Article

Securing Fine-grained Spatio-temporal Top-*k* Query in TMWSNs: a Novel Scheme

Jie Min [1]*, Xingpo Ma* [2], and Hongling Chen [3]

[1] School of Information Engineering, Xinyang Agriculture and Forestry University, Xinyang 464000, Henan, P.R.China;
[2] School of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, Henan, P.R.China;
[3] Guangdong Polytechnic of Science and Technology, Zhuhai 519090, Guangdong, P.R.China.
* Correspondence: dreamer@xyafu.edu.cn;maxingpo@xynu.edu.cn

Abstract: To ensure the security of spatial-temporal Top-*k* query in two-tiered wireless sensor networks, many schemes have been proposed in the literature in the past decade. However, most of them only consider the scenario where sensor nodes are static, and cannot achieve the security goal for spatial-temporal Top-*k* query in mobile sensor networks, because the mobility of the sensor nodes will affect the spatial-temporal relationships of the sensory data items generated by the sensor nodes. Although we have proposed some schemes for two-tiered mobile wireless sensor networks (TMWSNs) in our previous work, there is still large room to improve their performances. In this paper, we proposed a novel scheme named STQ-TMWSN for secure fine-grained spatial-temporal Top-*k* query in TMWSNs based on the virtual-grid construction and the size-order encryption binding. Theoretic analysis shows that STQ-TMWSN can achieve low computation complexity and high security performance. Simulation results indicate that STQ-TMWSN brings much lower communication cost than the state-of-the-art schemes on securing Top-*k* query in TMWSNs.

Keywords: Two-tiered mobile wireless sensor networks; Internet of Things; fine-grained spatial-temporal Top-*k* query; privacy preservation; completeness verification

1. Introduction

As one important component of Internet of Things (IoT) [1], wireless sensor networks (WSNs) [2] can be used in many application scenarios and are still being studied [3] by many researchers even though extensive research has been carried out on WSNs for the past two decades. In recent years, two novel variants of WSNs, namely two-tiered wireless sensor networks (TWSNs) [4] and two-tiered mobile wireless sensor networks (TMWSNs) [5], attract more and more attention from both the industial and the research communities since they perform much better than the traditional WSNs on the scalability, the flexibility and the robustness. Such advantages are mainly brought by the two-tiered architecture design of TMWSNs. Specifically, the lower tier of TMWSNs is composed of many mobile sensor nodes, while the upper tier consists of some storage nodes. The mobile sensor nodes at the lower tier are responsible of monitoring the physical environments around and generating sensory data items, which can be transmitted directly to the nearby storage nodes at the upper tier to get stored or processed. Users can retrieve the sensor data items which they are interested in by launching some kinds of queries, such as spatial-temporal Top-*k* query, to the storage nodes which will send the query results to the users through on-demand wireless links after processing the queries.

However, since TMWSNs are usually deployed in some insecure environments where human supervision is often lacked, the query processing on the storage nodes in TMWSNs are facing security threats. By compromising the storage nodes, the adversaries can launch many powerful attacks which are much more serious than by capturing some sensor nodes in TMWSNs. Once a storage node is compromised, all the sensory data items stored on it can be disclosed, and the procedures of query processing on it are no longer trustworthy either.

In this paper, we focus on the problem of secure spatial-temporal Top-k query in TMWSNs, considering the storage nodes may be compromised. A spatial-temporal Top-k query is defined as a query which aims to find out the qualified top k sensory data items generated in the queried region and the queried time interval [5]. Our aims are to preserve the privacy of the data items stored on the storage nodes and protect the integrity of the spatial-temporal Top-k query results.

To our best knowledge, there have been only a few works studying the problem of securing spatial-temporal Top-k query in TMWSNs at present. Most of the existing secure Top-k query processing schemes are proposed for cloud computing[6,7] and TWSNs[8,9]. Those proposed for cloud computing are not fit for TMWSNs because of the following reasons:

- Top-k queries in the cloud are generally securely processed based on the data which are outsourced on cloud servers by the same data owner. In cloud computing, the data owner knows all its outsourced data and thus can construct the tree-based index (e.g., IR-tree[10]), the binary heap[11]or other tree-like structures based on the whole data set to facilitate Top-k query without losing data privacy; while in TMWSNs, expect for the storage nodes which are considered as not fully trusted, there is no such a data owner who knows all the sensory data generated by all the sensor nodes and thus cannot construct the data-privacy-reservation index easily.
- Secure Top-k query schemes in the cloud are based on the strong processing ability and rich resources of the cloud servers. The nodes in TMWSNs, especially the sensor nodes are usually resource-limited and weak in computing, and the relatively weight encryption technologies which are used to preserve the privacy of the data in the cloud are not fit for TMWSNs.

In addition, the secure Top-k query processing schemes proposed for TWSNs are not fit for TMWSNs either, because they cannot preserve the integrity of the spatial-temporal Top-k query results in TMWSNs. In fact, attackers can launch much more covert attacks in TMWSNs than in TWSNs. When a mobile sensor node travels from the queried region to other regions or vice versa in the queried time interval, some sensory data generated by the sensor node may be in the queried region, and others may not. Obviously, the sensory data generated out of the queried region by the traveling sensor node are not the qualified ones which satisfy the requirements of the spatial-temporal Top-k query. However, few securing Top-k query schemes proposed in TWSNs consider this, which leaves leaks for the attackers to launch new kinds of covert attacks. For example, the attackers may replace the data items which are generated in the queried region by a sensor node with those produced out of the queried region by the same sensor node.

The above-mentioned reasons motivate us to make a profound study on securing spatial-temporal Top-k query in TMWSNs. In summary, the main contributions of this paper are three fold:

- It proposes a novel scheme named STQ-TMWSN (STQ is short for spatial-temporal Top-k query) to preserve the privacy of the data stored on storage nodes and protect the integrity of the spatial-temporal Top-k query results in TMWSNs. A series of protocols are developed in STQ-TMWSN for different components of TMWSNs.
- It provides sound theoretical analysis on the the security of STQ-TMWSN. It is proved in the paper that STQ-TMWSN is not only able to preserve the privacy of the sensory data items and their corresponding scores, but also detect the incomplete query results successfully for spatial-temporal Top-k query under the security model presented in this paper.
- Extensive simulations were conducted in the paper, and the results show that STQ-TMWSN is much more efficient than the related state-of-the-art schemes.

The remainder of this paper is organized as follows. Section II summaries the related schemes; Section III describes the system model, the security model, the definitions of some terminologies and the problem statement; Section IV presents the proposed scheme STQ-TMWSN in detail; Section V analyzes the security of STQ-TMWSN; In Section VI, STQ-TMWSN is compared with the related state-of-the-art schemes through extensive simulations; Section VII concludes this paper.

2. Related Works

2.1. Securing Top-k Queries in TWSNs

The study of securing Top-k queries in TWSNs was originally launched by the authors in [12], where three schemes are proposed to preserve the completeness of the Top-k query results in TWSNs. The three schemes were proposed based on the MAC (Message Authentication Code) technique, which requires each sensory data item to be attached with a MAC as its proof data. Then, many other schemes which use the similar technique appeared, such as those in [12–16]. However, the MAC-based technique is relatively less efficient because attaching a MAC to each sensory data item brings large quantity of extra data since a MAC takes almost 40% of the volume of a sensory data item according to [12].

Besides the MAC-based technique, some other methods were also proposed to ensure the privacy of the sensory data and the completeness of the Top-k query results in TWSNs, such as inserting digital watermarks or dummy readings into the normal ones [17] and constructing data aggregation trees [18,19]. However, inserting digital watermarks or dummy readings into the measure data makes it hard and complicated for the network owner to extract the normal readings from the hybrid ones, and it also brings a lot of redundant data, which further leads to the increase of the communication cost of both the sensor nodes and storage nodes.

What is more, one of the most important common points of these schemes is that they are all proposed for TWSNs where nodes are static, and they cannot perfectly treat the security threats faced by spatial-temporal Top-k query in TMWSNs as it is described in Section I.

2.2. Securing Top-k Queries in TMWSNs

The first work on securing Top-k queries in TMWSNs was done by F. Liu et al in 2015[20], when they presented a novel network architecture, namely TMWSNs, and proposed a scheme VTMSN to ensure the completeness of spatial-temporal Top-k query in TMWSNs. The main techniques used in VTMSN are symmetric encryption and information binding. Specifically, it binds the score of each sensory data item with its corresponding generation time, location, and value ranking order by concatenating and encrypting them with the kept symmetric key. Although VTMSN increases the difficulty for the attackers to undermine the completeness of the query results because of the binding relationships, it still has shortcomings. One is that it cannot preserve the privacy of the sensory data items since it leaves the data items disclosed to the storage nodes for ease of Top-k query processing on them; another one is that there should be large volume of location data transported together with the sensed readings, which greatly increases the communication cost of the sensor nodes and storage nodes.

To overcome the latter shortcoming of VTMSN, Haiqin Wu et al proposed a scheme named EVTopk [21] in 2016. EVTopk achieves completeness preservation of the Top-k query results by using the HMAC (Hash Message Authentication Code), which is formed by making hashing and encryption operations on the concatenated items including the score, the location, and the neighboring HMAC. However, since each sensory data item should be attached with an HMAC in EVTopk, the HMACs account for a large proportion of the data reports of the sensor nodes and the query results. Moreover, EVTopk is not able to achieve data privacy preservation either. In [22], a comparative study was made on the two schemes EVTopk and VTMSN. To further decrease the volume of the proof data in the data reports and the query results, in 2018, a scheme named VIP-TQ was proposed to preserve the integrity of the query results for spatial-temporal Top-*k* query in TMWSNs. In VIP-TQ, each sensory data is bound together with its location as well as its neighboring data score using pairwise-key-based encryption. Although the binding can effectively prevent the compromised storage nodes from undermining the integrity of the Top-*k* query results, it leaves the scores of the sensory data disclosed to the storage nodes, which increases the risk of divulging the privacy of the sensory data. In the same years, Xingpo Ma et al proposed two other schemes, namely SSSTQ1 and SSSTQ2 [5], for securing spatial-temporal Top-*k* in TMWSNs. However, one of the encryption technologies used in the two schemes is OPES [23] which has been proposed more than 15 years and is not strongly secure any more. Besides, large number of original locations associated with the sensory data items are added into the data reports and the query results for integrity verification, which heavily increases the communication cost of the systems. Such shortcomings in existing schemes motivate us to propose a novel, secure and efficient scheme for fine-grained spatial-temporal Top-*k* query in TMWSNs in this paper.

3. Models, Notations and Problem Statement

3.1. System Model

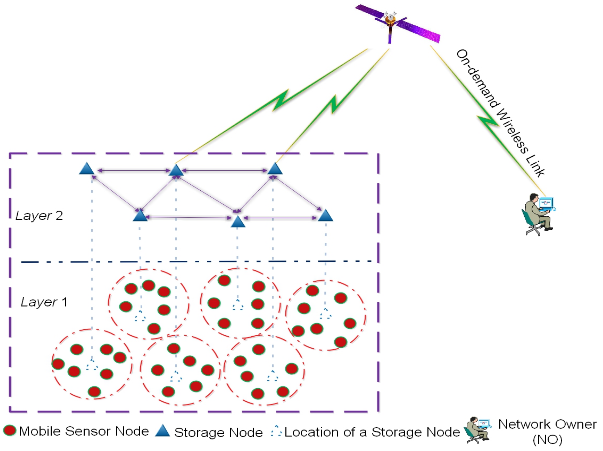


Figure 1. System model of TMWSNs.

The system model of TMWSNs is shown in Figure 1. The Nodes in TMWSNs are deployed into disjointed cells, each of which has more than one mobile sensor nodes and only one storage node. Mobile sensor nodes do not move all the time. They stay at some target locations for certain time intervals when they reach the positions, and go on moving to other target locations if it is necessary. We assume that the mobile sensor nodes only generate sensory data items when they are staying at their target locations. Moreover, we just consider the case that the mobile sensor nodes only move within their own cells in this paper, and leaves the more complex case that all sensor nodes can move anywhere of the deployment field in our future work. Since it will cost much more energy for the sensor nodes to move than do other tasks, such as computation and communication, a lot of energy will be saved if the sensor nodes do not move so far (e.g., moving just within the cells where the sensor nodes were originally deployed).

Time is divided into epochs. At the end of each epoch, all the sensor nodes in each cell transmit their sensory data to the corresponding storage node in their own cell. Each storage node converges and fuses the sensory data items collected in its own cell and stores them for further processing. Moreover, storage nodes also provide data-and-query processing services (e.g., processing and replying to spatial-temporal Top-*k* query) for the network owner, which can launch and send the queries to the storage nodes as well as get the query results from the storage nodes through the on-demand wireless links [12].

3.2. Notations and Terminologies

In this subsection, we introduce the notations and describe the definitions of some terminologies used in this paper. We use the set $\{D^t_{i,j,1}, D^t_{i,j,2}, ..., D^t_{i,j,\mu^t_{i,j}-1}, D^t_{i,j,\mu^t_{i,j}}\}$ to denote the sensory data items generated by sensor node S_i at its j^{th} target location in the t^{th} epoch T^t, where $\mu^t_{i,j}$ is the total number of the sensory data items generated by S_i at its j^{th} target location in T^t. For any sensory data item $D^t_{i,j,x}$, its corresponding data score $d^t_{i,j,x}$ can be worked out using a public scoring function $f(*)$ [24], namely $d^t_{i,j,x} = f(D^t_{i,j,x})$. Without loss of generality, we assume different sensory data items have distinct scores. Moreover, in order to facilitate presentation, we assume the ranking orders of the sensory data items generated by any sensor node at a target location are consistent with their subscript digital numbers. For example, there is $D^t_{i,j,1} < D^t_{i,j,2} < ... < D^t_{i,j,\mu^t_{i,j}-1} < D^t_{i,j,\mu^t_{i,j}}$, where i and j are the node ID and the target location ID of S_i respectively. The specific meanings of the notations used in this paper are listed in Table 1.

Table 1. Notations and Their Meanings.

Notations	Meanings
S_i	The sensor node whose ID is $i(0 < i \leq N)$
N	Totoal number of sensor nodes in one cell
T^t	The t^{th} epoch
λ^t_i	Total number of target locations of S_i in T^t
$Loc^t_{i,j}$	The j^{th} target location of S_i during T^t
$\mu^t_{i,j}$	Total data item numbers of S_i generated at $Loc^t_{i,j}$ in T^t
$n^t_{i,j}$	Total number of the qualified Top-k data items generated by S_i at $Loc^t_{i,j}$ in T^t
Q^t	A spatial-temporal Top-k query
R^t	The query result of Q^t
I_{Q^t}	The ID of Q^t
I_C	The ID of a given cell C
QR_{I_C}	The queried region in cell I_C
Key^t_i	The pairwise key shared by S_i and the network owner in T^t
$RT^t_{S_i}$	The data report generated by S_i in T^t
$E_{Key^t_i}\{*\}$	Symmetric encrypting operation with Key^t_i based on [25]
$E_{OPE}\{*\}$	Encrypting operation based on the OPE encryption scheme [26]
$RST^t_{S_i}$	The processed result of $RT^t_{S_i}$
Ω_i	Total number of the queried locations encrypted in $RST^t_{S_i}$
$\gamma^t_{i,j}$	Total number of the sensory data items encrypted in $DPP^t_{i,j}$
R_{tpk}	Set of the qualified Top-k data items extracted from R^t

We define the terminologies used in this paper as follows:

- **Fine-grained Spatial-temporal Top-k Query:** Given a cell whose ID is I_C in TMWSNs, an epoch T^t, and a parameter k, a fine-grained spatial-temporal Top-k query is defined as the query which tries to find out the top k sensory data items that have the biggest (or the smallest) scores among all the sensory data items generated in QR_{I_C} in T^t, where QR_{I_C} is a sub-region of the cell whose ID is I_C. The metalanguage of a fine-grained spatial-temporal Top-k query Q^t is described as in Eq.(1).

$$Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\} \tag{1}$$

- **Queried Node and Queried Location:** given a spatial-temporal Top-k query $Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}$, for anyone of any sensor node's target locations in epoch T^t, if it falls in QR_{I_C}, then it is called a queried location, and the corresponding sensor node is called a queried node.
- **Qualified Top-k Data Items:** given a spatial-temporal Top-k query $Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}$ if a sensory data item $D^t_{qualified}$ satisfies the following two conditions, it is called the qualified Top-k data item of Q^t: 1) $D^t_{qualified}$ was generated in QR_{I_C} and T^t; 2) Among all the sensory data items generated in QR_{I_C} and T^t, there are at least $N_{Q^t} - k$ data items whose scores are smaller (or bigger) than the score of $D^t_{qualified}$, where N_{Q^t} refers to the total number of the sensory data items generated in QR_{I_C} and T^t.
- **Data-proof Packet $DPP^t_{i,j}$:** for any target location $Loc^t_{i,j} (0 < j \leq \lambda^t_i)$ of any sensor node S_i $(1 \leq i \leq N)$, Data-proof Packet $DPP^t_{i,j}$ refers to the subreport produced by S_i for the sensory data generated at $Loc^t_{i,j}$ during T^t. Specifically, $DPP^t_{i,j}$ consists of the pairwise-key-encrypted sensory data items and the OPE-encrypted scores ('OPE' is short for 'Order-preserving Encryption'[26]) as well as some proof information generated by S_i at $Loc^t_{i,j}$ during T^t. More specific contents of $DPP^t_{i,j}$ will be shown in Algorithm 1 in Section IV.

3.3. Security model

In TMWSNs, storage nodes are assumed not only curious but also malicious, while other kinds of nodes are trustworthy. Comparing with the assumption that the edge server nodes are just curious in many existing schemes, this assumption is much closer to the cases in real applications. Specifically, we assume a curious storage node will try its best to disclose the sensory data items as well as the data scores computed based on the public scoring function, and a malicious storage node will do its best to undermine the completeness of the spatial-temporal Top-k query results. To achieve the malicious attack, a compromised storage node may put none or only part of the qualified top k data items into the Top-k query result, and it may also put some fabricated data items and/or the unqualified-but-real ones into the query result when processing a spatial-temporal Top-k query. For example, suppose the complete query result should be $\{D^t_1, D^t_2, D^t_3\}$. Then an incomplete query result may be $\{D^t_1\}$ or $\{D^t_1, D^t_4, D^t_{fabricated}\}$, where D^t_4 is a real-but-unqualified sensory data item and $D^t_{fabricated}$ is a fabricated data item. The attack model is illustrated in Fig 2.

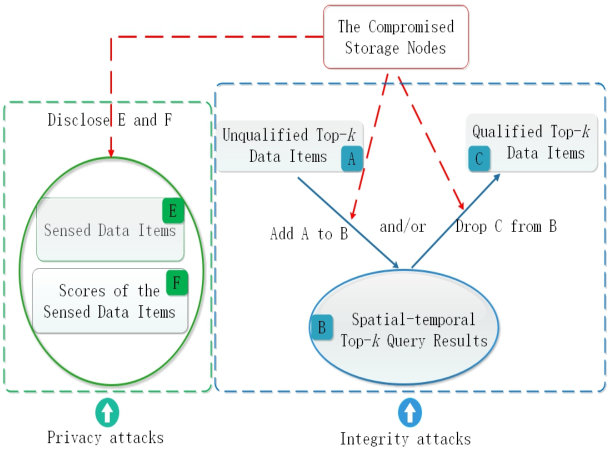

The Compromised Storage Nodes
Disclose E and F
Sensed Data Items
E
Scores of the Sensed Data Items
F
Unqualified Top-k Data Items
A
Qualified Top-k Data Items
C
Add A to B
and/or
Drop C from B
B
Spatial-temporal Top-k Query Results
Privacy attacks
Integrity attacks

Figure 2. Illustration of the attack model.

In our security model, the privacy of the sensory data items, which are generated by the mobile sensor nodes in TMWSNs, and their corresponding scores should be protected.

Other information, such as spatial-temporal Top-k query and the generation locations of the sensory data items, will be leaked to storage nodes. It is hard to enable storage nodes to process spatial-temporal Top-k query smoothly and successfully without such leaks. Fortunately, the leaked information brings little threat to the safety of the systems. Moreover, we assume each mobile sensor node is assumed to be equipped with the tamper-proof hardware, with the help of which the adversaries cannot disclose the encryption materials stored in the hardware even if they capture the sensor nodes [9].

3.4. Problem Statement and Design Goal

Under the system and the security models described above, the problem tackled in this paper can be presented in a word as follows: Given a spatial-temporal Top-k query $Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}$, how to obtain its query result from the storage nodes without disclosing the sensor data items and their corresponding scores to the storage nodes and verify the completeness of the query result correctly and efficiently in TMWSNs.

Specifically, our design goal is to propose a novel scheme which enables efficient, privacy-preservation and integrity-verifiable query processing for spatial-temporal Top-k query in TMWSNs. Specifically, three objects as follows should be achieved.

- *The privacy-preservation goal*: Our proposed scheme should preserve the privacy of the sensory data items and their scores collected from the mobile sensor nodes. Without losing the ability of processing spatial-temporal Top-k query, storage nodes in the systems must be not able to disclose the sensory data items and their scores.
- *The integrity-verification goal*: Our proposed scheme should enable the network owner to verify the integrity of spatial-temporal Top-k query results, no matter what attacking means introduced in the security model are adopted. We believe that accurate integrity verification is an effective way to force the adversaries dare not to destroy the integrity of the query results, because they do not want the curious or malicious storage nodes to be discovered.
- *The efficiency goal*: Our proposed scheme should be effective in communication and computation. It should greatly decrease the additional communication cost of both the sensor nodes and storage nodes, since those nodes, especially the sensor nodes, are energy limited. Here, the additional communication cost mainly refers to the cost of transmitting the proof data which are used by the network owner to verify the integrity of the query results.

4. Our Scheme STQ-TMWSN

In this section, we present our novel scheme STQ-TMWSN. As a whole, STQ-TMWSN consists of five parts: (1) Secret key predistribution; (2) virtual-location construction; (3) secure data preprocessing; (4) secure spatial-temporal Top-k query processing; (5) completeness verification of the query results. In the following subsections, we describe the five parts of STQ-TMWSN in great detail.

4.1. Secret Key Predistribution

STQ-TMWSN leverages two encryption methods: one is the order preservation encryption (OPE) encryption [26] and the other one is the pairwise-key-based encryption [25]. The former is used to encrypt the scores of the sensory data items, while the latter one is used to encrypt the sensory data items and the proof data, such as the target locations of the sensor nodes and the ranking orders of the sensory data items. Subsection 4.3 will show more detail about this.

As it is assumed in [12], we also assume each sensor node is pre-loaded with a distinct pairwise key shared with network owner. Moreover, some OPE-encryption materials [26] should also be pre-loaded in the tamper-proof hardware of each sensor node before being deployed. To achieve the forward security, pairwise keys should be updated periodically. Specifically, Key_i^{t+1} should be equal to $hash(Key_i^t)$, where $hash(*)$ is an one-way hash function.

4.2. Construction of the Virtual Grids

In STQ-TMWSN, the sensor deployment field is divided into many small virtual grids. In each virtual grid, all points approximately have the same location in real applications. Then, we design an ID-distribution law for the virtual grids. Based on the law, the real locations of each mobile sensor node can be worked out easily if the IDs of the virtual grids where it has moved to are known.

Specifically, the ID-distribution law is decribed as follows. Suppose the TMWSNs-deployed field is a $L * L$ square rectangle. STQ-TMWSN divides the rectangle into $\eta = (L/\zeta)^2$ small virtual grids, where ζ is a small digital number which can divide the length L with no remainder. Clearly, the smaller ζ is, the larger η is. Then, each virtual grid is given an ID which is a sequence number ranging from 1 to η. The virtual grids in the first row at the upper side of the rectangle are given the IDs 1, 2, 3, . . ., $L/\zeta - 1$, and L/ζ respectively from the left to the right in order; the IDs $L/\zeta + 1$, $L/\zeta + 2$, . . ., $2 * (L/\zeta) - 1$, and $2 * (L/\zeta)$ are assigned to those in the second row orderly;. . .; those in the last row have the IDs $\eta - L/\zeta + 1$, $\eta - L/\zeta + 2$, . . ., $\eta - 1$, and η respectively.

Using such an ID-distribution law, each sensor node first works out the IDs of the virtual grid where it has moved to, and then takes the IDs as the coordinate values of its target locations.

4.3. Secure Data Preprocessing

This subsection describes how each sensor node generates its data report, which will be uploaded to the corresponding storage node at the end of each epoch, based on its own sensory data items under the privacy-and-integrity preservation requirements. Specifically, for any sensor node $S_i (0 < i \leq N)$, the procedure of data report generation in STQ-TMWSN is shown in Algorithm 1.

In the protocol, S_i firstly computes the score of each sensory data item generated by itself based on the public scoring function; then, it works out $DPP^t_{i,j} (0 < j \leq \lambda^t_i)$ for each of its target locations which it has been moved to during epoch T^t. To do this, three cases are considered: $\mu^t_{i,j} = 0$, $\mu^t_{i,j} = 1$, and $\mu^t_{i,j} > 1$. If $\mu^t_{i,j} = 0$, $DPP^t_{i,j}$ should include $E_{Key^t_i}\{0, Loc^t_{i,j}\}$ to show that no sensory data was generated by S_i at $Loc^t_{i,j}$ in epoch T^t, where $E_{Key^t_i}\{*\}$ is a symmetric encrypting operation with Key^t_i based on [25]; if $\mu^t_{i,j} = 1$, $DPP^t_{i,j}$ should contain $E_{Key^t_i}\{1, Loc^t_{i,j}\}$ to indicate that only one sensory data item was generated by S_i at $Loc^t_{i,j}$ in epoch T^t, and it also needs to include both the pairwise-key-encrypted score and the OPE-encrypted score of the only data item. The former will be used as part of the proof information for integrity verification, and the latter will be used by storage nodes to process spatial-temporal Top-k query smoothly. The only sensory data item should also be encoded using the pairwise key and included in $DPP^t_{i,j}$. If $\mu^t_{i,j} > 1$, the contents of $DPP^t_{i,j}$ are a little complex. Specifically, it contains not only the OPE-encrypted scores and the pairwise-key-encrypted data items and scores, but also the chaining relationships of the ranked sensory data items. The chaining relationships, which are used to prevent the adversaries from destroying the integrity of the Top-k query results by dropping part of the qualified Top-k data items, are achieved by encrypting each sensory data item together with its ranking order number, which is called the sequence number in the following of this paper, using the pairwise key Key^t_i. Moreover, each sensory data item is bond together with its corresponding target location to further strengthen the integrity preservation of the Top-k query results. The final output $RT^t_{S_i}$ in Algorithm 1 is the very data report which will be uploaded to the corresponding storage node of S_i.

Algorithm∼1 Secure Data Preprocessing on $S_i (0 < i \leq N)$

```
Ensure: target location set $\{Loc^t_{i,1}, Loc^t_{i,2}, \cdots, Loc^t_{i,\lambda^t_i - 1}, Loc^t_{i,\lambda^t_i}\}$; all the sensory data
    items generated by $S_i$ in $T^t$; the symmetrc key $Key^t_i$; the encryption materials used
    for OPE;
```

```
Require: $RT^t_{S_i}$;
Compute the score of each sensory data item using the public scoring function;
for $j = 1$ to $\lambda^t_i$ do
    if $\mu^t_{i,j} = 0$ then
        set $DPP^t_{i,j}$ to $\left\{Loc^t_{i,j}, E_{Key^t_i}\{0, Loc^t_{i,j}\}\right\}$;
    end if
    if $\mu^t_{i,j} = 1$ then
        set $DPP^t_{i,j}$ to $\left\{Loc^t_{i,j}, E_{Key^t_i}\{1, Loc^t_{i,j}\}, E_{Key^t_i}\{d^t_{i,j,1}, Loc^t_{i,j}\},
        E_{OPE}\{d^t_{i,j,1}\}, E_{Key^t_i}\{D^t_{i,j,1}, Loc^t_{i,j}\}\right\}$;
    end if
    if $\mu^t_{i,j} > 1$ then
        Sort the sensory data items generated by $S_i$ at $Loc^t_{i,j}$ in $T^t$ according to their
        scores;
        set $DPP^t_{i,j}$ to $\left\{Loc^t_{i,j}, E_{Key^t_i}\{\mu^t_{i,j}, Loc^t_{i,j}\}, E_{Key^t_i}\{$
        $d^t_{i,j,1}, Loc^t_{i,j}\}, E_{OPE}\{d^t_{i,j,1}\}, E_{Key^t_i}\{1, D^t_{i,j,1},$
        $Loc^t_{i,j}\}, \cdots, E_{OPE}\{d^t_{i,j,\mu^t_{i,j}-1}\}, E_{Key^t_i}\{\mu^t_{i,j} - 1,$
        $D^t_{i,j,\mu^t_{i,j}-1}, Loc^t_{i,j}\}, E_{OPE}\{d^t_{i,j,\mu^t_{i,j}}\},$
        $E_{Key^t_i}\{D^t_{i,j,\mu^t_{i,j}}, Loc^t_{i,j}\}\right\}$;
    end if
end for
set $RT^t_{S_i}$ to $\left\{i, t, E_{key_{i,t}}\{Loc^t_{i,1}, Loc^t_{i,2} \ldots,$
$Loc^t_{i,\lambda^t_i-1}, Loc^t_{i,\lambda^t_i}\}, DPP^t_{i,1}, DPP^t_{i,2}, \ldots,$
$DPP^t_{i,\lambda^t_i-1}, DPP^t_{i,\lambda^t_i}\right\}$
return $RT^t_{S_i}$.
```

4.4. Secure Spatial-temporal Top-k Query Processing

This subsection presents how storage nodes process spatial-temporal Top-*k* query securely in our proposed scheme STQ-TMWSN. When a storage node receives a spatial-temporal Top-*k* query $Q_t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}$, it first determines whether it is located in cell I_C. If it is not located in cell I_C, it discards the query; otherwise, it processes every data report uploaded by the sensor nodes in cell I_C. At last, the storage node packets all the processing results of the data reports collected in the queried cell to form the final query result of the spatial-temporal Top-*k* query.

Algorithm 2 shows the procedure of processing a spatial-temporal Top-*k* query on the storage node in the queried cell. Specifically, Lines 1∼9 aim to find out the number of locations, which fall in QR_{I_C}, of each sensor node in cell I_C, and the corresponding $Data-proofPackets$ generated at those locations; from lines 12 to 42, there is a big $''for''$ loop, which is used to process every report generated in cell I_C in T^t. Line 14 shows the processing result of $RT^t_{S_i}$ considering the case that no target location of S_i falls in QR_{I_C} in T^t; Lines 16∼39 describe the procedure of processing $RT^t_{S_i}$ considering the case that there is at least one location of S_i falls in QR_{I_C} in T^t. In the above-mentioned latter case, all the $Data-proofPackets$ that correspond to the target locations located in QR_{I_C} are processed based on the exact values of μ^t_{i,x_j} and/or n^t_{i,x_j}, where μ^t_{i,x_j} and n^t_{i,x_j} denote the total data number and the qualified data number respectively corresponding to the location Loc^t_{i,x_j} which is supposed to be in the queried region QR_{I_C}. During the procedure of processing the $Data - proofPackets$, the OPE-encrypted items are all removed from the original $Data - proofPackets$ since the only use of them is to make storage nodes find out the qualified Top-k data items encrypted with the pairwise keys. Moreover, all the unqualified data items except for the one which follows the last qualified Top-k data item in each $Data - proofPacket$ are also removed from each original $Data - proofPacket$, and the reserved one will be used for completeness verification of the spatial-temporal Top-k query results.

Algorithm~2 Secure Spatial-temporal Top-k Query Processing on the storage node in Cell I_C

```
Ensure: $\{RT^t_{S_1}, RT^t_{S_2}, ..., RT^t_{S_{N-1}}, RT^t_{S_N}\}$; $Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}$;
Require: $R^t$;
for $i = 1$ to $N$ do
    $n[i] = 0$;
    for $j = 1$ to $\lambda^t_i$ do
        if $Loc^t_{i,j}$ is in $QR_{I_C}$ then
            put $DPP^t_{i,j}$ into set $\Theta$;
            $n[i] = n[i] + 1$;
        end if
    end for
end for
find out the pairwise-key-encrypted qualified Top-$k$ data items among all the pairwise-key-encrypted data items in set $\Theta$ according to their corresponding OPE-encrypted scores;
calculate $n^t_{i,j}$ for each $i \in [1, N]$ and $j \in [1, \lambda^t_i]$;
for $i = 1$ to $N$ do
    if $n[i] = 0$ then
        set $RST_{S_i}$ to $\left\{i, t, E_{key_{i,t}}\{Loc^t_{i,1}, Loc^t_{i,2} \ldots,\right.$
        $\left.Loc^t_{i,\lambda^t-1}, Loc^t_{i,\lambda^t}\}\right\}$
    else
        for $j = 1$ to $n[i]$ do
            if $\mu^t_{i,x_j} = 0$ then
                set $DPP^t_{i,x_j}$ to $\left\{E_{Key^t_i}\{0, Loc^t_{i,x_j}\}\right\}$;
            end if
            if $n^t_{i,x_j} = 0, \mu^t_{i,x_j} > 0$ then
                set $DPP^t_{i,x_j}$ to $\left\{E_{Key^t_i}\{d^t_{i,x_j,1}, Loc^t_{i,x_j}\}\right\}$;
            end if
            if $0 < n^t_{i,x_j} = \mu^t_{i,x_j} \leq k$ then
                if $n^t_{i,x_j} = 1$ then
                    set $DPP^t_{i,x_j}$ to $\left\{E_{Key^t_i}\{1, Loc^t_{i,x_j}\}, E_{Key^t_i}\{D^t_{i,x_j,1}, Loc^t_{i,x_j}\}\right\}$;
                end if
                if $n^t_{i,x_j} > 1$ then
                    set $DPP^t_{i,x_j}$ to $\left\{n^t_{i,x_j}, E_{Key^t_i}\{\mu^t_{i,x_j}, Loc^t_{i,x_j}\},\right.$
                        $E_{Key^t_i}\{1, D^t_{i,x_j,1}, Loc^t_{i,x_j}\}, ...,$
                        $E_{Key^t_i}\{\mu^t_{i,x_j} - 1, D^t_{i,x_j,\mu^t_{i,x_j}-1}, Loc^t_{i,x_j}\},$
                    $E_{Key_i^t}\{D^t_{i,x_j,\mu^t_{i,x_j}}, Loc^t_{i,x_j}\}\Big\};$
                end if
            end if
            if $0 < n^t_{i,x_j} \leq k, \mu^t_{i,x_j} > n^t_{i,x_j}$ then
                if $\mu^t_{i,x_j} = n_{x_i,t} + 1$ then
                    set $DPP^t_{i,x_j}$ to $\Big\{n^t_{i,x_j}, E_{Key_i^t}\{\mu^t_{i,x_j}, Loc^t_{i,x_j}\},$
                    $E_{Key_i^t}\{1, D^t_{i,x_j,1}, Loc^t_{i,x_j}\}, ...,$
                    $E_{Key_i^t}\{n^t_{i,x_j}, D^t_{i,x_j,n^t_{i,x_j}}, Loc^t_{i,x_j}\},$
                    $E_{Key_i^t}\{D^t_{i,x_j,\mu^t_{i,x_j}}, Loc^t_{i,x_j}\}\Big\};$
                end if
                if $\mu^t_{i,x_j} > n^t_{i,x_j} + 1$ then
                    set $DPP^t_{i,x_j}$ to $\Big\{n^t_{i,x_j}, E_{Key_i^t}\{\mu^t_{i,x_j}, Loc^t_{i,x_j}\},$
                    $E_{Key_i^t}\{1, D^t_{i,x_j,1}, Loc^t_{i,x_j}\}, ...,$
                    $E_{Key_i^t}\{n^t_{i,x_j}, D^t_{i,x_j,n^t_{i,x_j}}, Loc^t_{i,x_j}\},$
                    $E_{Key_i^t}\{n^t_{i,x_j} + 1, D^t_{i,x_j,n^t_{i,x_j}+1}, Loc^t_{i,x_j}\}\Big\};$
                end if
            end if
        end for
        set $RST^t_{S_i}$ to $\Big\{i, t, E_{Key_i^t}\{Loc^t_{i,1}, Loc^t_{i,2}, \ldots,$
            $Loc^t_{i,\lambda_i^t - 1}, Loc^t_{i,\lambda_i^t}\}, DPP^t_{i,x_1}, DPP^t_{i,x_2}, \ldots,$
            $DPP^t_{i,x_{n[i]-1}}, DPP^t_{i,x_{n[i]}}\Big\};$
    end if
end for
return set $\Big\{I_{Q^t}, RST^t_{S_1}, RST^t_{S_2}, ..., RST^t_{S_{N-1}}, RST^t_{S_N}\Big\}.$
```

4.5. Completeness Verification of the Query results

This subsection introduces how the completeness of the spatial-temporal Top-k query results are verified by the network owner. Specifically, we present the completeness verification strategy in Algorithm 3, the output of which is the value of the boolean variable $completeness$. In particular, if $completeness$ is $false$, R^t is verified as incomplete; otherwise, R^t is complete and the final R_{tpk}, which refers to the R_{tpk} in Algorithm 3 when Algorithm 3 returns $completeness$, is composed of all the qualified Top-k data items corresponding to the spatial-temporal Top-k query Q^t .

The main idea of Algorithm 3 to verify the completeness of R^t is to find out the minimal data score of the qualified Top-k data items and the maximal score of the unqualified ones generated in the queried region from R^t, and compare them with each other. Normally, the former one should be bigger than the latter one if the query aims to find out the biggest top k data items. If this condition does not hold in R^t, R^t is considered incomplete. However, it is not correct yet to declare that R^t has integrity even if such a condition holds in R^t. Before doing such a comparison, it is necessary to check whether each sensor report was processed properly by the compromised storage node (lines 2∼53 in Algorithm 3) based on the proof information included in R^t. To achieve this, each $Data - proofPacket$ in R^t should be checked. When checking the $Data - proofPackets$, three cases need to be considered, namely $\gamma^t_{i,x_j} = 0$ (lines 16∼25),

$\gamma^t_{i,x_j} = 1$ (lines 26∼32), and $\gamma^t_{i,x_j} > 1$ (lines 33∼ 51). If $\gamma^t_{i,x_j} = 0$, either S_i did not generate any data items at Loc^t_{i,x_j} in T^t or no data item generated by S_i at Loc^t_{i,x_j} in T^t is the qualified Top-k data item. Thus, in such a case, either $E_{Key^t_i}\{d^t_{i,x_j,1}, Loc^t_{i,x_j}\}$ or $E_{Key^t_i}\{0, Loc^t_{i,x_j}\}$ should be originally included in DPP^t_{i,x_j} in R^t. If $\gamma^t_{i,x_j} = 1$, the data item included in DPP^t_{i,x_j} should be a qualified Top-k data item according to lines 24∼26 in Algorithm 2. If $\gamma^t_{i,x_j} > 1$, according to lines 27∼38 in Algorithm 2, the storage node must have made some illegal query-processing operations if any of the following cases happens (lines 33∼35 in Algorithm 3): a) $n^t_{i,x_j}$is not included in DPP^t_{i,x_j} in R^t; b) no sensory data item in DPP^t_{i,x_j} is encrypted with a sequence number; c) the sequence numbers encrypted in DPP^t_{i,x_j} are not sorted in ascending order from 1; d) any sensory data item encrypted in DPP^t_{i,x_j} is not originally encrypted with Loc^t_{i,x_j}; e) $E_{Key^t_i}\{\mu^t_{i,x_j}, Loc^t_{i,x_j}\}$ is not originally included in DPP^t_{i,x_j}. Moreover, in the case that $\gamma^t_{i,x_j} > 1$, γ^t_{i,x_j} should be equal to either n^t_{i,x_j} or $n^t_{i,x_j} + 1$ according to lines 27∼38 in Algorithm 2 where n^t_{i,x_j} is included in R^t. Thus, in lines 36∼50 in Algorithm 3, the above-mentioned two cases are considered respectively to detect the integrity of R^t.

Algorithm∼3 Integrity Verification of the Query Result R^t on a network owner

```
Ensure: $R_t = \left\{I_{Q^t}, RST^t_{S_1}, RST^t_{S_2}, ..., RST^t_{S_{N-1}}, RST^t_{S_N}\right\}; Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}; \{Key^t_1, Key^t_2, .$
Require: $Completeness$.
 1: $R_{tpk} = \emptyset; V_{nonTop} = \emptyset; Completeness = true;$
 2: for $i = 1$ to $N$ do
 3:     if $(RST^t_{S_i} \notin R_t) || (RST^t_{S_i}$ contains no pairwise-key-encrypted target locations) then
 4:         set $Completeness = false$; return $Completeness$;
 5:     end if
 6:     decrypt all the ciphertext in $RST^t_{S_i}$ with $Key^t_i$;
 7:     if The network owner cannot decrypt the ciphertext normally then
 8:         $Completeness = false$; return $Completeness$;
 9:     end if
10:     calculate the value of $\Omega_i$ which is the total number of the queried locations in $RST^t_{S_i}$;
11:     for $j = 1$ to $\Omega_i$ do
12:         if $DPP^t_{i,x_j}$ is not originally in $RST^t_{S_i}$ ($DPP^t_{i,x_j}$ is a Data-proof Packet corresponding to $Loc^t_{i,x_j}$ which is in $QR_{I_C}$) then
13:             $Completeness = false$; return $Completeness$;
14:         end if
15:         calculate the value of $\gamma^t_{i,x_j}$ which is the total number of the sensory data items in $DPP^t_{i,x_j}$;
16:         if $\gamma^t_{i,x_j} = 0$ then
17:             if $E_{Key^t_i}\{d^t_{i,x_j,1}, Loc^t_{i,x_j}\}$ is originally in $DPP^t_{i,x_j}$ in $R^t$ then
18:                 $V_{nonTop} = V_{nonTop} \cup \{d^t_{i,x_j,1}\}$;
19:                 continue;
20:             else if $E_{Key^t_i}\{0, Loc^t_{i,x_j}\}$ is originally in $DPP^t_{i,x_j}$ in $R^t$ then
21:                 continue;
22:             else
23:                 $Completeness = false$; return $Completeness$;
```

```
            end if
        end if
        if $\gamma^t_{i,x_j} = 1$ then
            if $DPP^t_{i,x_j} \neq \left\{E_{Key^t_i}\{1, Loc^t_{i,x_j}\}, E_{Key^t_i}\{D^t_{i,x_j,1}, Loc^t_{i,x_j}\}\right\}$ then
                $Completeness = false$; return $Completeness$;
            end if
            $R_{tpk} = R_{tpk} \cup \{D^t_{i,x_j,1}\}$;
            continue;
        end if
        if ($n^t_{i,x_j}$ is not included in $DPP^t_{i,x_j}$ in $R^t$)||(no sensory data item in $DPP^t_{i,x_j}$ is encrypted with a sequence number)||(the sequence numbers encrypted in $DPP^t_{i,x_j}$ are not sorted in ascending order from 1)||(any sensory data item encrypted in $DPP^t_{i,x_j}$ is not originally encrypted with $Loc^t_{i,x_j}$)||($E_{Key^t_i}\{\mu^t_{i,x_j}, Loc^t_{i,x_j}\}$ is not originally included in $DPP^t_{i,x_j}$) then
            $Completeness = false$; return $Completeness$;
        end if
        if $n^t_{i,x_j} = \gamma^t_{i,x_j}$ then
            if $\gamma^t_{i,x_j} \neq \mu^t_{i,x_j}$ then
                $Completeness = false$; return $Completeness$;
            else
                $R_{tpk} = R_{tpk} \cup \{D^t_{i,x_j,1}, D^t_{i,x_j,2}, \ldots, D^t_{i,x_j,\gamma^t_{i,x_j}}\}$;
            end if
        else if $n^t_{i,x_j} = \gamma^t_{i,x_j} - 1$ then
            if ($E_{Key^t_i}\{D^t_{i,x_j,\mu^t_{i,x_j}}, Loc^t_{i,x_j}\}$ is included in $DPP^t_{i,x_j}$)&&($\gamma^t_{i,x_j} \neq \mu^t_{i,x_j}$) then
                $Completeness = false$; return $Completeness$;
            end if
            $R_{tpk} = R_{tpk} \cup \{D^t_{i,x_j,1}, D^t_{i,x_j,2}, \ldots, D^t_{i,x_j,n^t_{i,x_j}}\}$;
            $V_{nonTop} = V_{nonTop} \cup \{f(D^t_{i,x_j,\gamma^t_{i,x_j}})\}$;
        else
            $Completeness = false$; return $Completeness$;
        end if
    end for
end for
if $(V_{nonTop} = \emptyset) \| (SIZE(R_{tpk}) \neq k)$ then
    $Completeness = false$; return $Completeness$;
end if
if $f(MIN(R_{tpk})) < MAX(V_{nonTop})$ then
    $Completeness = false$; return $Completeness$;
end if
return $Completeness$.
```

5. Security Analysis

In this section, the performances of STQ-TMWSN on both privacy preservation and completeness verification are analyzed thoroughly in the form of theorems and proofs.

5.1. Analysis of STQ-TMWSN on Privacy Preservation

Theorem 1. *TMWSNs aided by our scheme STQ-TMWSN are able to preserve the privacy of both the sensory data items and their scores for spatial-temporal Top-k query under the attack model presented in Section III.*

Proof. According to Algorithm 1, before being uploaded to storage nodes, all sensory data items are encrypted with the pairwise keys and all the data scores are also encrypted using OPE [26] by the sensor nodes in TMWSNs. Meanwhile, according to the security model presented in this paper and the key pre-distribution method used in STQ-TMWSN, the sensor nodes and the STP which keep the pairwise keys and the OPE-encryption materials are trustworthy, the curious or malicious storage nodes are not able to obtain the keys or the OPE-encryption materials from them so that they cannot disclose the values of the sensory data items and their scores. Thus, Theorem 1 holds. □

5.2. Analysis of STQ-TMWSN on Integrity Verification

Theorem 2. *Suppose a queried node $S_i(\forall i \in [1, N])$ generated $\mu^t_{i,j}(\mu^t_{i,j} > 0)$ data items at a queried location $Loc^t_{i,j}(\forall j \in [1, \lambda^t_i])$ in epoch T^t, where there are $n^t_{i,j}(0 < n^t_{i,j} \leq k)$ qualified Top-k data items. If at least one of those qualified Top-k data items was deleted from $DPP^t_{i,j}$ by the storage node when producing $RST^t_{S_i}$ of R^t which is the query result of the spatial-temporal Top-k query $Q^t = \{I_{Q^t}, T^t, k, I_C, QR_{I_C}\}$, under the security model described in Section III, the incomplete R^t must be detected by any network owner with a 100% successful rate in TMWSNs aided by our scheme STQ-TMWSN.*

Proof. Since the storage node does not know Key^t_i, if it inserts the sensory data items which are encrypted with some other keys rather than Key^t_i into $DPP^t_{i,j}(\forall i \in [1, N], \forall j \in [1, \lambda^t_i])$, the incomplete R^t must be detected by the network owner according to lines 6 $\sim$ 9 in Algorithm 3. Moreover, according to lines 33 $\sim$ 35 in Algorithm 3, R^t must be also considered as incomplete if the storage node puts any encrypted data item, which was generated by S_i in T^t at some other location rather than $Loc^t_{i,j}$, into $DPP^t_{i,j}$. Thus, in the following of this proof, we need only to consider the situation that all the encrypted sensory data items left in $DPP^t_{i,j}$ after being processed by the storage node are the real ones which were generated by $S_i(\forall i \in [1, N])$ at $Loc^t_{i,j}$ in T^t (but some or all of them may not be the qualified ones). Then, if at least one qualified sensory data items generated by S_i at $Loc^t_{i,j}$ in T^t is discarded by the storage node, one of the following two cases must appear: 1) the storage node has deleted all the sensory data items from $DPP^t_{i,j}$ when producing $RST^t_{S_i}$ of R^t; 2) the storage node only discarded part of the sensory data items from $DPP^t_{i,j}$, and the discarded data items contain some qualified one/ones.

First of all, consider the case that the storage node has deleted all the sensory data items from $DPP^t_{i,j}$. In this case, the storage node should leave $E_{Key^t_i}\{d^t_{i,j,1}, Loc^t_{i,j}\}$ in $DPP^t_{i,j}$ in $RST^t_{S_i}$ of R^t to avoid being detected according to lines 16 $\sim$ 25 in Algorithm 3, because it cannot generate the legal encryption item $E_{Key^t_i}\{0, Loc^t_{i,j}\}$. Then, $d^t_{i,j,1}$ should be put into V_{nonTop} according to lines 17 $\sim$ 18 in Algorithm 3, and some real but unqualified sensory data items generated in QR_{I_C} and T^t must be put into R_{tpk} to make the number of the elements in R_{tpk} equal to k according to lines 53 $\sim$ 55 in Algorithm 3. If the discarded sensory data items contain some qualified one/ones, $d^t_{i,j,1}$ must be the score of a qualified Top-k data item. Then, $f(MIN(R_{tpk}))$ must be smaller than $MAX(V_{nonTop})$ because the score of any qualified Top-k data item must be bigger than that of any real but unqualified one generated in QR_{I_C} and T^t consuming all data scores are distinct. Thus, according to lines 56 $\sim$ 58 in Algorithm 3, the incomplete R^t must be detected by the network owner.

doi:10.20944/preprints202102.0398.v1

Then, consider the case that the storage node deletes part of the sensory data items from $DPP^t_{i,j}$, and the deleted data items contain some qualified one/ones. In this case, two situations should be discussed. One is that all the sensory data items encrypted with sequence order numbers are deleted from $DPP^t_{i,j}$, while the other one is that at least one sensory data item encrypted with a sequence number is left in $DPP^t_{i,j}$ after being processed. In the first situation, $E_{Key^t_i}\{D^t_{i,j,\mu^t_{i,j}}, Loc^t_{i,j}\}$ must be left in $DPP^t_{i,j}$ after being processed, and there must be $DPP^t_{i,x_j} \neq \left\{E_{Key^t_i}\{1, Loc^t_{i,j}\}, E_{Key^t_i}\{D^t_{i,j,1}, Loc^t_{i,j}\}\right\}$ since $\mu^t_{i,j} \neq 1$ in this situation and $E_{Key^t_i}\{1, Loc^t_{i,j}\}$ must not be included in DPP^t_{i,x_j}. According to lines $26 \sim 29$ in Algorithm 3, the incomplete R^t must be detected by the network owner. Then, consider the second situation. To make the sequence numbers encrypted with the sensory data items in $DPP^t_{i,j}$ in $RST^t_{S_i}$ of R^t ascends from 1 orderly (Lines $33 \sim 35$ in Algorithm 3), the storage node must delete all the sensory data items in one of the sets $\Phi_1, \Phi_2, \Phi_3, \Phi_4$, and Φ_5 from $DPP^t_{i,j}$. The five sets are shown in Eq.(2), where $1 < w < \mu^t_{i,j} - 1$.

$$\begin{cases} \Phi_1 = \{D^t_{i,j,w}, D^t_{i,j,w+1}, ..., D^t_{i,j,\mu^t_{i,j}-1}\} \\ \Phi_2 = \{D^t_{i,j,\mu^t_{i,j}-1}\} \\ \Phi_3 = \{D^t_{i,j,w}, D^t_{i,j,w+1}, ..., D^t_{i,j,\mu^t_{i,j}}\} \\ \Phi_4 = \{D^t_{i,j,\mu^t_{i,j}-1}, D^t_{i,j,\mu^t_{i,j}}\} \\ \Phi_5 = \{D^t_{i,j,\mu^t_{i,j}}\} \end{cases} \tag{2}$$

If the storage node discards the sensory data items/item in set Φ_1 or Φ_2 from $DPP^t_{i,j}$ when processing $DPP^t_{i,j}$, $E_{Key^t_i}\{D^t_{i,j,\mu^t_{i,j}}, Loc^t_{i,j}\}$ and $E_{Key^t_i}\{1, D^t_{i,j,1}, Loc^t_{i,j}\}$ must be left in $DPP^t_{i,j}$ after being processed, which means that $\gamma^t_{i,j}$ is bigger than 1. According to lines $36 \sim 50$ in Algorithm 3, the storage node has to either set $n^t_{i,j}$ to $\gamma^t_{i,j}$ or $\gamma^t_{i,j} - 1$ in $DPP^t_{i,j}$ in $RST^t_{S_i}$ of R^t to prevent the incomplete R^t from being detected. Even though, the incomplete R^t must be also detected by the network owner according to lines $36 \sim 38$ and $42 \sim 45$ in Algorithm 3, because $\gamma^t_{i,j}$ must not be equal to $\mu^t_{i,j}$ in this case and $E_{Key^t_i}\{D^t_{i,x_j,\mu^t_{i,j}}, Loc^t_{i,j}\}$ is included in $DPP^t_{i,j}$ at the same time.

If the storage node deletes the sensory data items/item in set Φ_3, Φ_4 or Φ_5 from $DPP^t_{i,j}$ when processing $DPP^t_{i,j}$, the encryption item $E_{Key^t_i}\{\gamma^t_{i,j}, D^t_{i,j,\gamma^t_{i,j}}, Loc^t_{i,j}\}$ should be left in $DPP^t_{i,j}$ after being processed. Then, if $\gamma^t_{i,j} = 1$, the incomplete R^t must be detected by the network owner according to lines $26 \sim 29$; if $\gamma^t_{i,j} > 1$, since $\gamma^t_{i,j} \neq \mu^t_{i,j}$ in this case, the storage node has to set $n^t_{i,j}$ to $\gamma^t_{i,j} - 1$ in $DPP^t_{i,j}$ in $RST^t_{S_i}$ of R^t to make the incomplete R^t free from being detected according to lines $36 \sim 50$ in Algorithm 3. Then, $f(D^t_{i,j,\gamma^t_{i,j}})$ will be put into set V_{nonTop} according to lines $42 \sim 47$ in Algorithm 3. Because some dropped sensory data item/items is/are qualified Top-k data item/items, $D^t_{i,j,\gamma^t_{i,j}}$ must also be a qualified Top-k data item. Since the number of the sensory data items in R_{tpk} should be k, some real but unqualified Top-k data items whose scores are smaller than $f(D^t_{i,j,\gamma^t_{i,j}})$ must be put into set R_{tpk}. Thus, there must be $f(MIN(R_{tpk})) < MAX(V_{nonTop})$, and the incomplete R^t must be detected by the network owner according to lines $56 \sim 58$ in Algorithm 3.

Thus, if the storage node deletes at least one qualified sensory data items from $DPP^t_{i,j}$, the network owner in TMWSNs aided by STQ-TMWSN is able to detect the incomplete R^t with a successful rate of 100%, and Theorem 2 holds. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202102.0398.v1

Theorem 3. *Under the security model described in Section III, any network owner in TMWSNs aided by our scheme STQ-TMWSN can detect the incomplete spatial-temporal Top-k query results with a 100% successful rate.*

Proof. According to the security model described in Section III, the curious or malicious storage node cannot produce fabricated and pairwise-key-encrypted sensory data items which cannot be detected by the network owner because it is not pre-loaded with the legal pairwise keys. Thus, for any spatial-temporal Top-k query Q^t, if its query result R^t is incomplete, at least one qualified sensory data item must be discarded by the storage node when producing R^t. In other words, there must be at least one queried sensor node $S_i(\forall i \in [1, N])$ whose corresponding $Data - proofPacket\ DPP^t_{i,j}$ at location $Loc^t_{i,j}(\forall j \in [1, \lambda^t_i])$ satisfies the following condition: at least one qualified sensory data item was deleted from $DPP^t_{i,j}$ by the storage node when producing $RST^t_{S_i}$ of R^t. Then, according to Theorem 2, the incomplete R^t must be detected by the network owner in TMWSNs aided by STQ-TMWSN. Thus, Theorem 3 holds.
□

6. Computation Complexity Analysis

This section analyzes the computation complexity of the three schemes presented above.

Firstly, the computation complexity of Algorithm 1 is analyzed as follows. Since most of the statements in Algorithm 1 are the loop body of the "for" loop statements in Algorithm 1, the computation complexity of Algorithm 1 should be that the loop numbers multiply the computation complexity of the loop body. In the loop body, there are only three conditional statements. Thus, the computation complexity of the loop body depends on the pairwise-key encryption methods used in STQ-TMWSN and the total length of the data that need to be encrypted as well as the computation complexity of OPE. Although different pairwise-key cryptography methods, such as [25] and [27], may have different computation complexities, they are considered lightweight generally and fit for the resource-limited sensor nodes [28,29], let alone the storage nodes which are much more powerful than the sensor nodes. Moreover, OPE also has low computation complexity according to [26]. For each $DPP^t_{i,j}(0 < i \leq N, 0 < j \leq \mu^t_{i,j})$, the length of the data that need to be encrypted varies according to $\mu^t_{i,j}$, which symbolizes the total number of the sensed-data items generated by S_i at $Loc^t_{i,j}$ in T^t. Let l_D and l_d denote the bit length of a sensed-data item and that of a data score respectively, l_n symbolize not only the bit length of a sequence number but also that of $\mu^t_{i,j}$, l_{Loc} refers to the bit length of a virtual location, and $l^{OPE}_{i,j}$ and $l^{PW}_{i,j}$ denote the bit length of the data that need to be encrypted using OPE and that of those encoded adopting the pairwise-key encryption method respectively in $DPP^t_{i,j}$. Then, the values of $l^{OPE}_{i,j}$ and $l^{PW}_{i,j}$ can be worked out using equations (3) and (4) respectively according to Algorithm 1.

$$l^{OPE}_{i,j} = \begin{cases} 0 & \text{if } \mu^t_{i,j} = 0, \\ l_d & \text{if } \mu^t_{i,j} = 1, \\ \mu^t_{i,j} \times l_d & \text{if } \mu^t_{i,j} \geq 2 \end{cases} \tag{3}$$

$$l^{PW}_{i,j} = \begin{cases} l_n + l_{Loc} & \text{if } \mu^t_{i,j} = 0, \\ l_n + l_d + l_D + 3l_{Loc} & \text{if } \mu^t_{i,j} = 1, \\ (l_n + l_D + l_{Loc})\mu^t_{i,j} + l_d + 2l_{Loc} & \text{if } \mu^t_{i,j} \geq 2 \end{cases} \tag{4}$$

Secondly, pay attention to Algorithm 2. The computation complexity of lines $1 \sim 9$ is O($\sum_{i=1}^{N} \lambda^t_i$); the computation complexity of line 10 depends on the adopted sorting algorithm and the total number of sensory data items generated in T^t and QR_{I_C}; that of line 11 is O($\sum_{i=1}^{N} \lambda_i^t$); that of lines 12 $\sim$ 43 in Algorithm 2 is O(N) in the best case (e.g., $n[i]$ is always 0 for each $i \in [1, N]$), and is O($\sum_{i=1}^{N} n[i]$) in the worst case (e.g., $n[i]$ is not equal to 0 for each $i \in [1, N]$).

Finally, it is the turn of Algorithm 3, which mainly consists of one outer "for" loop whose loop body contains an inner "for" loop. In the loop body of the outer loop, the computation complexity of line 6 is the highest among all the statements which are in the loop body of the outer loop and out of the inner loop. If decrypting one encryption item $E_{Key_i^t}\{*\}$ is taken as one operation, the operation number of line 6 should be $n[i] + 1$ according to line 40 in Algorithm 2. Then, the computation complexity of Algorithm 3 should be O($\sum_{i=1}^{N}(n[i] + 1 + \Omega_i)$).

7. Performance Evaluation

In this section, we evaluate the performances of our proposed scheme STQ-TMWSN through extensive simulations taking OMNET++ as the simulation tool.

7.1. Metrics and Experimental Setup

Since the sensory data items always need to be transmitted to storage nodes no matter what kind of methods are used to ensure the security of spatial-temporal Top-k query, the performance of STQ-TMWSN on energy efficiency is evaluated mainly by testing the additional communication cost. Specifically, the metrics used in our simulations are listed as follows.

- Additional communication cost in a cell (C_{cell}): Total energy consumed by transmitting all the proof data produced in a cell and an epoch to the storage node in the cell;
- Proof-data ratio (R_{vs}): The ratio of C_{cell} to $C_{reports}$. Here, $C_{reports}$ refers to the total energy consumed by transmitting all the reports generated in a cell and an epoch to the storage node in the cell, where the data reports include both the sensory data items and the proof data generated by all the sensor nodes in the cell and the epoch;

The parameters used in our simulation and their own default values are shown in Table 2, where the default values of some parameters are set by referencing [12]. In fact, static sensor nodes are also allowed to be existed in TMWSNs. In the simulation, we adjust the ratio of the mobile sensor nodes to the total ones in the systems by changing the value of r_{mobile}.

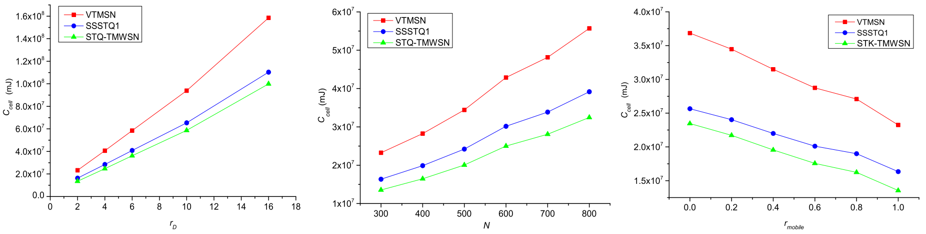
Table 2. Parameters and Their Default Values.

Parameters	Default value
N	300
T (Length of each epoch)	100 s
T_{mobile} (Period for a sensor node to keep moving)	5 s
T_{static} (Period for a sensor node to keep static)	5 s
m_{speed} (Moving speed of each mobile sensor node)	5 m/s
r_{mobile} (Ratio of the mobile sensor nodes to the total ones)	100 %
C_{size} (Cell size)	400×400 m^2
R (Sensor communication radius)	50 m
r_D (Data generation rate of each sensor node)	2 items/s
q_{period} (Period for the network owner to launch a query)	5 s
q_{radius} (Radius of the queried region which is a circle)	50 m
l_D (Length of a sensory data item)	400 bits
l_d (Length of a data score)	20 bits
l_n (Length of a sequence number)	10 bits
l_{id} (Length of an ID number)	10 bits
l_t (Length of a time data)	32 bits
l_{Loc} (Length of each two-dimensional location)	128 bits
l_{VLoc} (Length of each target location)	16 bits
e_{send} (Cost of sending one bit data)	1 mJ
$e_{receive}$ (Cost of receiving one bit data)	1 mJ

7.2. Simulation Results

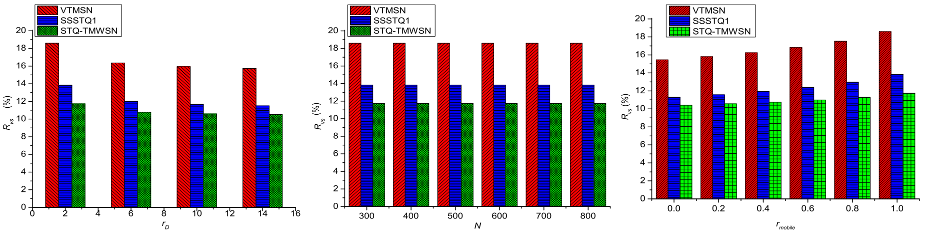
This subsection presents the simulation results of C_{cell} and R_{vs} with different settings of r_D, N and r_{mobile} respectively. We compare our scheme with VTMSN [20] and SSSTQ1[5] in this section. VTMSN, which was proposed in 2015, is the earliest work on securing spatial-temporal Top-k query in TMWSNs, while SSSTQ1 can be considered as the state-of-the-art scheme proposed for securing spatial-temporal Top-k query in TMWSNs. Figure 3 shows the simulation results of C_{cell} under different settings of r_D, N and r_{mobile}, and Figure 4 illustrates the simulation results of R_{vs} with different settings of r_D, N and r_{mobile} respectively. From Figure 3, we can see that the C_{cell} lines of STQ-TMWSN are all lower than those of VTMSN and SSSTQ1. This indicates that our proposed scheme STQ-TMWSN is more energy-efficient than the other two schemes. The C_{cell} lines in Figures 3 (a) and 3 (b) are on a upward trend because the quantity of sensory data items rises as r_D or N becomes larger and larger which causes the increase of the proof data, while those in Figure 3 (c) are on a downward trend as r_{mobile} rises from 0 to 1 because the sensor nodes are assumed to generate sensory data items only when they are static or arrive at their target locations and the quantity of the sensory data items and the corresponding proof data must decrease when more sensor nodes are set to be mobile.

Thanks to the technology of virtual-location construction proposed in this paper, fewer bits of location information are included in the proof data in STQ-TMWSN than the other two schemes, which decreases the ratio of the proof data to the whole data including both sensory data items and their proof. From Figure 4, we can see that the values of R_{vs} of STQ-TMWSN are all under 12% which is within the acceptable range in real applications and also lower than those of the other two schemes.



(a) (b) (c)

Figure 3. C_{cell} with different settings of r_D (a), N (b) and r_{mobile} (c)



(a) (b) (c)

Figure 4. R_{vs} with different settings of r_D (a), N (b) and r_{mobile} (c)

8. Conclusions

Secure spatial-temporal Top-k query has been well-studied in TWSNs in the past two decades. However, such a problem is less-studied in TMWSNs, where mobile sensor nodes exist. This paper presents a novel scheme named STQ-TMWSN to ensure secure processing of spatial-temporal Top-k query in TMWSNs. Thorough security analysis shows that STQ-TMWSN is able to preserve the privacy of both the sensory data items and the data scores. Meanwhile, we have proved in this paper that TMWSNs aided by STQ-TMWSN can detect the incomplete query-processing results of the spatial-temporal Top-k queries with a 100% successful rate under the attack model presented in this paper. Moreover, extensive simulation experiments are conducted to evaluate the performances of STQ-TMWSN. The simulation results demonstrate that STQ-TMWSN is much more efficient than the related state-of-the-art schemes, and this is very important for the resource-limited TMWSNs.

Funding: This work was supported by the Natural Science Foundation of China (No. 61972090), the Key Specialized Research and Development Project in Henan Province (No. 202102210161), and Planning Subject for the 13th Five Year Plan of National Education Sciences (No.2019GXJK272).

Acknowledgments: The authors would like to thank the reviewers for their valuable comments and suggestions

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, X.; Ma, Z.Z.J.L.Y.Z.L.Z.N. An Effective Edge-Assisted Data Collection Approach for Critical Events in the SDWSN-Based Agricultural Internet of Things. *Electronics* **2020**, *9*, 907.
2. Liu, A.; Liu, X.; Long, J. A Trust-Based Adaptive Probability Marking and Storage Traceback Scheme for WSNs. *Sensors* **2016**, *16*. doi:10.3390/s16040451.
3. Sun, Y.; Rehfeldt, D.; Brazil, M.; Thomas, D.; Halgamuge, S. A Physarum-Inspired Algorithm for Minimum-Cost Relay Node Placement in Wireless Sensor Networks. *IEEE/ACM Transactions on Networking* **2020**, *28*, 681–694. doi:10.1109/TNET.2020.2971770.
4. Zhang, Q.; Xie, C.; Jiang, L. A Novel Tree-Based Genetic Algorithm for the Multicast Protocol in Two-Tiered WSNs. *Chinese Journal of Electronics* **2020**, *29*, 852–858. doi:10.1049/cje.2020.07.007.
5. Ma, X.P.; Liang, J.B.; Wang, J.X.; Wen, S.; Wang, T.; Li, Y.; Ma, W.P.; Qi, C.D. Secure fine-grained spatio-temporal Top-k queries in TMWSNs. *Future Generation Computer Systems* **2018**, *86*, 174–184.

6. Negi, D.; Ray, S.; Lu, R. Pystin: Enabling Secure LBS in Smart Cities With Privacy-Preserving Top-k Spatial–Textual Query. *IEEE Internet of Things Journal* **2019**, *6*, 7788–7799. doi:10.1109/JIOT.2019.2902483.
7. Ding, X.; Liu, P.; Jin, H. Privacy-Preserving Multi-Keyword Top-*k* k Similarity Search Over Encrypted Data. *IEEE Transactions on Dependable and Secure Computing* **2019**, *16*, 344–357. doi:10.1109/TDSC.2017.2693969.
8. Kui, X.; Feng, J.; Zhou, X.; Du, H.; Ma, X. Securing top-k query processing in two-tiered sensor networks. *Connection Science* **2020**, pp. 1–19.
9. Li, R.; Liu, A.X.; Xiao, S.; Xu, H.; Bruhadeshwar, B.; Wang, A.L. Privacy and Integrity Preserving Top- *k* Query Processing for Two-Tiered Sensor Networks. *IEEE/ACM Transactions on Networking* **2017**, *25*, 2334–2346. doi:10.1109/TNET.2017.2693364.
10. Su, S.; Teng, Y.; Cheng, X.; Xiao, K.; Li, G.; Chen, J. Privacy-Preserving Top-k Spatial Keyword Queries in Untrusted Cloud Environments. *IEEE Transactions on Services Computing* **2018**, *11*, 796–809. doi:10.1109/TSC.2015.2481900.
11. Quan, H.; Wang, B.; Zhang, Y.; Wu, G. Efficient and Secure Top-k Queries With Top Order-Preserving Encryption. *IEEE Access* **2018**, *6*, 31525–31540. doi:10.1109/ACCESS.2018.2847307.
12. Zhang, R.; Shi, J.; Liu, Y.; Zhang, Y. Verifiable Fine-Grained Top-k Queries in Tiered Sensor Networks. 2010 Proceedings IEEE INFOCOM, 2010, pp. 1–9. doi:10.1109/INFCOM.2010.5461927.
13. Liao, X.; Li, J.; Lei, Y. Secure and efficient Top-k query processing in two-tiered sensor network. *Journal of Computer Research and Development* **2013**, *50*, 490–497.
14. He, R.; Dai, H.; Yang, G.; Wang, T.; Bao, J. An efficient Top-k query processing with result integrity verification in two-tiered wireless sensor networks. *Mathematical Problems in Engineering* **2015**, *2015*, 1–8. doi:10.1155/2015/538482.
15. Hua, D.; Geng, Y.; Fu, X.; Qiang, Z. EVTQ: An Efficient Verifiable Top-k Query Processing in Two-Tiered Wireless Sensor Networks. 2013 IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks, 2013, pp. 206–211. doi:10.1109/MSN.2013.26.
16. Liang, J.; Jiang, C.; Ma, X.; Wang, G.; Kui, X. Secure Data Aggregation for Top-k Queries in Tiered Wireless Sensor Networks. *Adhoc & Sensor Wireless Networks* **2016**, *32*, 51–78.
17. Li, R.; Lin, Y.; Yi, Y.; Xiong, S.; Ye, S. Security Top-k query protocol in two layer sensor networks. *Journal of Computer Research and Development* **2012**, *49*, 1947–1958.
18. Yu, C.M.; Tsou, Y.T.; Lu, C.S.; Kuo, S.Y. Practical and Secure Multidimensional Query Framework in Tiered Sensor Networks. *IEEE Transactions on Information Forensics and Security* **2011**, *6*, 241–255. doi:10.1109/TIFS.2011.2109384.
19. Chen, W.; Yu, L.; Gao, D. A Privacy Preserving Histogram Aggregation Algorithm with Integrity Verification Support. *Chinese Journal of Electronics* **2014**, *42*, 2268–2272.
20. Liu, F.; Ma, X.; Liang, J.; Lin, M.; Zhang, R.; Shi, J.; Liu, Y.; Zhang, Y. Verifiable Top-k Query Processing in Tiered Mobile Sensor Networks. *International Journal of Distributed Sensor Networks* **2015**, *11*, 437678. doi:https://doi.org/10.1155/2015/437678.
21. HaiqinWuandLiangminWang. Efficient and Secure Top-k Query Processing on Hybrid Sensed Data. *Mobile Information Systems* **2016**, *2016*, 1–10. doi:http://dx.doi.org/10.1155/2016/1685054.
22. M, X.; L, X.; Junbin L, e.a. A Comparative Study on Two Typical Schemes for Securing Spatial-Temporal Top-k Queries in Two-Tiered Mobile Wireless Sensor Networks. *Sensors* **2018**, *18*, 871. doi:https://doi.org/10.3390/s18030871.
23. Agrawal, R.; Kiernan, J.; Srikant, R.; Xu, Y. Order Preserving Encryption for Numeric Data. Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data; ACM: New York, NY, USA, 2004; SIGMOD '04, pp. 563–574. doi:10.1145/1007568.1007632.
24. Das, G.; Gunopulos, D.; Koudas, N.; Tsirogiannis, D. Answering Top-k Queries Using Views. Proceedings of the 32Nd International Conference on Very Large Data Bases. VLDB Endowment, 2006, VLDB '06, pp. 451–462.
25. Nagaraju, B.; Ramkumar, P. A New Method for Symmetric Key Cryptography. *International Journal of Computer Applications* **2016,**, *142*, 36–39.
26. Khoury, E.; Medlej, M.; Jaoude, C.A.; Guyeux, C. Novel order preserving encryption scheme for wireless sensor networks. 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), 2018, pp. 1–6. doi:10.1109/MENACOMM.2018.8371028.
27. Verma, S.; Choubey, R.; Soni, R.; Ogi, P. An Efficient Developed New Symmetric Key Cryptography Algorithm for Information Security. *International Journal of Emerging Technology and Advanced Engineering* **2012**, *2*, 18–21.
28. Roy, S.; Karjee, J.; Rawat, U.; N., D.P.; Dey, N. Symmetric Key Encryption Technique: A Cellular Automata based Approach in Wireless Sensor Networks. *Procedia Computer Science* **2016**, *78*, 408 – 414. 1st International Conference on Information Security & Privacy 2015, doi:https://doi.org/10.1016/j.procs.2016.02.082.
29. la Krishna.; Doja, M.N. Deterministic K-means secure coverage clustering with periodic authentication for wireless sensor networks. *International Journal of Communication Systems* **2017**, *30*, 1–16. doi:10.1002/dac.3024.