SECURE AND PRIVACY PRESERVING DATA CENTRIC SENSOR NETWORKS WITH MULTI-QUERY OPTIMIZATION

R Tanuja¹, S H Manjula², K R Venugopal³, L M Patnaik⁴
¹Assistant Professor, Dept. of Computer Science and Engg, UVCE, Bangalore, India
²Associate Professor, Dept. of Computer Science and Engg, UVCE, Bangalore, India
³Professor, Dept. of Computer Science and Engg, UVCE, Bangalore, India
⁴Honorary Professor, Indian Institute of Science, Bangalore, India

Abstract
In large-scale Wireless Sensor Networks the effective use of the vast amounts of data gathered will require scalable, self-organizing, and energy-efficient data dissemination and data retrieval techniques. Data Centric Sensor (DCS) networks is a better approach in which the sensor nodes send the sensed data to a designated sensor node whose name is associated with sensed data. However, due to unattended nature of Wireless Sensor Networks, these sensor nodes are susceptible to different types of attacks. The attacker may compromise these storage nodes and get data of his interest. In this paper we propose a Secure and Privacy Preserving Data Centric Sensor Networks that includes security and privacy support to DCS networks. We use multi level key structure and cryptographic algorithms to provide security. In addition, we propose a multi-query optimization technique that aggregates similar queries for a small periodic time and construct a query message. This reduces the number of messages required to serve multiple similar queries. Simulation and experimental results show that our work provides a secure privacy preserving data centric sensor network based on cryptographic keys and reduces the message overhead and incurs a minimum communication cost compared to previous works.

Keywords: Data Centric Sensor Networks, Privacy preserving, Query Optimization, Security, Steiner tree, Wireless Sensor Networks

1. INTRODUCTION
Wireless Sensor Network consists of hundreds to thousands of small sensor nodes that have limited computational, storage and communication capability. Wireless Sensor Nodes are severely limited in their capabilities and resources. For e.g., a typical sensor [1] operates at a frequency of 2.4 GHz, has a data rate of 250Kbps, 128 KB of flash memory, 512 KB of memory for measurements and communication range of 30 to 100m. Unlike traditional networks, sensor nodes are often deployed in accessible areas, which have the added risk of physical attack. These are unattended in nature that sense and store/forward the information about their surroundings.

WSNs are found to be extremely useful for a large area of applications. Sensors can be deployed for habitat monitoring, environmental monitoring including forest fire detection, air pollution and green house monitoring. They are extremely useful in military applications that demand high security for nuclear and chemical attack detection, battlefield surveillance. It has guaranteed its use in health care for patient monitoring and many other commercial applications [2]. Because sensor networks pose unique challenges, security techniques used in traditional networks cannot be applied directly.

The resources and information communicated in WSNs should be protected and as well as defend security attacks. Most important security measures that should be addressed are: (i) Data Confidentiality: where information can be interpreted only by the intended recipient and no one else. (ii) Data Integrity: ensures that information is not modified during communication. (iii) Data Availability: deals with the network services available even in presence of internal or external attacks like denial of service (DoS). The main goal is successful delivery of packets in network at all times. (iv) Data freshness: implies the data is new and ensures that no replay attacks from adversary are taking place. (v) Self-Organization: deals with a great challenging task of self organization and self healing in dynamic WSNs. Using public-key cryptographic techniques, one can ensure the safety of nodes operation. (iv) Authentication: ensures the communicating nodes are genius and are whom they claim to be. This is very important issue as many attacks use the identification of a real node to break-through the network and access the sensitive data readings.

With a large-scale sensor network the amount of data gathered is very high and the fact that these data are widely spread across the network demands an efficient data dissemination and access technique to extract the interested data from the network. One of the key challenges in wireless sensor networks is the storage and querying of useful sensor data, broadly referred to as data management. The term useful sensor data is application specific and has different...
meaning in different application scenarios. For eg., in a target localization application, users are interested in detecting vehicles and tracking their movement. Here, useful sensor data comprises time and location of the target objects as well as their tracks. The different factors that depend on how data can be retrieved are:
1) Where the data stored in a network? Is it stored locally at each sensor node (local storage), in a distributed manner within the network (distributed storage), or at the edge of the network (centralized storage)?
2) How are queries routed to stored data? Can we use attributes of the search to make it efficient?
3) How can the network deal with storage limitations of individual sensor nodes?

Many schemes have been proposed for data management in sensor networks. Accordingly Data Centric Sensor (DCS) network [3] model has a better approach in which data are named based on the event type and sensed data are passed to the corresponding named sensor node using Geographic Table (GHT) [4]. The advantage of this is that the queries for data with a specific event can be sent directly to the storage nodes rather than flooding the query through the entire network.

1.1 Motivation
For security reasons, the data sensed and transmitted by the sensor nodes must reach the destination and the information must be available to legitimate users and not to unprivileged users or attackers. In addition, the data storage and retrieval techniques should be energy efficient. The DCS [5] storage systems support efficient data storage and retrieval techniques as the data is stored only in storage nodes rather than the entire network. In a sensor network that is based on external base station storage the sensed data by all the nodes is forwarded to the base station. The energy in the nodes nearer to the base station gets depleted soon and they die. These networks lead to a single point of failure and the base station is most vulnerable to various types of attacks. In such a network scenario, DCS provides a better data storage and retrieval techniques. However, previous DCS does not support security and privacy. Securing sensor networks faces different challenges because these are small wireless sensor nodes deployed in very large numbers that are unattended in nature and lack tamper resistance. Due to this they are more prone to adversary attacks that need security.

1.2 Contributions
We present a Secure and Privacy Preserving Data Centric Sensor (SPDCS) network which provides security to the DCS network with an improved query optimization technique. The contributions of the paper are as follows:
(i) The sensed data by the sensor nodes/detection cells are encrypted and forwarded to storage cells. Even if an attacker compromises the detection cell he cannot obtain any previous data.
(ii) If an attacker attacks a storage cell, he cannot decrypt any data stored as they are encrypted with different detection cell keys.
(iii) It provides an efficient query optimization technique which combines different user queries of same event type through a gateway node that reduces the number of messages. Gateway uses a different set of user keys for secure communication of queries and replies.

1.3 Organization
The remaining part of the paper is organized as follows: Related works are presented in section II. Section III presents the background of the work. Preliminaries discussing the network model and adversary model are presented in section IV. Section V describes problem definition and algorithm. Implementation and performance evaluation is shown in section VI. Finally, section VII gives the conclusions of the paper.

2. RELATED WORKS
Recently, many researchers have been engaged in developing schemes that address the unique challenges in security of WSNs. Zia and Zomaya [6] have made an effort to document the security issues in WSNs. Attacks, countermeasures and threat models have been proposed in different layers. Key distribution for the sensor network is extensively studied. Generally two approaches are given.
One is key distribution at the initial network deployment stage through a Base station and the other is key pre-deployment. Zhu et al., [7] have proposed a key management protocol, Localized Encryption and Authentication Protocol (LEAP) that offers many security benefits to WSNs. It employs one base station and assumes it to be trustworthy. It does not include defense against hacked or compromised base station. Delan Alsoufi et al., [8] have proposed a solution in order to overcome the security issue with one base station by employing multiple base stations.

One of the main concerns when designing a key management scheme is the network scalability. Yu et al., [9] have proposed a new scalable key management scheme for WSNs, which provides good secure connectivity coverage. For this purpose, they make use of the unital design theory. They show that the basic mapping from unitals to key pre-distribution allows one to achieve high network scalability. Nonetheless, this naive mapping does not guarantee a high key sharing probability. Walid Bechkit et al., [10] propose an enhanced unital-based key pre-distribution scheme providing high network scalability and good key sharing probability. Chin-Yin Chow et al., [11] discuss privacy preserving location monitoring system where information gathered about personal location faces threats. The concept of aggregate location information is proposed from which individual identities are removed. Their algorithm is based on $k$ anonymity privacy concept. The resource-aware algorithm aims to minimize the communication cost and the quality aware algorithm aims to maximize the accuracy of the aggregate locations.

Recent works are being focused on query processing in sensor networks. Niki Trigoni et al., [12] consider multi-query optimization for aggregate queries on sensor networks. They develop a set of distributed algorithms for processing multiple queries that incur minimum communication considering the limitations of sensor nodes. They use linear reduction technique for processing multiple queries that incur minimum communication and show through analysis that it results in significant energy savings. Shili Xiang et al., [13] design a light-weight but effective scheme to support multiple data acquisition and aggregation queries in a wireless sensor network, in order to minimize the number of radio transmissions. As the base station is the interface of a wireless sensor network, they use the base station as a filter to reduce duplicate data accesses from the sensor network, and as a screen to hide the query dynamics as much as possible. They design a two-tier optimization scheme, base station optimization and in-network optimization.

In this paper we consider multi-query optimization for aggregate queries on sensor networks. We develop a set of distributed algorithms for processing multiple queries that incur minimum communication while observing the computational limitations of the sensor nodes. Our algorithms support incremental changes to the set of active queries and allow for local repairs to routes in response to node failures. A thorough experimental analysis shows that our approach results in significant energy savings, compared to the previous works.

3. BACKGROUND

Min Shao et al.,[14] have discussed about security and privacy for DCS networks with different levels of data privacy based on encryption. They use Euclidean Steiner Tree (EST) and Keyed Bloom Filter (KBF) for query optimization to minimize the query overhead. They use five types of symmetric keys for different mapping schemes to store the event E occurred. The proposed mapping schemes are capable of answering queries of different granularity and can achieve different levels of privacy. In the EST scheme, the query path is encoded as a Steiner tree. Each node id is presented by two bytes, thus only 12 cell ids can be encoded in each packet. In the KBF scheme, 25 bytes are used to encode the query path with a Bloom Filter, and it is expected to achieve an acceptable false positive rate.

Our work present security and privacy support for the DCS network with different levels of cryptographic keys for various purpose. Tanuja R et al.,[15] have discussed about security for DCS networks with different levels of data privacy based on encryption. Here a node compromised is revoked with different level of keys. We use the trajectory based routing in which the trajectory will be explicitly appended in the data packet using EST. In addition, we propose an efficient multi-query optimization technique based on equivalence class that reduces the number of queries sent to the network with maintaining the same level of security compared to the previous work.

4. PRELIMINARIES

This section presents our network model, assumptions made about the network and the anticipated adversary. Table 1 summarizes notations used in the algorithm.

4.1 Network Model

Similar to previous works on DCS systems [5] our algorithm assumes that the region under analysis is represented in the form of cells. Each cell contains one or more nodes. A cell may be a Detection cell or a Storage cell. Detection cells are those that sense data about an event type. Storage cells are the cells where the sensed data are stored. Each cell has a cell ID and each sensor node knows in which cell it is located through GPS. A gateway or a mobile sink is a node with special capabilities, which acts as an interface between user and the sensor network. Queries are injected to the network through gateway. Gateway is responsible for authenticating different user groups and for aggregating queries. In the proposed scheme we use three types of cryptographic keys.

- **Cell Key**: Each cell is assigned a key which is used for storing sensed data in encrypted form in the designated storage cell.
- **User Key**: Each user group is assigned a user key for secure communication between the user and gateway.
5.1 Objectives
(i) Security and Privacy for WSN: Security is provided through different levels of keys namely cell keys, user keys and master keys for encryption and authentication. Privacy is ensured by periodically changing the storage location for data of an event type. Storage cell to detection cell mapping is done in a secure way through hash function.
(ii) Query optimization through multi-query optimization technique: In multi-query optimization technique, queries are buffered for epoch time and aggregated based on event types. The aggregated query is injected to the sensor network. This helps in achieving the following:
• Result sharing among multiple users: If multiple users query about the same event then the result from the network is fetched at once and distributed to all users who does same query.
• Minimization of message overhead: As a result of query aggregation, number of queries injected into the network is minimized and number of hops traversed are reduced as compared to individual queries.
(iii) Compromised node revocation: Based on the reports of the forwarding nodes, malicious nodes are revoked by distributing new set of keys

5.2 Algorithm
The algorithm consists of three phases:
(i) Event detection phase: In this phase the nodes in the detection cell detect the event and collect the data of that event. It also detect the storage cell by the hash function. These data are then encrypted using its cell key of detection cell. The encrypted data are forwarded to the storage cell of the particular event.
(ii) Query forwarding phase: In this phase the user queries are sent to the gateway and forwarded. The user query is encrypted using the user key shared between the gateway and the user. At the gateway the queries are processed to form an aggregated query. This query is forwarded to each storage nodes.
A packet is destined to avoid traffic analysis compared with e. If e does not contain f t, a query packet is delivered to generate Ui.

This continues until the query packet reaches all the storage and IDs of its sibling node and forward it to its children.

The data packet is then forwarded to the Storage cell Si. The storage cell Si stores the data packet locally. In this phase Multi Node Routing technique is used as Anti Traffic analysis method where the data packet generated by a detection cell is forwarded to cells other than the destination storage cell. This is done in order to avoid traffic analysis performed by an attacker to know the location of sensed data.

### Table 2: SDCS: Event Detection Phase

<table>
<thead>
<tr>
<th>Begin</th>
<th>initialize set of cells as detection cells and storage cells for every detection cell Di</th>
</tr>
</thead>
<tbody>
<tr>
<td>if Di senses data about event Ei then</td>
<td></td>
</tr>
<tr>
<td>S = H(Ei)</td>
<td></td>
</tr>
<tr>
<td>dp = C(CT, Ei)</td>
<td></td>
</tr>
<tr>
<td>forward dp to Si which stores it locally</td>
<td></td>
</tr>
<tr>
<td>endfor</td>
<td></td>
</tr>
<tr>
<td>end</td>
<td></td>
</tr>
</tbody>
</table>

2) Query Forwarding phase: Let U = {U1, U2, U3, ..., Un} be n users which appear in epoch time t. Let e be a generalised event type initialised to NULL. Let Ui generate a query and encrypt it using the user key KGU and forward the query to the gateway. We can use one of the feasible encryption algorithms for sensor networks [15]. At the gateway, the query is decrypted. The event type of decrypted query is compared with e. If e does not contain event type of query then it is combined. After this the query packet is pushed into a recipient buffer. This is required to store information about the user who generated the query for that event. After an epoch time of t, a query packet is generated which contains the generalized event type e. Euclidean Steiner Tree (EST) is now constructed. The EST is similar to a minimum spanning tree, in addition there are steiner cells along with the storage cells. These steiner cells can also help to improve query privacy because they add noise to the set of storage cells. The gateway is at the root of steiner tree which contains the IDs of cells in the EST. The IDs are added to the destination field of the query packet and sent to the child cells. On receiving query packet, the child cell reconstructs an EST sub-tree by removing its ID and IDs of its sibling node and forward it to its children. This continues until the query packet reaches all the storage cells.

### Table 3: SDCS: Query Forwarding Phase

<table>
<thead>
<tr>
<th>Begin</th>
</tr>
</thead>
<tbody>
<tr>
<td>initialize e = NULL</td>
</tr>
<tr>
<td>for every user Ui</td>
</tr>
<tr>
<td>if Ui generates query qp then</td>
</tr>
<tr>
<td>Cqp = E(qp, Kgu)</td>
</tr>
<tr>
<td>forward Cqp to gateway</td>
</tr>
<tr>
<td>endif</td>
</tr>
<tr>
<td>endfor</td>
</tr>
<tr>
<td>for every query Cqp</td>
</tr>
<tr>
<td>Pqp = D(Cqp, Kgu)</td>
</tr>
<tr>
<td>compare eqp with e</td>
</tr>
<tr>
<td>if (eqp! = e) then</td>
</tr>
<tr>
<td>e = eqp</td>
</tr>
<tr>
<td>endif</td>
</tr>
<tr>
<td>push Pqp into recipient buffer, RB</td>
</tr>
<tr>
<td>endif</td>
</tr>
<tr>
<td>generate query packet with e as payload</td>
</tr>
<tr>
<td>construct EST and forward query packet</td>
</tr>
<tr>
<td>End</td>
</tr>
</tbody>
</table>

3) Result Forwarding Phase: A storage cell Si receives a query packet. If storage cell contains data of the event specified in the query packet then the data packet is forwarded to the gateway. At the gateway the data packet is decrypted and the information of the users who generated the query for the data packet is fetched from the recipient buffer. Now the data packet is encrypted using the user key and forwarded to the user. The user decrypts the data using the user key.

### Table 4: SDCS: Result Forwarding Phase

<table>
<thead>
<tr>
<th>Begin</th>
</tr>
</thead>
<tbody>
<tr>
<td>// processing at gateway</td>
</tr>
<tr>
<td>for every dp received</td>
</tr>
<tr>
<td>for every packet in RB</td>
</tr>
<tr>
<td>if (dp(e) == RB(e)) then</td>
</tr>
<tr>
<td>P = D(dp, Ci)</td>
</tr>
<tr>
<td>Cre = E(P, Kgu)</td>
</tr>
<tr>
<td>reply Cre to user having user key Kgu</td>
</tr>
<tr>
<td>discard the packet in Recipient Buffer</td>
</tr>
<tr>
<td>endif</td>
</tr>
<tr>
<td>endfor</td>
</tr>
<tr>
<td>end</td>
</tr>
</tbody>
</table>

### 5.3. Node Revocation

Here any compromised or malicious node’s behavior is reported from the sensor nodes of the network to the gateway or mobile sink. Whenever a node within a cell for eg., a cell in the second row and third column C(2,3) of the WSN wants to report the misbehavior of another node in the same cell to gateway, it uses its master key to send report and a MAC based on the report to mobile sink. When multiple nodes send such a report about a malicious node then the gateway node distribute a new cell key to C(2,3) that is encrypted by the remaining individual nodes master key of that cell.
6. IMPLEMENTATION AND PERFORMANCE
EVALUATION

6.1 Simulation Setup

We evaluate the performance of our scheme by simulation and compare it with earlier schemes. The simulation was run using MATLAB. Nodes are randomly deployed into 100×100m2. The region is divided into 10×10 cell units. Nodes are randomly distributed and ensured that each cell contains at least one node. User keys are generated prior to sensor network deployment. Cell keys are generated after network deployment. A Gateway node or mobile sink with special properties is deployed within the network. Events are randomly simulated in later stages. We assume two distinct user groups with different information access privileges. At any point in time m users among a group can send queries to gateway. We use two buffers at the gateway one for storing incoming queries until an epoch time t and another for storing the same queries in order to determine the sender who generated the query. We assume that the number of users generating queries during an epoch is always less than sum of sizes of the buffers.

6.2 Results and Analysis

We consider following four metrics:
• Number of query messages: The number of query messages is total number of queries actually injected to the sensor network by the gateway.
• Message Overhead: is the number of hops traversed by all messages in the network.
• Total Query Delay: is the sum of delays experienced by a query message within the network while reaching the destination cell.
• Communication cost: Communication cost includes cost of query and result messages.

We consider a 128 bytes with a 18 bytes header and 110 bytes payload. If the payload contains x bits of information then the communication cost is calculated as follows: Communication cost = (x/110)*128. We consider fixed size headers and fixed size packets. The result packets for each query is considered to be of 64 bytes.

In pDCS [14] with a single-query EST scheme, for every packet that arrives a EST is constructed each time and query packet is generated and forwarded along the nodes of EST. In SPDCS with multi-query EST the packets are buffered until an epoch time t and later aggregated and sent as a single packet. Therefore the number of query messages generated in multi-query EST remains constant while it increases linearly with increase in number of messages in single-query EST scheme. For example, if t=8ms and assuming a query packet arrives every 1ms at gateway, then for 8ms 8 query packets would arrive and number of query messages generated = 8 in case of single-query EST. But in case of multi-query EST number of query messages generated will be 1 as 8 query packets are buffered until 8ms before generating the query message.

In Fig. 2 we compare message overhead for pDCS using single query with our work. In Single-query EST since a query message is generated for each query packet, the number of hops that the message traverses to cover all storage cells increases simultaneously with increase in number of storage cells. Whereas with Multi-query EST since a query message is generated for certain number of query packets, the number of hops traversed is significantly less and therefore multi-query EST out performs Single-Query EST. For example, if no of storage cells = 4, number of steiner cells in EST = 2, number of query packets = 5 in an epoch of 8ms, then in single-query EST each query message traverses 6 hops to reach all storage cells. Therefore total no. of hops by 5 query messages = 30.

With multi-query EST since query packets are aggregated for an epoch, total no. of hops traversed by 5 query messages = 6. This accounts for a significant reduction in network resources usage.

In Fig. 3 we consider the query delay. In Single-Query EST the total query delay is low when compared to Multi-Query
EST since in multi-query the query packets are buffered for an epoch of time prior to sending to the sensor network. The average query delay is considerably less in Multi-Query EST when compared to Single-Query EST.

6.3 Communication Cost Analysis

We compare performance of SDCS with previous methods using an analytical model based on approximated communication costs. We consider a sensor network with n nodes distributed among cells. Let Etotal represent total number of events detected in a sensor network. Let Q denote number of event types for which queries have been issued and Eq denote number of events detected for event types queried. We assume an asymptotic notation of O(√n) as cost for node-node routing within the sensor network. We compare costs based on approximations for both query packets arriving at gateway and number of packets routed within the sensor network.

Secure and Privacy Preserving Data Centric Storage : SPDCS : Total cost: Q\sqrt{n} + Eq\sqrt{n} + Etotal\sqrt{n}.
Data Centric Storage : pDCS Total cost: Q\sqrt{n} + Eq\sqrt{n} + Etotal\sqrt{n}.

It is evident from the above evaluations that when Q=1, Data-Centric Storage approach used in pDCS is equivalent to our Secure Data-Centric Storage approach. But when there are more queries the cost of previous works increases. Therefore we conclude that Secure Data-Centric Storage SDCS is preferable when large number of queries are involved and maximum arrival time of a query for an event type within an epoch of t does not exceed the time interval between consecutive detections for the same event type.

7. CONCLUSION

In wireless sensor networks security and privacy support with minimum cost is a major concern. In this paper we propose a technique that supports security for DCS networks with a better key management scheme that uses two levels of keys to avoid different types of attacks and a multi-query optimization technique using Keyed Bloom filter that reduces the number of query messages by aggregating similar queries at the gateway and reduce number of replies within the network. Simulation results show that our algorithm will support security for the data stored in the DCS networks and can significantly reduce the message overhead for a set of aggregate queries without losing any query privacy. In future, we will address different types of attacks such as Denial of Service and flooding attacks. We can further study in terms of query optimization to reduce communication overhead with a tradeoff of accuracy for any recently made query that can be processed within the gateway itself.

REFERENCES

BIOGRAPHIES

Tanuja R is currently Assistant Professor, Dept. of Computer Science, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. Her research interests are in the field of Wireless Sensor Networks, Cryptography and Network security.

S H Manjula is currently Associate Professor, Dept. of Computer Science, UVCE, Bangalore University, Bangalore. She was awarded Ph.D. in Computer Science from Dr.MGR University, Chennai. Her research interests are in the field of Wireless Sensor Networks and Data mining.

K R Venugopal is currently the Principal, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has over 410 research publications in refereed Intl. Journals and Conference Proceedings.

L M Patnaik is currently Honorary Professor, Indian Institute of Science, Bangalore where he was a Professor since 1986 with the Department of Computer Science and Automation. During the past 35 years of his service at the Institute he has over 500 research publications in refereed Intl. Journals and Conference Proceedings.