storage node and the user in PriSec TopK I is determined by both of the security level and the  $k$  value in the top-k query. The communication cost measurement to the fixed security level of our proposed. The result represents the mean of 50 trials. Note that even for a larger  $k$  value given a fix security level, the communication cost does not increase too much which is because the threshold of the random selection function adjusts adaptively. Specially, in the 80 bits security level, the communication cost between a storage node and the user of the proposed scheme is  $1:7 \sim 103B$  for top-10 query which is  $1:8 \sim 103B$  for top- 20 queries. We can find a step change of the communication cost when the  $k$  value is large. The reason is that the main effect factor of communication factor changes from the threshold to the  $k$  value. Even though the step change, the additional communication consumption decreases, which indicates that the scheme we proposed is more efficient for the top-k query whose  $k$  value is larger.



**B. Authenticity/Completeness**

To guarantee the authenticity and completeness of the query result the storage node returns, we introduce the verification method based on hypothesis testing method with computing consumption derived from the computing commitment information, the sampled sensor node IDs and their corresponding sensing records. Also indicates the efficiency of our scheme with different k value and the number of sensor nodes. Compared to the VFTop-k proposed in which brings large amount of communication consumption performed on time-slot dataset, namely, more than 104B in our experiment environment, the scheme we proposed only introduces only 1/3 communication consumption than that of the VFTop-k as well as protecting the privacy which is not guaranteed in the VFTop-k. The impact of k on the communication cost which increases smoothly when k goes up, and the higher security level asks for more communication consumption. The similar result can be found in Figure. 2ii but the increment is smoother with the larger number of sensor nodes, which is because that with the number of sensor nodes rises up the length of the computing commitment information does not increase that much. Considering the detection rate of the verification method we proposed, Figure. 2iii shows that the detection rate always larger than 90% in the experiment environment and the increment of the k value and the security level brings a slimly higher detection rate. we determine that the verification method adapts a large-scale sensor network with the reason that the larger number of sensor nodes increases the detection rate while it also bring larger communication consumption as shown above.

**VIII. CONCLUSION**

In this paper, we proposed a general three-tier security framework for authentication and pair-wise key establishment between mobile sinks and sensor nodes. The proposed scheme, based on the polynomial pool-based key pre-distribution scheme substantially improved network resilience to mobile sink replication attacks compared to the single polynomial pool-based key pre-distribution approach. Using two separate key pools and having few stationary access nodes carrying polynomials from the mobile pool in the network may hinder an attacker from gathering sensor data, by deploying a replicated mobile sink. We explore the problem of top-k query on time slot data set in three-tier wireless sensor network, and establish a set of privacy and correctness requirements for such a secure top-k scheme to become practical. We propose three PriSec TopK schemes meeting different privacy and correctness requirements in consideration of three levels of threat models. Thorough analysis investigating privacy, detection rate and efficiency guarantee of proposed scheme is given, and experiments on the real-world dataset further show the efficiency of proposed schemes. We have further improved the security performance of the proposed scheme against stationary access node replication attack by strengthening the authentication mechanism between stationary access nodes and sensor nodes.

**REFERENCES**

- [1] Amar Rasheed, Rabi N. Mahapatra “The Three-Tier Security Scheme in Wireless Sensor Networks with Mobile Sinks” IEEE transactions on parallel and distributed systems, vol. 23, no. 5, may 2012
- [2] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless Sensor Networks: A Survey. Computer Networks, vol.38, no.4, March 2002
- [3] Xiaojing Liao , Jianzhong Li, “Privacy-preserving and Secure Top-k Query in Two-tier Wireless Sensor Network
- [4] T. Gao, D. Greenspan, M. Welesh, R.R. Juang, and A. Alm, “Vital Signs Monitoring and Patient Tracking over a Wireless Network,” Proc. IEEE 27th Ann. Int’l Conf. Eng. Medicine and Biology Soc. (EMBS), Sept. 2005.
- [5] M. Shao, S. Zhu,W. Zhang, and G. Cao.pDCS: Security and privacy support for data-centric sensor networks. In INFOCOM07, pp. 1298C1306. IEEE, 2007
- [6] Sheng B, Li Q. Verifiable privacy-preserving range query in two tiered sensor networks. In INFOCOM’08, pp.46-50. IEEE 2008
- [7] Shi J, Zhang R, Zhang Y. Secure range queries in tiered sensor networks. In INFOCOM’09. pp.197-206. IEEE, 2009
- [8] Fei C, Alex L. SafeQ: Secure and E\_icient query processing in sensor networks. In NFOCOM’10. IEEE, 2010
- [9] Rui Z, Jing S, Yunzhong L, et al. Verifiable fine-grained top-k queries in tiered sensor networks. In INFOCOM’10. IEEE, 2010