Efficient and Secure Top-k Query Result Verification Process in Tiered Sensor Networks

Sushma V G
Dept of CSE
TJIT Bangalore

Mrs. Suma R
Asst. professor, Dept of CSE
TJIT, Bangalore

Abstract:- **Sensor nodes in wireless sensor network (WSN) collects data and forward towards the storage node. This traditional approach consumes more amount of energy. So to make energy efficiency more in WSN, storage nodes are expected to be placed as an intermediate (middle) tier of large scale sensor networks. These storage nodes can be used for caching the collected sensor readings. The storage nodes responds to user queries with benefits of power and storage, saving for ordinary sensors. An important issue is that the compromised storage node may not only cause the privacy problem, but also return fake/incorrect query processed results. So there is a requirement of strong verification scheme in WSN. We propose a simple and effective dummy/fake reading based anonymization process, under which the query result integrity can be guaranteed by the proposed verifiable top-*k* query VQ system. Compared with existing works, the VQ system have a fundamentally different design technique and achieve the lower communication complexity and saving bandwidth with improved throughput. For better result we introduced a buffered controller in the storage node. When a query request arrived at the storage node then it checks and if result is present then served from the storage node and validated based on the time to live. Several tests, experiment and prototype implementations are conducted to demonstrate the practicality of the proposed methods.**

Keywords— Query; Result Verification; Tiered Sensor Networks; VQ Systems

I. RELATED WORK

In paper [2], the author has discussed the work on The major problem of the wormhole attack. To over come this problem he introduced a general mechanism, called packet leashes. With a specific protocol, called TIK, that implements leashes. This mechanism takes place on host CPU rather than in the MAC layer and is ignored because of the authentication of each packet.

In paper [5], the author introduced a general method i.e. digital signature scheme in which the public key is fixed & secret signing key is updated. Forward security is used and the key is altered frequently but not efficient.

In paper [6], the author in order to achive differential privacy. Privacy-preserving data analysis approach differs in the statistics, databases, theory, and cryptography communities, in that a formal privacy is guarantee. The key privacy guarantee that has emerged is differential privacy. For the purpose the differential privacy was not achived upto the consideration because the attack on the differential privacy key was reconized.

In paper [9], the author discuses on the Fast Privacy-Preserving Top-*k* Queries using Secret Sharing. A set of parties hold private lists of key-value pairs. And uses the secure multiparty computation (MPC) techniques to solve this problem and design two MPC protocols, PPTK and PPTKS. IP addresses and port numbers were considered for the privacy purpose.

In paper [4], the author discusses on the Asymmetric Concealed Data Aggregation Techniques in Wireless Sensor Networks. Traditional security algorithms are infeasible in WSNs due to the limited computing, communication power, storage, band width and energy of sensor nodes. Concealed Data Aggregation (CDA) based on privacy homomorphism (PH) gives a critical solution for energy efficient secure data aggregation in WSNs.

In paper [10], the author NEMESYS. large-scale mobile botnets, smart-phones can also be used to launch attacks on mobile networks. NEMESYS will gather and analyze information about the nature of cyber-attacks targeting mobile users and the mobile network. The honeypots and a honeyclient.

In paper [11], the author Order-Preserving Encryption. OPeE was enhanced by other two methods ROPF and MOPE. ROPF(random order-preserving function) encryption leaks neither the precise value of any plaintext nor the precise distance between any two of them. Encryption MOPE with a random shift cipher. MOPE improves the security of OPE in a sense, as it does not leak any information about plaintext location**.**

In paper [3], the author discusses about the safeQ. This is of type two-tiered sensor networks, where storage nodes serve as an intermediate tier between sensors and a sink. SafeQ, a protocol that prevents attackers . SafeQ also allows a sink to detect compromised storage nodes when they misbehave.

In paper [7], the author reaches about the Secure Range Queries in Tiered Sensor Networks. 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. A compromised master node may leak hosted sensitive data to the adversary. Efficient range-query processing is the technique, prevents compromised master nodes from reading hosted data and also ensures high query efficiency.

In paper [8], the author brings about the Order Preserving Encryption for Numeric Data to the proposed system. Once encrypted, data can no longer be easily queried aside from exact matches. Presents an order-preserving encryption

scheme. Any data is being compromised then further cannot encrypt or decrypt the DATA

II. INTRODUCTION

A. Tiered Sensor Networks

In tiered sensor networks for data collection, there could be unstable connection between the authority (or network owner) and network, a intermediate(middle) tier with the purpose of caching the sensed data for data archival and query response becomes necessary. The network model is illustrated in Fig. 1, where the authority can forward queries to retrieve the sensor readings. The intermediate(middle) tier is composed of a small number of storage-abundant nodes [24], called *storage nodes*. The bottom tier consists of a large number of resource-constrained ordinary sensors that sense the environment.

In the above tiered architecture, sensor nodes are usually partitioned into disjoint groups, each of which is associated to a particular storage node. Each group of sensor nodes is called a *cell*. The sensor nodes in a cell form a multi-hop network and always forward the sensor readings to the associated storage node. The storage node saves a copy of received query results from the sensor readings and is responsible for answering the queries from the authority. An example of the tiered architecture can be found in Fig. 1.

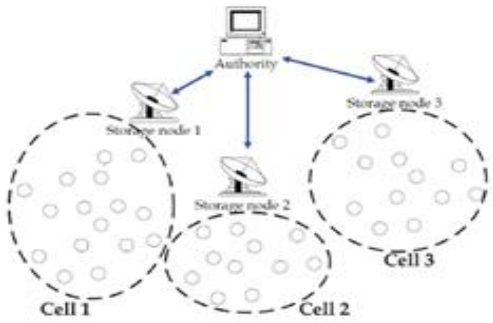

Fig 1: Tired sensor network architecture.

B. Security Issues in Tiered Sensor Networks

In tiered sensor networks, the authority issues proper queries to retrieve the desired portion of sensed data. Top-k query [29] can be used to extract the extreme results from the sensor nodes readings. So easily, the storage nodes also easily become the targets to be compromised because of their significant role in responding to queries.

By compromising storage nodes, the adversary can also return the falsely extracted readings to the authority. The most challenging is that the compromised storage nodes can violate query result completeness, creating an incorrect query result for the authority by replacing some portions of the query result with the other genuine readings. For example, once the storage node 1 is compromised by the adversary, the storage node 1 can be configured by the adversary to always return unqualified sensor readings to the authority. Should identify that data integrity usually refers to both data authenticity and completeness.

C. Existing Works on Verifiable Queries

Two methods, additional evidence and crosscheck verification, were proposed [36] in as solutions for securing top-*k* query in tiered sensor networks. Where the former generates hashes for each and every consecutive pair of sensed data by sensor nodes for verification purpose, the later performs on the network-wide broadcast such that the information about the readings is distributed to all over the network and therefore the query result cannot be manipulated.

The concept behind additional evidence is that if each consecutive pair of sensed data is associated with a hash, once an unqualified sensor reading is used to replace the genuine query result, the authority may know because it can find that there are some missing sensor readings for hash verification.

The goal behind crosscheck verification is that the genuine top-k results are distributed over all several sensor nodes. With certain consideration, the authority will find a query result incompleteness by checking the other sensor nodes' sensor readings.

Hybrid method [36] is a combination of additional evidence and crosscheck verification, attempting to balance the communication cost and the query result incompleteness detection capability.

D. Efficiency and Security Gap

Despite the further works on verifiable query results, here still have the following concerns:

• In a network of *n* sensor nodes, Hybrid method [36] incurs tremendous *O (n2)* communications.

• Although SMQ [3] can be adapted to verify the top-*k* query result, an aggregation process tree not only needs to be constructed but also needs to remain intact and unchanged for further. The exact information about the tree topology is also required by the authority. In real world deployment, of these requirements are difficult to meet.

Because these methods [36] do not handle the data privacy issue. On the other hand, the bucket index used in SMQ [35] leaks the possible value range for each sensor reading, which could be valuable data information, to the adversary.

E. Naïve Approaches

Although the method [36] can be extended in some straightforward way to the methods with data confidentiality and integrity guarantee, such extension actually implies some of the other severe back holds, which are unacceptable in the design of a verifiable query system. Consider the case where the sensor readings are encrypted by popular encryption functions, like DES algorithm and AES algorithm. In this case, the storage node were unable to answer for the top-*k* query issued by the authority due to for want of the numeric order of sensor readings. On the other hand, consider the case that order-preserving encryption algorithm (OPE) [3] is used to encrypt sensor readings. In this case, the numeric order of sensor readings is preserved. Nevertheless, this is method achieves by all of the sensors sharing a common OPE privacy key. The reason for doing so is that once a sensor is compromised, the OPE key is exposed to the adversary and the data confidentiality and integrity is completely breached.

Verifiable query processing is also considered in the context of range query. In [24], the query result completeness is achieved by requiring sensors to send cryptographic one-

way hashes to the storage node even when they do not have satisfying readings. In [27], [37], crosscheck was also utilized to secure range query, as in [36]. By converting the verification of whether a number is in a range to several verification of whether two numbers are equal, SafeQ [9] offered an alternative for data retrieval in encryption domain. In SMQ [35], each sensor applies hash operation to the received data and its own data, generating a verifiable object of the sensor readings of the entire network. The basic idea behind SMQ is to construct an aggregation tree over the sensor nodes. Afterward, each sensor node simply aggregates and forwards the sensor readings of all its descendant nodes to its parent node. The notion of stream cipher is used in [28] to have a design of more efficient encrypted data retrieval. The database community also conducted research on the completeness verification. Nevertheless, similar to [1], all the data to be queried are generated by the single entity. In addition, the prior works on top-*k* query in [6], [29] focus on the privacy issue, rather than integrity issue.

F. Contributions

The Verifiable top-*k* Query (VQ) system based on the novel dummy reading-based anonymization framework are proposed for privacy preserving top-*k* query result integrity verification in tiered sensor networks. In particular, its being considered with the following contributions:

• A randomized and distributive version of Order Preserving Encryption, rdOPE, is proposed to be the privacy innovation of this methods.

• AD-VQ-static aimed with the lower communication complexity at the cost of slight detection capability degradation, which could be of both theoretical and practical interest works.

• Analytical study, numerical simulation, and prototype implementation are implemented to demonstrate the practicality condition of the proposed methods.

IV. SYSTEM MODEL

A. Network Model

As shown in Fig. 1, the sensor network considered in this paper is composed of a large number of resource-constrained sensors nodes and a few storage abundant storage nodes. A cell (the dashed circle in Fig. 1) is a connected multihop network composed of a storage node and a number of ordinary sensor nodes. Storage nodes can communicate with the authority A via direct or multi-hop communications, and are assumed to know their particulars multihop cells. Time on the nodes has been synchronized and is divided into equal time i.e. epochs [17], [26].

B. Security Model

After node are being compromised, all the data/information stored in the compromised nodes will be fused to the adversary. The adversary takes full control of the storage node and now is in the ability to manipulate its computation result and communication process list. The goal of the adversary is to breach atleast one of the data privacy, authenticity, completeness, and integrity of the information. Since this approach focuses on the design for securing top-*k* query, we assume that the other security issues such as broadcast authentication process, key establishment [31] method, and anomaly detection [2], [12], [13] concept are applicable.

- *Authority*

The Authority receives query as shown in fig 2 from the user, query is sent to the storage nodes to fetch the query result.

- *Buffer memory*

The query is sent to the buffer memory. The buffer memory searches for the result, where the query result is being saved according no of search of the particular query. If the query result fetched then within no time the result will be fetched and given to the authority. if the result not being fetched then it sends to the storage nodes.

- *Storage nodes*

Storage nodes receives query as shown in Fig 2 from the authority and forwards it to the sensor nodes further.

- *Sensor nodes*

As shown in the fig 2 the authority as and when receives the query, by the further process as explained it forwards to the particular sensor node.

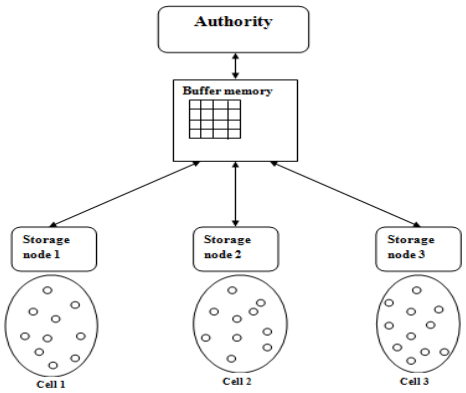

Fig 2: System architecture.

Sensor nodes individually will search for the result, when it gets the file, the file will be encrypted and sent to the storage node. Storage node will decrypt the encrypted file and send to the authority. The storage node will also command the sensor node to search for the file and to upload the file. When the particular query is surfed many times it will be added to the buffer memory by using the concept of the **rank function** method.

And further if sensor nodes are compromised then storage detects that compromised sensor nodes and removes tat sensor node then replaces tat with the related sensor node.

C. Query Model

For the top-*k* query, although its sensible to consider a ranking function method [29], which is used to take the outcome the ranking scores of data items searched, to ease the presentation, we assume that *A* instead asks *sM* to return the data readings with the first k^{th} highest values of the corresponding cell.

D. Problem Statement

Suppose *A* issues a top-*k* query to *sM*. Let $B = \{di, j \mid 1 \le i \le n, 1 \le j \le \mu i\}$ be the set of sensor readings of entire network.

The objective is to obtain the top-*k* result _*k* that fulfills the following requirements:

• **privacy**: β cannot be known by *sM*, and moreover, {*di, j* $|1 \leq j \leq \mu i$ } can only be known by *si* .

•**authenticity**: $\Omega k \subseteq \beta$. The readings in _*k* are from sensors {*si* }*ni* =1.

•**completeness**: min $\Omega k \geq \max(\beta \setminus \Omega k)$. No readings smaller than the minimum element in _*k* will be accepted by *A*.

E. Performance Metrics

The following performance metrics used here evaluates the integrity verification methods:

• **detection probability**, *P*X det: the probability that an inauthentic or incomplete query result is detected by *A* in the X scheme.

•**communication cost**, *C*X: the communication cost *C*X of the X scheme is defined as:

$CX = CX\ T + \beta CX\ V$,

where CX_T , $CX_{,V}$, and β denote the number of bits transmitted between sensors and *sM* in data submission phase (in cell communication cost), the number of bits transmitted between *sM* and *A* in query response phase (query communication cost), and the query frequency of *A* issuing the queries to retrieve data, respectively. For example, β = 0.01 means that on average the issues a query for every 100 epochs.

V. PROPOSED METHODS

In proposed methods, a novel use of randomized and distributed OPE (rdOPE) algorithm, is first developed to accomplish the privacy protection guarantee in the proposed Verifiable top-*k* Query VQ system. Study evolves in a number of successive forward steps; by presenting Global Dummy reading-based VQ (GD-VQ) and Local Dummy reading-based VQ (LD-VQ), which constitute the foundation of proposed dummy reading-based anonymization process. Afterward, they are enhanced to be Advanced Dummy reading-based VQ (AD-VQ) system, which reduces the communication overhead significantly.

A. The rdOPE Scheme

1) Motivation: OPE has been applied widely to encrypted database retrieval purpose. Unfortunately, in the literature, the data are all assumed to be generated and encrypted by a particular single authority, which is not the case in the consideration. In addition, because the number of possible sensor readings could be non unlimited and known from hardware specifications, the relation between plaintexts and ciphertexts could be revealed up. For example, if the sensors can only generate 20 kinds of possible outputs, then practically the adversary can derive the OPE key by investigating the numerical order of the eavesdropped ciphertexts despite the theoretical security threat guarantee.

2) Algorithmic Description of rdOPE: Solution is a novel use of OPE algorithm, called rdOPE, which provides the randomness in the encryption output's and is suitable for the case of distributed data generation with limited input value range. The technical challenge of rdOPE design is to maintain the numerical orders of encryptions from different sensors that use different OPEs. With the observation that the possible mapping between plaintexts and ciphertexts are fixed in advance, the ciphertexts can be determined prior to sensor deployment such that the numerical orders of ciphertexts in different sensors can be preserved.

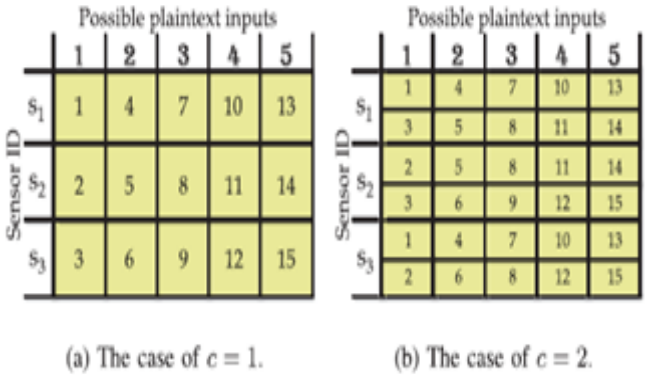

Fig 3: Example of rdOPE

B. The GD-VQ system

Basic Idea of GD-VQ The basic idea of GD-VQ is that the privacy, authenticity, and completeness are guaranteed by rdOPE, cryptographic hash, and the insertion of dummy readings, respectively. In particular, once the adversary cannot distinguish between genuine and dummy readings, the malicious removal of query results may cause the lose of dummy readings that are supposed to be included in the query result.

Note that the "dummy readings" of the sensor *si* are defined as those readings sent from *si* to the storage node, generated by the program of *si* itself, but not collected from the sensor hardware to reflect the environment condition. This enables *A* to detect the query result incompleteness.

1) Algorithmic Description of GD-VQ: The μi sensor readings are encrypted by *si* with rdOPE key *k(i)* to form $ei,1 < \cdots < ei,\mu i$. Let $\alpha gdvq$ be a security parameter of GD-VQ. Each sensor additionally generates $\alpha gdvq$ distinct random dummy readings from [1, *b*], resulting in $\hat{}ei,1 < \cdots < \hat{}ei,\mu i+\alpha gdvq$, where μi of them are $ei,1, \ldots, ei,\mu i$, and $\alpha gdvq$ of them are the dummy readings. This can be implemented by calculating $hGDVQ(\tilde{}\ ki\ ||1), \ldots, hGDVQ(\tilde{}\ ki\ ||\alpha gdvq)$, where the output range of $hGDVQ(\cdot)$ is [1, *b*]. To ease the analysis, the dummy readings are assumed to not collide with $\{ei,1, \ldots, ei,\mu i\}$. An illustrative example is shown in Fig. 4a, where the rdOPE ciphertexts are generated from rdOPE key in Fig. 3b.

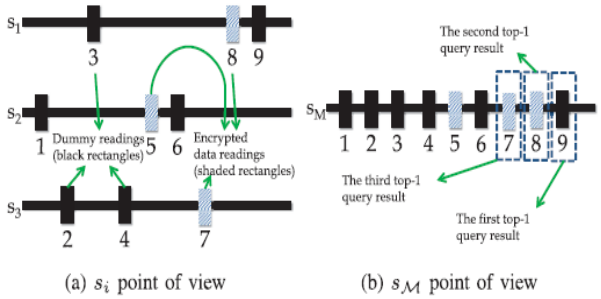

Fig 4: Example of GD_VQ

Since the dummy readings are generated randomly from [1, *b*], they could collide with the legitimate ciphertext that *si* does not sense the corresponding reading. Without particular treatments, this kind of collision makes *A* accept false readings. For example, as shown in Fig. 4a, the dummy reading 9 generated by *s*1 can be pruned easily by *A* because no entry 9 is in *k(1)* of Fig. 3b. Nonetheless, *s*3 generates dummy readings 2 and 4 but actually does not have the corresponding sensor readings 1 and 2. This results in a circumstance where the dummy readings collide with encrypted sensor readings.

Under this circumstance, *A* may falsely accept 1 and 3 as *s*1's readings. The remedy is to generate different hashes depending on whether the reading is dummy.

2) Top-k Query Processing of GD-VQ: In GD-VQ, to retrieve $\_k$, *A* instead needs to issue top-1 queries repeatedly until *A* obtains $\_k$. In other words, from *sM* point of view, for each received top-1 query, the current top-1 query is applied to the sensor readings excluding the previously returned top-1 results. An example of the query response phase of GD-VQ is shown in Fig. 3a, where *A* obtains two genuine readings, 7 and 8, by repeatedly issuing three top-1 queries.

On the other hand, from *A* point of view, for each received top-1 result, it follows the algorithm in Fig. 4 to verify the query result integrity and determine whether further top-1 query is needed.

This can be accomplished because with the knowledge of $k_{(i)}$, *A* can generate all of the dummy readings by also calculating $h_{GDVQ}(\tilde{}\ k_i\ ||1), \ldots, h_{GDVQ}(\tilde{}\ k_i\ ||\alpha_{gdvq})$, $1 \le i \le n$. For example, when receiving the second top-1 result, 8, *A* checks whether it had received 9 with the knowledge of 1, 2, 3, 4, 6, and 9 being dummy. Subsequently, *A* checks whether $e_{\pi,j}$ is dummy by calculating $h^{\tilde{}}{}_{k\pi}(e_{\pi,j}\ ||\ \tilde{}\ k\pi)$ and $h^{\tilde{}}{}_{k\pi}(e_{\pi,j})$ (the second **if** statement in Fig. 4). $e_{\pi,j}$ is genuine sensor reading if $h_{\pi,j} = h^{\tilde{}}{}_{k\pi}(e_{\pi,j}\ ||\ \tilde{}\ k\pi)$, is dummy if $\pi, j = h^{\tilde{}}{}_{k\pi}(e_{\pi,j})$, and is inauthentic otherwise. Finally, depending on whether *A* has collected enough number of genuine sensor readings, *A* issues Fig. 5.

Parameter: $\Omega k = \emptyset$ is set for the first execution

```
1   if all dummy readings ≥e(Π,  j) have been received
2     if e(Π,  j) is a genuine sensor reading
3       Ωk = Ωk {D(k(Π))(e(Π,  j))}
4         if |Ωk| <  k
5           issue one more top-1 query
6         else stop issuing top-1 query
7       elseif e(Π,  j) is dummy reading
8         issue one more top-1 query
9       else alarm of inauthentic query result
10    else alarm of incomplete query result
```

$$= 1 - \frac{k}{k'}\frac{k-1}{k'-1}\cdots\frac{k-(x-1)}{k'-(x-1)} = 1 - \prod_{\ell=0}^{x-1}\frac{k-\ell}{k'-\ell}.$$

Fig 5: Algorithm of GD-VQ

2) Detection Probability of GD-VQ: Assume that $k\_ \ge k$ top-1 queries are issued by *A* to retrieve $\_k$. From the above description, one can know that among these $k\_$ query results, *k* genuine and dummy readings, the only option for the adversary is to randomly choose and replace *x* of $k\_$ query result by the other smaller readings. The detection probability *PGDVQ* det of GD-VQ can be formulated as:

PGDVQ det = *Pr*[at least one of *x* choices are dummy]

3) Weakness of GD-VQ: Though the dummy reading insertion enables A to verify the result completeness, as the dummy readings are distributed over [1, b], GD-VQ is in fact very communication inefficient, because 1) A is required to issue an uncertain number of top-1 queries to obtain the genuine top-k result, and 2) all of the dummy readings need to be returned in the worst case, leading to the overwhelming communication burden. Subsequently, a local dummy reading based scheme is proposed to conquer these two performance problems.

C. The LD-VQ Scheme

1) Basic Idea of LD-VQ: The LD-VQ design is the same as the GD-VQ design except that the dummy reading generation is dependent on the sensor readings and distributed over a limited range. By further taking advantage of the observation that the compromised storage node in most cases is unable to eavesdrop on sensor communications, such design has two benefits:

1) *A* can issue a single query to retrieve the genuine top-*k* result, reducing the need of two way communication between *sM* and *A*.

2) Even in the worst case, *CV* can still be limited. Though the level of anonymization of LD-VQ is weaker than GD-VQ, *A* is provided with an efficient way for the retrieval of $\_k$.

2) Algorithmic Description of LD-VQ: The μi sensor readings of *si* are encrypted by *si* with rdOPE key *k(i)* to form $ei,1 < \cdots < ei,\mu i$. Let $\alpha ldvq$ be a security parameter of LD-VQ. In LD-VQ, each sensor additionally generates $\alpha ldvq - 1$ distinct local dummy readings for each encryption, resulting in $\hat{}ei,1 < \cdots < \hat{}ei,\alpha ldvq\mu i$, where μi of them are $ei,1, \ldots, ei,\mu i$ while $(\alpha ldvq - 1)\mu i$ of them are dummy. The dummy reading generation on each sensor *si* in LD-VQ is that, for each reading ei, j , $\alpha ldvq$ distinct dummy readings are selected randomly from $[ei, j - \delta 2\ ,\ ei, j + \delta 2\ ]$. This can be implemented by calculating $hLDVQ(\tilde{}\ ki\ ||1), \ldots, hLDVQ(\tilde{}\ ki\ ||\alpha ldvq)$, where the output range of $hLDVQ(\cdot)$ is $[1, \delta]$ with δ being a system parameter affecting security and communication cost. The dummy readings are local in the sense that they are distributed over a restricted range. An illustrative example of LD-VQ is shown in Figs. 4a and 4b. The local dummy readings might also lead to the collision mentioned in Sec. IV-B. The technique can be utilized here to resolve the collision problem. We omit the repeated description for saving space.

The rationale behind the ID information removal is that once the adversary can identify the sources of readings, it can remove all of the readings from the sensors generating the top-*k* result without being detected. For example, in Fig. 4a, if the adversary knows 7, 8, and 9 are from *s*1, then it can return the incomplete result 1, 2, 3, 4, 5, and 6 that will succeed in the integrity verification below.

3) Top-k Query Processing of LD-VQ: In LD-VQ, to retrieve _*k*, *A* instead needs to issue a top-*kαldvq* query to *sM* because, in the worst case, the dummy readings induced by *ei, j* are all larger than *ei, j* . Let R = {*(ei , hi)*|1 ≤ *i* ≤ *kαldvq* }, *e*1 ≤ · · · ≤ *ekαldvq* , be the received top-*kαldvq* result. *A* performs the following procedures to verify its integrity.

LD-VQ can also fulfill the privacy, authenticity, and completeness requirements defined in Sec. I due to the similarity between LD-VQ and GD-VQ.

D. The AD-VQ Scheme

While GD-VQ incurs overwhelming communication burden, the security of LD-VQ completely relies on the assumption of local adversary. Moreover, the above two proposals share a common weakness that all of the readings, including genuine and dummy, need to be sent explicitly. In AD-VQ, we offer an alternative that can conquer the above problems simultaneously.

1) Basic Idea of AD-VQ: Its observed that a property of top-*k* result that the readings of neighboring sensors of the sensors generating _*k* are either smaller than min_*k* or are included in _*k*, as shown in Fig. 5a where 5 and 6, 4 and 5, and 7 and 8 in *s*6, *s*7, and *s*9, respectively, are smaller than the top-1 result, 9, in *s*8. A straightforward method for the query result completeness verification is to enable *A* to also have the readings of the neighboring sensors of the sensors claiming to generate _*k* . Nevertheless, this method is flawed in that the *global adversary* that monitors every single communication of the entire network may exhaustively search for the "*hill sensor*", which is defined as the sensor whose maximum reading is larger than all of the readings of its neighboring sensors, but is not in _*k*. For example, *s*3 in Fig. 5a is the hill
sensor for top-1 query because its reading 5 is the maximum of the readings in the proximity.

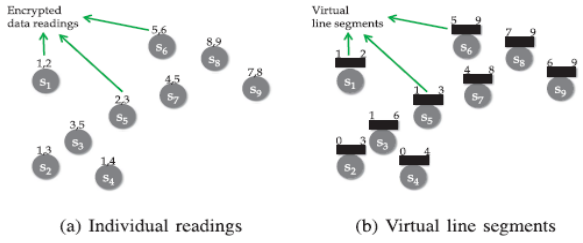

Fig 5: The conceptual illustration of AD-VQ

2) Algorithmic Description of AD-VQ: Each sensor *si* has the sensed data *di*,1 < · · · < *di,μi* and their encryptions *ei*,1 < · · · < *ei,μi*. Let *η* be a system parameter denoting the difference between the maximum and minimum encrypted readings within an epoch.2 Then, *si* constructs a virtual line segment *Li* = _*Li,L* ,*Li,U* _ with *Li,L* = *ei,μi* − *η* and *Li,U* = *ei,μi*, where *Li,L* and *Li,U* are used to represent the starting and ending points of *L*, respectively.

VI. CONCLUSIONS

A novel dummy reading-based anonymization process is proposed to design Verifiable top-*k* Query VQ system. In particular, AD-VQ-static system achieves the lower communication complexity with only minor detection capability penalty, which could be of both theoretical and practical intension. With only symmetric cryptography involved and their low implementation difficulty, the VQ systems are suitable and practical for current sensor networks. Malicious sensor nodes can be fetched by creating a key(k) in the storage nodes while forwarding the query, the sensor nodes will first send the key, the key generated by sensor node will be checked with the key of the storage node, if the keys get matched the sensor node is the verified one or else it is a malicious sensor node.

REFERENCES

[1] R. Zhang, Y. Zhang, and C. Zhang, "Secure top-k query processing via untrusted location-based service providers," in *Proc. 24th IEEE Conf. Comput. Commun.*, Mar. 2012, pp. 1170–1178.
[2] O. H. Abdelrahman, E. Gelenbe, G. Görbil, and B. Oklander, "Mobile network anomaly detection and mitigation: The NEMESYS approach," in *Proc. 28th ISCIS*, Oct. 2013, pp. 429–438.
[3] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in *Proc. ACM SIGMOD*, 2004, pp. 63–574.
[4] A. Boldyreva, N. Chenette, Y. Lee, and A. O'neill, "Order-preserving symmetric encryption," in *Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Tech.*, 2009, pp. 224–241.
[5] A. Boldyreva, N. Chenette, and A. O'neill, "Order-preserving encryption revisited: Improved security analysis and solutions," in *Proc. Int. Cryptol. Conf. CRYPTO*, 2011, pp. 1–18.
[6] M. Burkhart and X. Dimitropoulos, "Fast privacy preserving top-k queries using secret sharing," in *Proc. 19th ICCCN*, 2010, pp. 1 7.
[7] D. Boneh, C. Gentry, B. Lynn, and H. Shacham, "Aggregate and veriably encrypted signatures from bilinear maps," in *Proc. Int. Conf. Theory Appl. Cryptograph. Tech. Adv. Cryptol.*, 2003, pp. 416–432.
[8] M. Bellare and S. K. Miner, "A forward-secure digital signature scheme," in *Proc. 19th Annu. Int. Cryptol. Conf. Adv. Cryptol.*, 1999, 431–438.
[9] F. Chen and A. X. Liu, "SafeQ: Secure and efficient query processing in sensor networks," in *Proc. 24th IEEE Conf. Comput. Commun.*, Mar. 2010, pp. 1–9.
[10] C. Dwork, "Differential privacy," in *Proc. ICALP*, 2006, pp. 1–12.
[11] P. Desnoyers, D. Ganesan, and P. Shenoy, "TSAR: A two tier sensor storage architecture using interval skip graphs," in *Proc. ACM 3rd Int. Conf. Embedded Netw. Sensor Syst.*, 2005, pp. 39–50.
[12] E. Gelenbe, G. Görbil, D. Tzovaras, S. Liebergeld, D. Garcia, M. Baltatu, *et al.*, "NEMESYS: Enhanced network security for seamless service provisioning in the smart mobile ecosystem," in *Proc. ISCIS*, Oct. 2013, pp. 369–378.
[13] E. Gelenbe and G. Loukas, "A self aware approach to denial of service defence," *Comput. Netw.*, vol. 51, no. 5, pp. 1299–1314, 2007.
[14] [Onllie].Available: http://www.hpl.hp.com/news/2009/octdec/ cense.html
[15] Y.-C. Hu, A. Perrig, and D. Johnson, "Packet leashes: A defense against wormhole attacks in wireless networks," in *Proc. 22nd Annu. Joint Conf. IEEE Comput. Commun. INFOCOM*, Apr. 2003, pp. 1976–1986.
[16] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. ICPS*, Jul. 2005, pp. 88–97.
[17] Q. Li and D. Rus, "Global clock synchronization in sensor networks," in *Proc. IEEE Conf. Comput. Commun. INFOCOM*, Jan. 2004, pp. 1–11.
[18] D. Ma and G. Tsudik, "A new approach to secure logging," *ACM Trans. Storage*, vol. 5, no. 1, pp. 1—3, 2009.
[19] J. Nesome, E. Shi, D. Song, and A. Perrig, "The sybil attack in sensor network: Analysis & defense," in *Proc. 3rd Int. Symp. ISPN*, Apr. 2004, pp. 259–268.
[20] B. Parno, A. Perrig, and D. Johnson, "Distributed detection of node replication attacks in sensor networks," in *Proc. IEEE Symp. Security Privacy*, May 2005, pp. 49 63.
[21] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and D. Tygar, "SPINS: Security protocols for sensor networks," in *Proc. ACM Conf. Mobile Comput. Netw.*, 2001, pp. 521–534.
[22] L. Sweeney, "*k*-Anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl. Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.

[23] B. Schneier and J. Kelsey, "Cryptographic support for secure logs on untrusted machines," in *Proc. 7th USENIX Security Symp.*, 1998, pp. 53–62.
[24] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proc. 24th IEEE 27th Conf. Comput. Commun.*, Apr. 2008, pp. 743–766.
[25] B. Sheng, Q. Li, and W. Mao, "Data storage placement in sensor networks," in *Proc. 7th ACM Int. Symp. Mobile Ad Hoc nNetw. Comput.*, 2006, pp. 344–355.
[26] K. Sun, P. Ning, C. Wang, A. Liu, and Y. Zhou, "TinySeRSync: Secure and resilient time synchronization in wireless sensor networks," in *Proc. 13th ACM Conf. CCS*, Feb. 2006, pp. 264–277.
[27] J. Shi, R. Zhang, and Y. Zhang, "Secure range queries in tiered sensor networks," in *Proc. 24th IEEE Conf. Comput. Commun.*, Jan. 2009, pp. 1–9.
[28] Y.-T. Tsou, C.-S. Lu, and S.-Y. Kuo, "Privacy- and integrity-preserving range query in wireless sensor networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2012, pp. 328–334.
[29] J. Vaidya and C. Clifton, "Privacy-preserving top-k queries," in *Proc. IEEE 21st ICDE*, Apr. 2005, pp. 545–546.
[30] M. Wu, J. Xu, X. Tang, and W.-C. Lee, "Top-k monitoring in wireless sensor networks," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 7, pp. 962–976, Jul. 2007.
[31] C.-M. Yu, C.-S. Lu, and S.-Y. Kuo, "Noninteractive pairwise key establishment for sensor networks," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 556–569, Sep. 2010.
[32] F. Ye, H. Luo, S. Lu, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," in *Proc. IEEE 23rd Annu. Joint Conf. Comput. Commun. INFOCOM*, Mar. 2004, pp. 2446 2457.
[33] C.-M. Yu, G.-K. Ni, I.-Y. Chen, E. Gelenbe, and S.-Y. Kuo, "Top-k query result completeness verification in sensor networks," in *Proc. IEEE Int. ICC Workshops*, Jun. 2013, pp. 1026–1030.
[34] T.-H. You, W.-C. Peng, and W.-C. Lee, "Protecting moving trajectories with dummies," in *Proc. Int. Conf. Mobile Data Manag.*, May 2007, pp. 278–282.
[35] C.-M. Yu, Y.-T. Tsou, C.-S. Lu, and S.-Y. Kuo, "Practical and secure multidimensional query framework in tiered sensor networks," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 2, pp. 241–255, Jun. 2011.
[36] R. Zhang, J. Shi, Y. Liu, and Y. Zhang, "Verifiable fine-grained topk queries in tiered sensor networks," in *Proc. 24th IEEE INFOCOM*, Mar. 2010, pp. 1–9.
[37] R. Zhang, J. Shi, and Y. Zhang, "Secure multidimensional range queries in sensor networks," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2009, pp. 197–206.