Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 538482, 8 pages
http://dx.doi.org/10.1155/2015/538482

Research Article

An Efficient Top-k Query Processing with Result Integrity Verification in Two-Tiered Wireless Sensor Networks

Ruiliang He,[1] Hua Dai,[1] Geng Yang,[1] Taochun Wang,[2] and Jingjing Bao[3]

[1]*College of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2]*College of Mathematics and Computer Science, Anhui Normal University, Wuhu 241003, China*
[3]*Tongda College of Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

Correspondence should be addressed to Hua Dai; daihua@njupt.edu.cn

Received 9 April 2015; Revised 13 July 2015; Accepted 27 July 2015

Academic Editor: Mark Leeson

In two-tiered wireless sensor networks, storage nodes take charge of both storing the sensing data items and processing the query request issued by the base station. Due to their important role, storage nodes are more attractive to adversaries in a hostile environment. Once a storage node is compromised, attackers may falsify or abandon the data when answering the query issued by the base station, which will make the base station get incorrect or incomplete result. This paper proposes an efficient top-k query processing scheme with result integrity verification named as ETQ-RIV in two-tiered sensor networks. According to the basic idea that sensor nodes submit some encoded message containing the sequence relationship as proof information for verification along with their collected sensing data items, a data binding and collecting protocol and a verifiable query response protocol are proposed and described in detail. Detailed quantitative analysis and evaluation experiments show that ETQ-RIV performs better than the existing work in both communication cost and query result redundancy rate.

1. Introduction

Since the traditional multihop architecture is not suitable for large-scale wireless sensor networks (WSNs), a novel two-tiered architecture has been proposed. The two-tiered wireless sensor networks (TWSNs) introduce storage nodes that are abundant in energy, memory, and computing power to traditional multihop WSNs. In such a two-tiered architecture, the storage node serves as an intermediate tier between the base station and the sensor nodes, which is shown in Figure 1. The whole network is partitioned into several cells, each of which consists of a storage node and some sensor nodes nearby. The sensing data are sent to and stored in the storage node in the same cell after being collected by the sensor nodes. After receiving the queries issued by the base station, the storage node processes the query over the data items that are received from the sensor nodes and then returns the query result to the base station. The two-tiered architecture is also known to be indispensable for increasing network capacity and scalability, reducing system complexity, and prolonging network lifetime [1–3].

In TWSNs, storage nodes not only store all the sensing data of all the sensor nodes in the same cell, but also respond to the query requests issued by the base station, which makes the storage nodes more attractive to adversaries. Once a storage node is compromised, attackers may falsify or abandon the data when the storage node is answering the query issued by the base station, which will make the base station get incorrect or incomplete result. In an application that depends on the query result, an incomplete result may lead to wrong decisions. Therefore, it is of great significance to construct a verifiable query processing scheme, by which authenticity and integrity of query results can be verified by the base station.

Top-k query is frequently used in WSNs aiming at getting the k highest or lowest data in a specified region during a specified time epoch. For instance, "to get the 5 highest temperature data of the second warehouse from 12:00 to

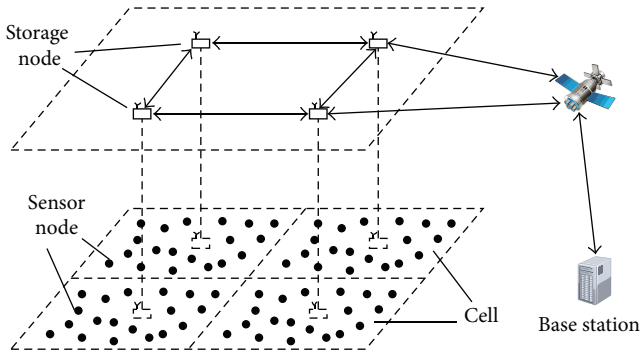


Figure 1: Network model of TWSN.

13:00" is a typical top-k query that could be used as fire detection. This paper proposes an efficient top-k query processing scheme with result integrity verification in TWSNs, named as *ETQ-RIV*, the basic idea of which is as follows. The sensor nodes submit some encoded message containing the sequence relationship along with their collected sensing data items to the storage nodes. When the base station issues queries, storage nodes send all the satisfied data items along with the encoded message as proof information to the base station. Then the base station can compute the result and verify the authenticity and integrity of it.

ETQ-RIV is an improvement scheme of EVTQ proposed in our previous work [4]. The improvements are as follows.

(i) In the data collecting process, each sensor node does not need to submit the out-bound HMAC, which further reduces the in-cell communication.

(ii) With processing query, the storage node needs to send one HMAC for each unqualified sensor node in EVTQ. However, in ETQ-RIV, only one HMAC for all noncontributed sensor nodes is required, which will greatly reduce the query communication cost.

The main contributions of this paper are as follows.

(i) We propose a sequence relationship encoding based method, which makes storage nodes unable to falsify or omit data items without being noticed.

(ii) We give the two concrete protocols, named as *data binding and collecting protocol* and *verifiable query response protocol*, to achieve *ETQ-RIV*, which make the base station capable of verifying the authenticity and integrity of the top-k query result.

(iii) We present the quantitative analysis of communication cost and redundancy rate of *ETQ-RIV* in detail and conduct experiments to evaluate the performance of this work compared to present methods.

The rest of this paper is organized as follows. The related works are presented in Section 2. Section 3 presents the preliminaries including related models, problem description, and performance evaluation index. Section 4 presents the two protocols proposed in this paper. Section 5 presents the theoretical analysis of the communication cost and redundancy rate. Section 6 presents the result and analysis of the simulation experiments. In last section, we make a conclusion for this paper.

2. Related Works

The traditional multihop model is not suitable for large scale wireless sensor networks; thus a novel two-tiered architecture was proposed, which is indispensable for increasing network capacity and scalability, reducing system complexity, and prolonging network lifetime [1–3].

In TWSNs, the storage nodes play an important role and are more attractive to adversaries, which may bring serious data security problems. Therefore, data privacy and security have been widely discussed in recent research work.

Attentions have been paid to secure top-k query processing in two-tiered sensor networks recently. A novel verifiable fine-grained top-k query processing scheme was proposed by Zhang et al. [5], which is the first work of verifiable top-k query scheme in two-tiered sensor networks. The basic idea of Zhang's scheme is that each sensor node generates *hashed message authentication code* (HMAC) [6] for every three consecutive data items to make query results verifiable. However, the long bits of HMACs result in high cost of communications. Liao and Li proposed a secure top-k query scheme named PriSecTopk [7] based on order preserving encryption [8] and message authentication code (MAC) [9], which could not lower the communication cost either. Ma et al. proposed a novel fine-grained verification scheme for top-k queries named VSFTQ [10] which uses symmetric encryption instead of message authentication. VSFTQ reduces the communication cost to a certain extent. On the basis of Zhang's scheme and Ma's scheme, Dai et al. proposed an efficient verifiable top-k query processing scheme in our previous work [4], which has a better performance in communication cost. Yu et al. proposed a dummy reading-based anonymization framework [11, 12], under which the query results can be guaranteed by verifiable top-k query (VQ) schemes they proposed. However, the VQ scheme requires each sensor node to send the neighboring sensor node IDs and hashes corresponding to individual genuine encrypted readings, which may lead to high communication cost of sensor nodes. Moreover, to obtain the top-k query result, the storage node is required to submit a top-$(\eta - 1 + k)$ query result due to the dummy readings, where η is a system parameter denoting the difference between the maximum and minimum encrypted readings within an epoch.

There is a special kind of top-k query when k equals one, which is also known as max/min query. For max/min query processing, Yao et al. proposed a preliminary privacy-preserving scheme [13] based on prefix membership verification (PMV) mechanism [14, 15] to compute the maximum or minimum value over encrypted data items. Dai et al. proposed an energy-efficient privacy-preserving MAX/MIN query processing solution [16] based on 0-1 encoding technique [17] which has a better performance compared to [13].

Besides top-k query, secure range query has been widely discussed recently [18–23]. The schemes proposed in [18–23]

ask for data items falling into specified ranges without data privacy disclosure and make query result verifiable.

3. Models and Problem Statement

3.1. Models. In this paper we consider the two-tiered network model as shown in Figure 1. The whole network is partitioned into *cells*, each of which consists of a storage node M and n homogeneous sensor nodes $\{s_1, s_2, \ldots, s_n\}$. M is abundant in resources such as energy, storage, and computation, which is in charge of not only storing of all sensing data collected by the sensor nodes in the same cell but processing and answering of the query request issued by the base station. Limited in resource, the sensor nodes just collect sensing data and transmit them to the storage node in the cell. As in [4], we take the assumption that storage nodes know the topological information of the whole cell, while a sensor node knows the locations of its neighboring nodes in one hop as well as the storage node's location of affiliated cells. In addition, the base station knows the topological information of the whole network.

In TWSN, a top-k query can be denoted as a three-element tuple $Q = (c, t, k)$, where c represents the query region, t represents the time epoch of the query, and k is the quantity of required data items. For simplicity of description, we discuss the specific top-k query that covers only one cell in one time epoch. It is easy to extend the proposed method to achieve queries including multiple cells over multiple time epochs.

3.2. Problem Statement. The threat model considered in this paper is similar to [4]. We assume that an arbitrary number of storage nodes could be compromised and instructed to respond with falsified and/or incomplete data as a top-k query result to the base station. In TWSN, both sensor nodes and storage nodes could be compromised. In general, a sensor node only has very little information and the vast majority in the whole network are uncompromised ones. However, the storage node plays a more important role in the network, for it stores all data items collected in the whole cell. Once a storage node is compromised, the adversary may falsify or abandon the data when the storage node answering the query issued by the base station, which will make the base station get incorrect or incomplete result and mislead the user into making wrong decisions. Therefore, storage nodes tend to be more attractive and vulnerable to adversaries. Although a compromised storage node may lead to leakage of data privacy, it is not concerned in this paper. In practice, in some application of WSNs, it is the data integrity instead of data privacy that is more important. For instance, video surveillance in a sensor network for building security is known to adversaries thus requiring no privacy. Therefore, we focus on the verifiable top-k query in this paper.

Given a top-k query $Q = (c, t, k)$, we assume the storage node M returns a data set R as query result. The problem of interest is how the base station can verify whether R satisfies both the *authenticity* and *integrity* requirements, which means the following two rules must hold.

(1) *Authenticity Rule*. All data items in R are surely collected by sensor nodes in the query cell c during time epoch t.

(2) *Integrity Rule*. There are exactly k data items that are indeed the k highest data items among what are collected by sensor nodes in the query cell c during time epoch t, or equivalently

$$\forall r_i \in R_t \wedge \forall d_j \in D_t \setminus R_t \longrightarrow r_i > d_j, \tag{1}$$

where D_t represents the data set consisting of all data items collected by sensor nodes in the query cell c during time epoch t.

3.3. Evaluation Metrics. In this paper, the following two metrics are used to evaluate the performance of the proposed scheme.

(1) Communication cost: including in-cell communication cost C_I and query processing communication cost C_Q. C_I represents the size in bits of data items transmitted from all the sensor nodes to the storage in a cell in the data binding and collecting procedure, while C_Q stands for the size of data items transmitted from the storage node in the query cell to the base station during the verifiable query response procedure.

(2) Redundancy rate of query result: the proportion of the total size in bits of additional proof information used to enable verifiable top-k queries to the total size in bits of response message in final query result. This rate, denoted as γ, indicates the efficiency of the query processing. A lower redundancy rate means less additional communication cost required for verification. The redundancy rate will be calculated by the following equation:

$$\gamma = \frac{l_s - l_r}{l_s} \times 100\%, \tag{2}$$

where l_s is the size in bits of the total data items returned by the storage node, while l_r is the total data size of final query result.

4. Verifiable Top-k Query Processing

4.1. Assumptions and Definition. Given a top-k query $Q_t = (c, t, k)$, we assume that the query covers just a cell c in a time epoch t, which we have mentioned in previous sections. There are n sensor nodes in the cell c, which can be denoted as $\Gamma = \{s_1, s_2, \ldots, s_n\}$. Assume that an arbitrary sensor node s_i $(1 \leq i \leq n)$ collects N data items denoted as $D_i = \{d_{i,1}, d_{i,2}, \ldots, d_{i,N}\}$ in each time epoch and sorts them into descending order. s_i shares a secret key K_i with the base station, which is used to encode the proof information for each data item by a HMAC function.

4.2. Data Binding and Collecting Protocol. In each time epoch, sensor node s_i collects data and encodes them before sending to the storage node M in the same cell for storage. The detailed procedure is as follows.

(1) s_i sorts the N data items $D_i = \{d_{i,1}, d_{i,2}, \ldots, d_{i,N}\}$ that it collects in time epoch t. Without loss of generality, we assume $d_{i,1} > d_{i,2} > \cdots > d_{i,N}$ after being sorted.

(2) According to the sequence of the sensing data items, s_i computes the message verification code $V(d_{i,j})$ for each data item. $V(d_{i,j})$ of the jth data item $d_{i,j}$ $(1 \le j \le N)$ can be computed as follows:

$$V\left(d_{i,j}\right) = \mathrm{HMAC}_{K_i}\left(t \parallel d_{i,1} \parallel d_{i,2} \parallel \cdots \parallel d_{i,j}\right), \quad (3)$$

where $\mathrm{HMAC}_{K_i}(\cdot)$ denotes encoding the corresponding data using a HMAC function with key K_i and $\parallel$ is the concatenation operator.

(3) s_i constructs a data collecting message MSG_C according to the following format and sends to the storage node M:

$$\mathrm{MSG}_C = \langle \mathrm{id}(s_i), t, d_{i,1}, V(d_{i,1}), d_{i,2}, V(d_{i,2}), \ldots, d_{i,N}, V(d_{i,N}) \rangle. \quad (4)$$

(4) M receives and stores the data collecting message of s_i. We denote the data set consisting of all the data items from all the sensor nodes in the cell as

$$\mathrm{DS}_t = \bigcup_{s_i \in \Gamma} \{d_{i,1}, \ldots, d_{i,N}\}. \quad (5)$$

4.3. Verifiable Query Response Protocol. Query processing and verification requires collaboration of the base station and the storage node, which works as follows. According to the query issued by the base station, the storage node M processes this query and responds with the corresponding data items. At last, the base station computes the query result and verifies the authenticity and integrity of the result. The protocol is described in detail as follows.

Phase 1 (query request transmission). The base station sends the query $Q_t = (t, k, c)$ to M and waits for its feedback.

Phase 2 (query message feedback). (1) After Q_t has been received, M computes the k highest data items, denoted as $\mathrm{top}k(\mathrm{DS}_t)$, according to all the sensing data DS_t which was sent during the time epoch t from the sensor nodes $\{s_1, s_2, \ldots, s_n\}$ in cell C. $\mathrm{top}k(\mathrm{DS}_t)$ satisfies the following condition:

$$\left|\mathrm{top}k\left(\mathrm{DS}_t\right)\right| = k \wedge \forall d_i, \quad d_j\left(d_i \in \mathrm{top}k\left(\mathrm{DS}_t\right) \wedge d_j \in \mathrm{DS}_t \setminus \mathrm{top}k\left(\mathrm{DS}_t\right)\right) \longrightarrow d_i > d_j. \quad (6)$$

Assuming that there are δ_i data items sent by s_i within $\mathrm{top}k(\mathrm{DS}_t)$, that is to say, $\{d_{i,1}, d_{i,2}, \ldots, d_{i,\delta_i}\} \subseteq \mathrm{top}k(\mathrm{DS}_t)$, then the following condition holds:

$$\delta_i = \left|D_i \cap \mathrm{top}k\left(\mathrm{DS}_t\right)\right|, \quad 0 \le \delta_i \le \min\{N, k\}. \quad (7)$$

(2) M constructs the following response message according to δ_i.

(i) Given a sensor node s_i where $\delta_i > 0$, we call it the contributed node, the response message of which is named as $\mathrm{msg}_1(s_i)$ and it should be computed as

$$\mathrm{msg}_1(s_i) = \begin{cases} \left\langle \mathrm{id}(s_i), d_{i,1}, \ldots, d_{i,\delta_i+1}, V\left(d_{i,\delta_i+1}\right)\right\rangle, & 0 < \delta_i \le N-1, \\ \langle \mathrm{id}(s_i), d_{i,1}, \ldots, d_{i,N}, V(d_{i,N})\rangle, & \delta_i = N, \end{cases} \quad (8)$$

where d_{i,δ_i+1} is the maximum data item collected by s_i that is not in $\mathrm{top}k(\mathrm{DS}_t)$ and we call it the out-bound of the s_i. If all the data items collected by s_i are in $\mathrm{top}k(\mathrm{DS}_t)$, its out-bound does not exist. Otherwise, there will be one and only one out-bound as for s_i.

(ii) We call the sensor node where $\delta_i = 0$ *noncontributed node*. There is only one response message msg_0 for all noncontributed nodes which should be computed as

$$\mathrm{msg}_0 = \left\langle \{\mathrm{id}(s_i), d_{i,1} \mid \delta_i = 0 \wedge s_i \in \Gamma\}, \bigoplus_{\delta_i = 0 \wedge s_i \in \Gamma} V(d_{i,1}) \right\rangle, \quad (9)$$

where $\bigoplus$ is the exclusive or operator.

(3) M summarizes all the response message generated in step (2), constructs the query feedback message MSG_Q, and sends it to the base station:

$$\mathrm{MSG}_Q = \langle \{\mathrm{msg}_1(s_i) \mid \delta_i > 0 \wedge s_i \in \Gamma\}, \mathrm{msg}_0 \rangle. \quad (10)$$

Phase 3 (query result computation and verification). (1) Upon receiving the query feedback message from M, the base station will do the preprocessing to confirm all the message $\{\mathrm{msg}_1(s_i) \mid \delta_i > 0 \wedge s_i \in \Gamma\}$ of each contributed node and the message msg_0 of all noncontributed nodes.

(2) The base station sorts all data items in the response message and gets the k highest data items, which is the top-k query result. We denote the top-k query result by R_t, the minimum data item of which is denoted as $\min(R_t)$.

(3) The base station checks whether the following conditions hold in sequence. If and only if all conditions are satisfied, the query result R_t is an authentic and complete result. Otherwise, the query result is abnormal.

Condition 1. All sensor ID in message $\{\mathrm{msg}_1(s_i) \mid \delta_i > 0 \wedge s_i \in \Gamma\}$ and msg_0 construct a set, named Ω, so that $\Omega = \{\mathrm{id}(s_i) \mid s_i \in \Gamma\}$ holds.

Condition 2. Assume that s_i is an arbitrary sensor node that contributes data items to message $\{\mathrm{msg}_1(s_i) \mid \delta_i > 0 \wedge s_i \in \Gamma\}$, and the response message of s_i is $\langle \mathrm{id}(s_i), d_{i,1}, d_{i,2}, \ldots, d_{i,\mu_i}, H_i\rangle$, where $d_{i,1} > d_{i,2} > \cdots > d_{i,\mu_i}$. The base station then computes $\mathrm{HMAC}_{K_i}(t \parallel d_{i,1} \parallel \cdots \parallel d_{i,\mu_i})$, where K_i is a distinct key known only to s_i and the base station. Then one of the following two conditions must hold:

(1) $\mu_i = N \wedge H_i = H_{k_i}(t \parallel d_{i,1} \parallel \cdots \parallel d_{i,N} \parallel \Psi) \wedge d_{i,l_i} \ge \min(R_t)$;

(2) $2 \le \mu_i < N \wedge H_i = H_{k_i}(t \parallel d_{i,1} \parallel \cdots \parallel d_{i,N} \parallel \Psi) \wedge d_{i,l_i}$.

Condition 3. Let $\{s_1, s_2, \ldots, s_p\}$ be all the noncontributed nodes and their response message msg_0 is $\langle \text{id}(s_1), d_1, \text{id}(s_2), d_2, \ldots, \text{id}(s_p), d_p, H_0 \rangle$, in which all the data items construct a set $\text{DS}' = \{d_1, d_2, \ldots, d_p\}$. The base station computes $\text{HMAC}_{K_1}(t \parallel d_1)$, $\text{HMAC}_{K_2}(t \parallel d_2)$, …, $\text{HMAC}_{K_p}(t \parallel d_p)$ using the key $K_1, K_2, \ldots, K_p$ shared with $s_1, s_2, \ldots, s_p$, respectively, and the following condition holds:

$$\left(H_0 = \bigoplus_{d_j \in \text{DS}'} \text{HMAC}_{K_j}\left(t \parallel d_j\right) \right) \wedge \left(\forall d_j \in \text{DS}' \longrightarrow d_j < \min\left(R_t\right) \right). \tag{11}$$

As described in previous two protocols, the sensing data items collected by a sensor node are sorted in descending order and encoded with a HMAC function. Each sensor node shares a secret key only with the base station, and the storage nodes know nothing about the keys. In such a case, it is computationally infeasible for the storage nodes to falsify the message verification code, as long as we choose a considerably complex HMAC algorithm such as SHA-1 [24]. Thus, it is impossible for storage nodes to falsify or conceal data items in the query result without being detected by the base station. Therefore, the scheme proposed in this paper enables authenticity and integrity verification of the top-k query result.

5. Protocol Analysis

5.1. Communication Cost Analysis. From *data binding and collecting protocol* and *verifiable query response protocol* we learn that there are two kinds of communication costs in two-tiered sensor networks, named as *in-cell communication cost* C_I and *query processing communication cost* C_Q. Assume that there are n sensor nodes in a cell and each node ID has l_{id} bits. Each sensor node collects N data items in every time epoch, each data item has l_d bits, and each HMAC code number has l_h bits. In addition, each time epoch number has l_t bits. Assume that there are L hops between each sensor node and the storage node on average. We assume that there are μ contributed nodes and n-μ noncontributed nodes, and λ of the μ contributed nodes contributes all the N data items to the query result. Obviously, we have $0 \le \lambda \le N$ and $\lambda \le k$ hold. According to such two protocols, we have

$$\begin{aligned} C_I &= n \cdot \left(l_{\text{id}} + l_t + N \cdot \left(l_d + l_h\right)\right) \cdot L, \\ C_Q &= \left(\mu \cdot l_{\text{id}} + \left(k + \mu - \lambda\right) \cdot l_d + \mu \cdot l_h\right) \\ &\quad + \left(\left(n - \mu\right) \cdot l_{\text{id}} + \left(n - \mu\right) \cdot l_d + l_h\right) \\ &= n \cdot l_{\text{id}} + \left(n + k - \lambda\right) \cdot l_d + \left(\mu + 1\right) \cdot l_h. \end{aligned} \tag{12}$$

5.2. Redundancy Rate Analysis. We define the redundancy rate γ of query result as the proportion of the total size of additional proof information used for query result verification to the total size of response message in final query result.

Table 1: Default evaluation parameters.

Para.	n	l_{id} (bit)	l_t (bit)	l_d (bit)	l_h (bit)	N	k
Val.	250	32	32	32	128	20	10

We take the same assumption as Section 5.1. According to our proposed protocols, we have

$$\begin{aligned} \gamma &= \frac{C_Q - \left(\mu \cdot l_{\text{id}} + k \cdot l_d\right)}{C_Q} \times 100\% \\ &= \frac{\left(n - \mu\right) \cdot l_{\text{id}} + \left(n - \lambda\right) \cdot l_d + \left(\mu + 1\right) \cdot l_h}{n \cdot l_{\text{id}} + \left(n + k - \lambda\right) \cdot l_d + \left(\mu + 1\right) \cdot l_h} \times 100\%. \end{aligned} \tag{13}$$

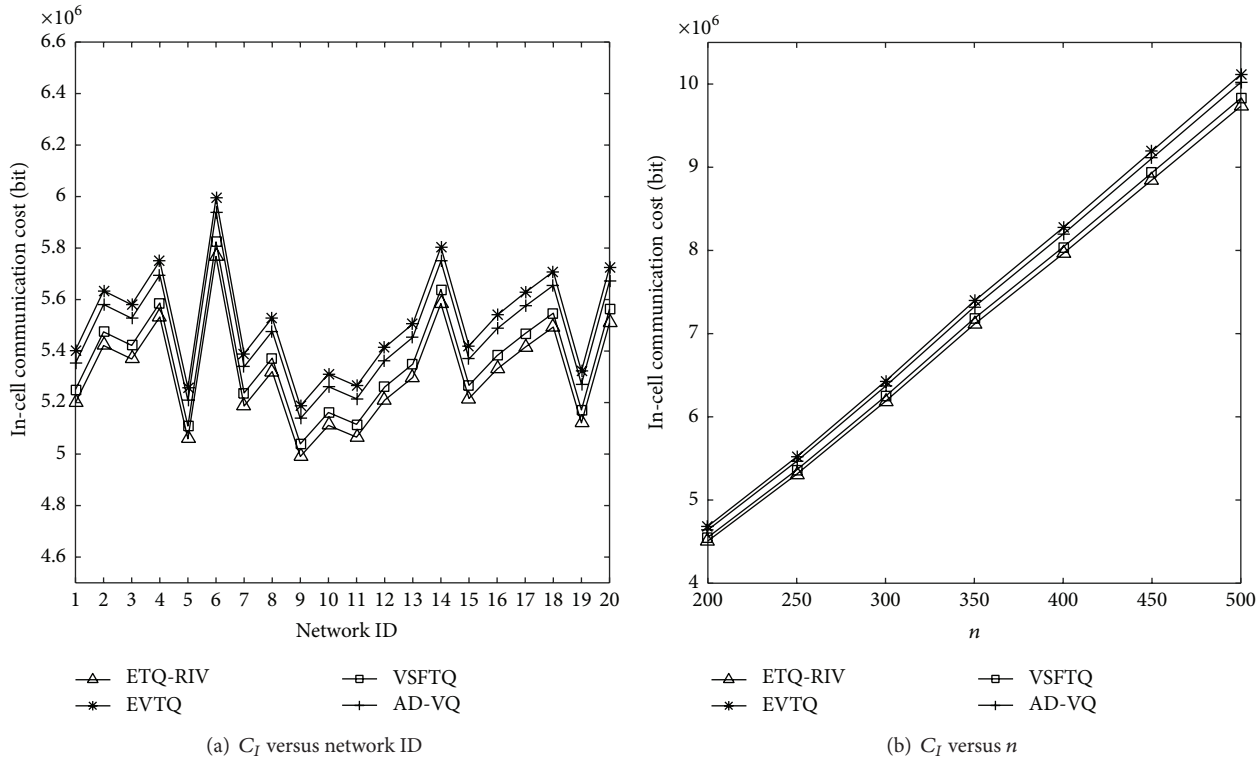
6. Performance Evaluation

In this section, we evaluate the performance of the scheme ETQ-RIV proposed in this paper and compare it with the schemes VSFTQ, EVTQ, and AD-VQ, which are proposed in [10], [4], and [11, 12], respectively. The four schemes VSFTQ, EVTQ, ETQ-RIV, and AD-VQ are implemented on the simulator of [25] with random sensor data items. We compare the in-cell communication cost C_I, query processing communication cost C_Q, and the query result redundancy rate γ of the four schemes, respectively. We also assume that the packet transmissions are both collision-free and error-free in our experiments.

We take the assumption that we carry out the top-k query just in one cell with one storage node and n sensor nodes. The n sensor nodes are distributed uniformly over a two-dimensional region which covers an $80 * 80\,\text{m}^2$ area. Each sensor node collects N data items during each time epoch and its communication radius is 10 meters. The bit lengths of sensor node ID, time epoch number, sensing data item, and HMAC code are represented by l_{id}, l_t, l_d, and l_h, respectively. The default parameters are listed in Table 1, which are used in the evaluations unless otherwise specified.

In each measurement, we generate 20 different networks with different network IDs. In each network, the sensor nodes are distributed randomly with different topology. The measurement result is the average of 20 networks.

6.1. In-Cell Communication Cost Evaluation. With default parameters, we evaluate the in-cell communication costs of VSFTQ, EVTQ, ETQ-RIV, and AD-VQ in different networks, the result of which is shown in Figure 2(a). Figure 2(a) indicates that our proposed scheme ETQ-RIV takes a little less communication costs than VSFTQ and AD-VQ, while EVTQ takes the highest communication costs. Compared with EVTQ, AD-VQ, and VSFTQ, ETQ-RIV saves about 3.77%, 2.86%, and 0.97% of in-cell communication costs on average, respectively. The reason is that in each scheme sensor nodes are required to submit every data item along with some proof information. EVTQ, AD-VQ, and ETQ-RIV use HMAC as proof information. In EVTQ, each sensor node submits one more out-bound HMAC besides every data item's proof information. In AD-VQ, each sensor requires submitting some information of neighboring nodes besides



(a) C_I versus network ID

(b) C_I versus n

Figure 2: Evaluation results of in-cell communication costs.

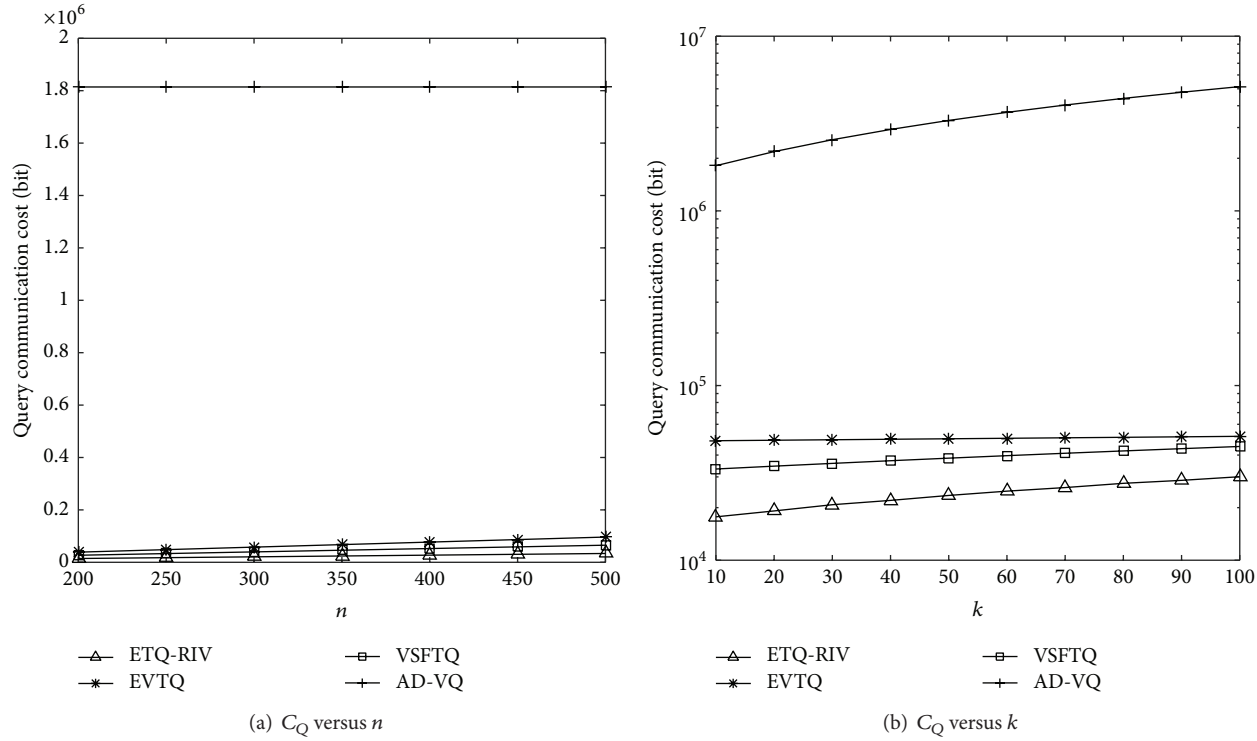
the HMAC information. In ETQ-RIV, the sensor nodes do not need the out-bound HMAC and neighboring nodes information, so ETQ-RIV performs better than EVTQ and AD-VQ. Different from the other three schemes, VSFTQ replaces the HMAC with symmetric encryption of data item's order, score, and the time epoch as proof information. In our evaluation, the total bit length of proof information in VSFTQ is more than ETQ-RIV. As a result, ETQ-RIV performs better than VSFTQ.

Figure 2(b) indicates that the in-cell communication costs of the three schemes increase as the sensor node number n increases. ETQ-RIV has the best performance as before and saves about 3.77%, 2.86%, and 0.97% of in-cell communication costs on average compared with EVTQ, AD-VQ, and VSFTQ, respectively. The reason is same as the description in the previous section.

6.2. Query Communication Cost Evaluation. We evaluate top-k query communication costs of VSFTQ, EVTQ, ETQ-RIV, and AD-VQ as n as well as k increases. Figure 3(a) indicates that the query communication costs of the four schemes all increase as n increases. The reason is obvious: the larger n, the more noncontributed sensor nodes, and the more verification information for noncontributed nodes required to be submitted. Figure 3(a) also shows that ETQ-RIV has the lowest query communication cost, followed by VSFTQ, EVTQ, and then AD-VQ. In detail, the query communication cost of ETQ-RIV is about 98.67% lower than AD-VQ, 64.30% lower than EVTQ, and 47.69% lower than VSFTQ. In AD-VQ, to obtain the top-k query result, the storage node is required to submit a top-$(\eta - 1 + k)$ query result due to the dummy readings, where η is a system parameter denoting the difference between the maximum and minimum encrypted readings within an epoch. That is why AD-VQ has the highest query communication cost. In EVTQ, the storage node is required to send a HMAC as proof information for each unqualified sensor node that has no data item satisfying the query, so its query communication cost is high. In VSFTQ, the storage node needs to send symmetric encryption of data item's order, score, and the time epoch as proof information, which takes more communication cost than ETQ-RIV. However, in ETQ-RIV, only one HMAC is needed for all noncontributed sensor nodes, which greatly reduces the query communication cost.

Figure 3(b) shows that the query communication costs of the four schemes all increase as k increases, which is because the larger k, the more data items that need to be returned by the storage node. Similar to Figure 3(a), ETQ-RIV still has the lowest query communication cost, which is about 99.31% lower than AD-VQ, 51.70% lower than EVTQ, and 38.44% lower than VSFTQ. The reason is as mentioned before: only one HMAC is required for all noncontributed sensor nodes in ETQ-RIV.

6.3. Redundancy Rate Evaluation. We evaluate top-k query result redundancy rate of VSFTQ, EVTQ, ETQ-RIV, and AD-VQ as k increases. Figure 4 indicates that the redundancy rates of the four schemes all decrease as k increases. Furthermore, AD-VQ has the highest redundancy rate followed by EVQT, VSFTQ, and ETQ-RIV. The redundancy rate of



(a) C_Q versus n (b) C_Q versus k

Figure 3: Evaluation results of query communication costs.

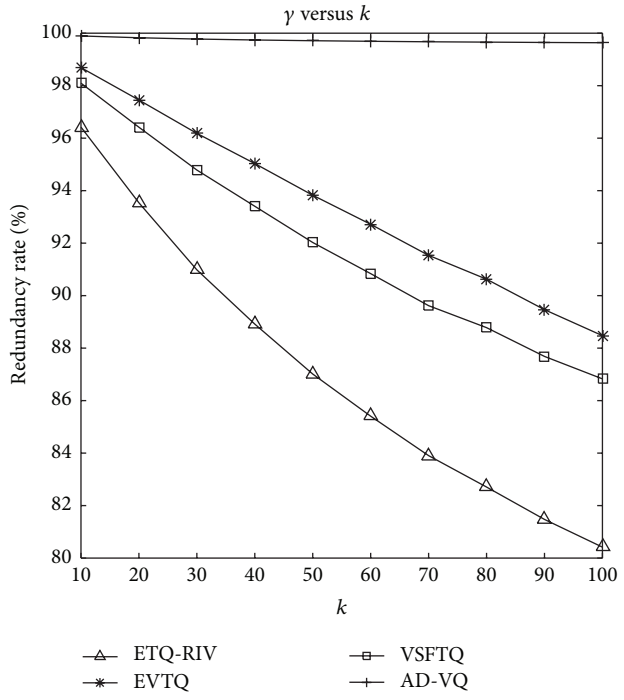


Figure 4: Evaluation results of query result redundancy rate.

ETQ-RIV is about 12.68% lower than AD-VQ, 6.77% lower than EVTQ, and 5.19% lower than VSFTQ. The reason is as follows. A top-k query result can be divided into two parts: the satisfied data items of query result and the verification information. A larger k implies more satisfied data items and a lower redundancy rate. Compared with the other three schemes, only one HMAC is required for all noncontributed sensor nodes in ETQ-RIV, which makes the ETQ-RIV lowest redundancy rate.

According to the above evaluations and analysis, we can conclude that our proposed ETQ-RIV has better performance than the existing works [4, 5, 7, 10–12] both in communication costs and redundancy rate.

7. Conclusions

In this paper, we focus on the problem of verifiable top-k query in two-tiered wireless sensor networks and propose an efficient top-k query processing scheme with result integrity verification which is denoted as ETQ-RIV. To make the query result verifiable, each sensor node should submit some encoded message containing the sequence relationship as proof information for verification along with their collected sensing data items. Evaluation results show that ETQ-RIV can decrease the redundancy rate of query result and thus decrease both in-cell and query communication costs and performs better than the existing works in communication costs.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under the Grants nos. 61300240, 61402014, 61472193, 61373137, 61373138, 61272084, and 61201163, the Natural Science Foundation of Jiangsu Province under the Grants nos. BK20151511 and BK20141429, the Project of Natural Science Research of Jiangsu University under Grants nos. 11KJA520002 and 14KJB520027, the Postdoctoral Science Foundation of China under Grant no. 2013M541703, and the Postdoctoral Science Foundation of Jiangsu Province under Grant no. 1301042B.

References

[1] O. Gnawali, K. Y. Jang, J. Paek et al., "The tenet architecture for tiered sensor networks," in *Proceedings of the 4th ACM Conference on Embedded Networked Sensor Systems*, pp. 153–166, Boulder, Colo, USA, 2006.

[2] P. Desnoyers, D. Ganesan, and P. Shenoy, "TSAR: a two tier sensor storage architecture using interval skip graphs," in *Proceedings of the 5th ACM Conference on Embedded Networked Sensor Systems*, pp. 39–50, San Diego, Calif, USA, November 2005.

[3] Y. Diao, D. Ganesan, G. Mathur, and P. J. Shenoy, "Rethinking data management for storage-centric sensor networks," in *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR '07)*, pp. 22–31, Asilomar, Calif, USA, January 2007.

[4] H. Dai, G. Yang, F. Xiao, and Q. Zhou, "EVTQ: an efficient verifiable top-k query processing in two-tiered wireless sensor networks," in *Proceedings of the 9th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN '13)*, pp. 206–211, IEEE, Dalian, China, December 2013.

[5] R. Zhang, J. Shi, Y. Liu et al., "Verifiable fine-grained top-k queries in tiered sensor networks," in *Proceedings of the 29th IEEE International Conference on Computer Communications*, pp. 1199–1207, San Diego, Calif, USA, 2010.

[6] H. Krawczyk, R. Canetti, and M. Bellare, "HMAC: keyed-hashing for message authentication," Tech. Rep. RFC 2104, Internet Society, Reston, Va, USA, 1997.

[7] X. Liao and J. Li, "Privacy-preserving and secure top-k query in two-tier wireless sensor network," in *Proceedings of the Global Communications Conference*, pp. 335–341, Anaheim, Calif, USA, 2012.

[8] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order-preserving encryption for numeric data," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 563–574, Paris, France, June 2004.

[9] R. Rivest, "The MD5 message-digest algorithm," Tech. Rep. RFC 1321, Internet Society, Reston, Va, USA, 1992.

[10] X. Ma, H. Song, J. Wang, J. Gao, and G. Min, "A novel verification scheme for fine-grained top-k queries in two-tiered sensor networks," *Wireless Personal Communications*, vol. 75, no. 3, pp. 1809–1826, 2014.

[11] C. M. Yu, N. K. Guo, Y. Chen et al., "Top-k query result completeness verification in sensor networks," in *Proceeding of the IEEE Communications Workshops (ICC '13)*, pp. 1026–1030, Budapest, Hungary, 2013.

[12] C. M. Yu, N. K. Guo, Y. Chen et al., "Top-k query result completeness verification in tiered sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 109–124, 2014.

[13] Y. Yao, N. Xiong, J. H. Park, L. Ma, and J. Liu, "Privacy-preserving max/min query in two-tiered wireless sensor networks," *Computers & Mathematics with Applications*, vol. 65, no. 9, pp. 1318–1325, 2013.

[14] J. Cheng, H. Yang, S. H. Wong, P. Zerfos, and S. Lu, "Design and implementation of cross-domain cooperative firewall," in *Proceedings of the IEEE International Conference on Network Protocols (ICNP '07)*, pp. 284–293, Beijing, China, October 2007.

[15] A. X. Liu and F. Chen, "Collaborative enforcement of firewall policies in virtual private networks," in *Proceedings of the 27th ACM Symposium on Principles of Distributed Computing*, pp. 95–104, ACM, August 2008.

[16] H. Dai, G. Yang, and X. Qin, "EMQP: an energy-efficient privacy-preserving MAX/MIN query processing in tiered wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, 11 pages, 2013.

[17] H. Y. Lin and W. G. Tzeng, "An efficient solution to the millionaires' problem based on homomorphic encryption," in *Proceedings of the 3rd International Conference on Applied Cryptography and Network Security*, pp. 97–134, New York, NY, USA, 2005.

[18] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proceeding of the 27th IEEE International Conference on Computer Communications*, pp. 46–50, IEEE, Phoenix, Ark, USA, April 2008.

[19] B. Sheng and Q. Li, "Verifiable privacy-preserving sensor network storage for range query," *IEEE Transactions on Mobile Computing*, vol. 10, no. 9, pp. 1312–1326, 2011.

[20] J. Shi, R. Zhang, and Y. Zhang, "Secure range queries in tiered sensor network," in *Proceedings of the 28th IEEE International Conference on Computer Communications*, pp. 945–953, Rio de Janeiro, Brazil, 2009.

[21] J. Shi, R. Zhang, and Y. Zhang, "A spatiotemporal approach for secure range queries in tiered sensor networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 264–273, 2011.

[22] F. Chen and A. X. Liu, "SafeQ: secure and efficient query processing in sensor networks," in *Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM '10)*, pp. 1–9, IEEE, San Diego, Calif, USA, March 2010.

[23] Y. Yi, R. Li, F. Chen, A. X. Liu, and Y. Lin, "A digital watermarking approach to secure and precise range query processing in sensor networks," in *Proceedings of the International Conference on Computer Communications (IEEE INFOCOM '13)*, pp. 1950–1958, Turin, Italy, April 2013.

[24] D. Eastlake and P. Jones, "US secure hash algorithm 1 (SHA1)," RFC 3174, Internet Society, Reston, Va, USA, 2001.

[25] A. Coman, J. Sander, and M. A. Nascimento, "Adaptive processing of historical spatial range queries in peer-to-peer sensor networks," *Distributed and Parallel Databases*, vol. 22, no. 2-3, pp. 133–163, 2007.