Secure Data Aggregation in Wireless Sensor Networks: A Survey

Patel Swapnil[1]
[1]Computer Engineering Department, Merchant Engineering College
Gujarat Technological University, Ahmedabad, India
patelswapnil79@gmail.com

Abstract: ***Recent advances in wireless sensor networks (WSNs) have led to many new applications including habitat monitoring and target tracking. Sensor nodes spend most of their energy during data transmission. With data aggregation, one can eliminate the redundant data transmission and so reduce the energy consumption. Sensor nodes are deployed in a very remote or hostile environment. They are often compromise. So security issues such as confidentiality, integrity, authentication and availability becomes crucial. There is currently research potential in securing data aggregation in the WSN. In this paper, we have presented existing secure data aggregation methods with advantages and disadvantages.***

Keywords: security, aggregation, wireless sensor network, sensor node, energy consumption

1. Introduction

A Wireless Sensor Network (WSN) usually contains thousands or hundreds of sensors which are randomly deployed. Wireless sensor nodes are made up of small electronic devices which are capable of sensing, computing and transmitting data from harsh physical environments like a battle field. These sensor nodes mostly depend on batteries for energy, which get depleted at a faster rate because of the computation and communication operations they have to perform.

A. Wireless Sensor Network Model

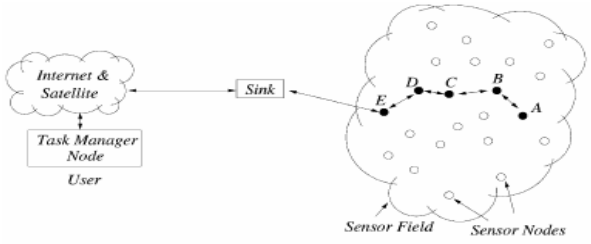

Figure 1: Wireless Sensor Network Communication Structure

Wireless sensor network is differing from other wireless ad-hoc network in the sense that they are resource limited, they are prone to failures, and they are deployed densely. The number of nodes in wireless sensor network is several orders higher than that of ad hoc networks. In wireless sensor network, network topology is constantly changing. They use a broadcast communication medium. The major components of a typical sensor network are: sensor nodes, the sensor field, the sink and the task manager. The area in which we expect a particular phenomenon to occur is called a sensor field i.e. the area in which the nodes are placed. Sensor nodes are the heart of the network. Sensor node collects data and routing this information back to a sink. Sinks are also known as data aggregation points. They receive process and store data from the other sensor nodes. Such points are usually assigned dynamically by the network. They reduce the overall energy requirements of the network by reducing the total number of messages to be sent.

The task manager node or base station is serves as gateway to other networks. The base station is a centralized point of control within the network. It extracts information from the network and disseminates control information back into the network. It serves as a powerful data processing or storage centre and an access point for a human interface. Hardware wise base station is either a laptop or a workstation.

B. Motivation

Sensor nodes have limited battery lifetime. The goal is to increase the battery lifetime. Most of the battery power is used to transmit and receive packets. A reduction in the number of transmissions and receptions will increase the available bandwidth. The number of communications can be reduced using aggregate queries. These are queries where the result is computed based on the data received from every node. For example, if a query is asking for the average temperature in a region, there is no need for the base station to receive all of the values. A node can get the values from its neighbors, compute the average and send the average to the upper nodes. There are several applications in which data aggregation can be employed. But mostly, in sensor databases, data aggregation is a crucial technique for performing aggregate queries.

2. Requirement for Data Aggregation Security

The data security requirements in the WSNs are similar to those in traditional networks. Following security properties are required to strengthen the security in aggregation:

(1) Data Confidentiality

It ensures that information content is never revealed to anyone who is not authorized to receive it. It can be divided into a hop by- hop basis and an end-to-end basis. In the hop by-hop basis, any aggregator point needs to decrypt the received encrypted data, apply some sort of

aggregation function, encrypt the aggregated data, and send it to the upper aggregator point. This kind of confidentiality implementation is not practical for the WSN since it requires extra computation. In an end to end basis, the aggregator does not need to decrypt and encrypt data. It needs to apply the aggregation functions directly on the encrypted data.

(2) Data Integrity

It ensures that the content of a message has not been altered during transmission process. An adversary near the aggregator point will be able to change the aggregated result sent to the base station by adding some fragments or manipulating the packet's content without detection.

(3) Data Freshness

It ensures that the data are recent and that no old messages have been replayed to protect data aggregation schemes against replay attacks. A passive adversary is able to listen to even encrypted messages transmitted between sensor nodes can replay them later on and disrupt the data aggregation results.

(4) Data Availability

It ensures that the network is alive and that data are accessible. Once an attacker gets into the WSN by compromising a node, the attack will affect the network services and data availability especially in those parts of the network where the attack has been launched. Moreover, the data aggregation security requirements should be carefully implemented to avoid extra energy consumption. If no more energy is left, the data will no longer be available.

(5) Authentication

There are two types of authentication; entity authentication, and data authentication. Entity authentication allows the receiver to verify if the message is sent by the claimed sender or not. Data authentication guarantees that the reported data is the same as the original one.

(6) Non-repudiation

In secure aggregation schemes, once the aggregator sends the aggregation results, it should not be able to deny sending them. This gives the base station the opportunity to determine what causes the changes in the aggregation results.

3. Types of Attacks on WSN Aggregation

WSNs are vulnerable to different types of attacks due to the nature of the transmission medium (broadcast), remote and hostile deployment location, and the lack of physical security in each node. Major attacks that might affect the aggregation in the WSN are as follows:

(1) Denial of Service Attack (DoS)

It is a standard attack on the WSN by transmitting radio signals that interfere with the radio frequencies used by the WSN and is sometimes called jamming. As the adversary capability increases, it can affect larger portions of the network.

(2) Node Compromise

It is where the adversary is able to reach any deployed sensor and extract the information stored on it which is sometimes called supervision attack. Considering the data aggregation scenario, once a node has been taken over, all the secret information stored on it can be extracted.

(3) Sybil Attack

In this attack, the attacker is able to present more than one identity within the network. It affects aggregation schemes in different ways. Firstly, an adversary may create multiple identities to generate additional votes in the aggregator election phase and select a malicious node to be the aggregator. Secondly, the aggregated result may be affected if the adversary is able to generate multiple entries with different readings. However, an adversary can launch a Sybil attack and generate n or more witness identities to make the base station accept the aggregation results.

(4) Selective Forwarding Attack

A compromised node forwards some of the packets while drops the remaining. In the aggregation context, any compromised intermediate nodes have the ability to launch the selective forwarding attack and this subsequently affects the aggregation results.

(5) Replay Attack

In this case an attacker records some traffic from the network without even understanding its content and replays them later on to mislead the aggregator and consequently the aggregation results will be affected.

(6) Stealthy Attack

The adversary aims to inject false data into the network without revealing its existence. In a data aggregation scenario, the injected false data value leads to a false aggregation result.

4. Related Work

In this section we have discuss existing method that are used for secure data aggregation.

(1) Secure Hop By Hop Data Aggregation Protocol (SDAP)

How can the base station obtain a good approximation of the fusion result when a fraction of sensor nodes are compromised? To answer this question, a secure hop-by-hop data aggregation protocol for sensor networks (SDAP) is proposed in [1]. The design of SDAP is based on the principles of divide-and-conquer and commit-and- attest. By using divide-and conquer, they partition the aggregation tree into groups to reduce the importance of high-level nodes in the aggregation tree. The design of SDAP is motivated by the following observation.

During a normal hop-by-hop aggregation process in a tree topology, we need to place more trust on the high-level nodes (i.e., nodes closer to the root) than the low-level nodes, because the aggregated result calculated by a high-level node is due to a larger number of sensor nodes. In other words, if a compromised node is closer to the root, the bogus aggregated data from it will have a larger impact on the final result computed by the root.

However, in reality none of these low-cost sensors should be more trustable than others. As such, SDAP takes the approach of reducing the trust on high-level nodes, which is realized by the principle of divide-and-conquer. More specifically, by using a probabilistic grouping method, SDAP dynamically partitions the topology tree into multiple logical groups (sub trees) of similar sizes. Since fewer nodes will be under a high level node in a logical sub tree, the potential security threat by a compromised high-level node is reduced. To preserve the efficiency of per-hop aggregation, SDAP performs hop-by-hop aggregation in each logical group and generates one aggregate from each group. In addition, based on the principle of commit-and-attest, SDAP enhances an ordinary hop-by-hop aggregation protocol with commitment capability, which ensures that once a group commits its aggregate this group cannot deny it later. After the BS has collected all the group aggregates, it then identifies the suspicious groups based on a bivariate multiple-outlier detection algorithm. Finally, each group under suspect participates in an attestation process to prove the correctness of its group aggregate. The BS will discard the individual group aggregate if a group under attestation fails to support its earlier commitment made in the collection phase; the final aggregate is calculated over all the group aggregates that are either normal or have passed the attestation procedure using commit-and-attest, the BS has a way to verify the aggregates. SDAP is a general-purpose secure aggregation protocol applicable to multiple aggregation functions. Simulation results show that the protocol is efficient with much less communication overhead.

(2) Secure Hierarchical In Network Aggregation

The overall algorithm presented in [2] has three main phases: query dissemination, aggregation-commit, and result- checking.

(a) Query dissemination: The base station broadcasts the query to the network. An aggregation tree has the base station at the root.

(b) Aggregation commits: In this phase, the sensor nodes iteratively construct a commitment structure resembling a hash tree. First, the leaf nodes in the aggregation tree send their data values to their parents in the aggregation tree. Each internal sensor node in the aggregation tree performs an aggregation operation whenever it has heard from all its child sensor nodes. Whenever a sensor node s performs an aggregation operation, s creates a commitment to the set of inputs used to compute the aggregate by computing a hash over all the inputs. Both the aggregation result and the commitment are then passed on to the parent of s. After the final commitment values are reported to the base station, the adversary cannot subsequently claim a different aggregation structure or result.

(c) Result checking: The result-checking phase is a novel distributed verification process. Once the querier has received the final commitment values, it disseminates them to the rest of the network in an authenticated broadcast. Each sensor node is responsible for checking that its own contribution was added into the aggregate. If a sensor node determines that its data value was indeed added towards the final sum, it sends an authentication code up the aggregation tree towards to the base station. Authentication codes are aggregated along the way with the XOR function for communication efficiency. When the querier has received the XOR of all the authentication codes, it can then verify that all the sensor nodes have confirmed that the aggregation structure is consistent with their data values. If so, then it accepts the aggregation result.

(3) Secure Data Collection Using Mobile Sink

Secure data collection scheme in wireless sensor network with mobile sink is presented in [3]. The proposed scheme divides the sink's data collection path into grids, sensors in each grid, uses secret keying information and collision-resistant hash functions to authenticate the source of beacons. It allows a sensor to authenticate the data request message to ensure reliable data collection. Simulations and analysis results show proposed scheme is robust against severe wireless sensor network attacks, such as wormhole attacks and HELLO flood attacks. The author has consider the case in which mobile sink follows a predetermined path to gather sensor data. The case in which mobile sink follows a random path to gather sensor data is not discuss.

(4) Reliable Private Data Aggregation Scheme(REBIVE)

Proposed approach in [4] is cluster-based approach that undergoes the steps; query launches & clusters formation, data aggregation at sensor nodes, and post-aggregation at query server.

(a) Query Launch & Clusters Formation: The query server

triggers the formation of clusters by launching a query. Thereafter all the sensor nodes collaborate to form clusters within the network. The server launches query with a HELLO message. Upon the reception of the HELLO message, each sensor decides to be a cluster head. If a node becomes a cluster head, it forwards the HELLO message to its neighbors. Otherwise, it just waits for a predefined time span and finally joins one of the clusters by sending a JOIN message. Thus multiple clusters are formed in the network.

(b) Data Aggregation within Clusters: Now the nodes within a cluster collaborates with each other to perform data aggregation. Here data aggregation in WSNs with reliability and privately is a major concern. Every sensor's individual data (raw data) should be known only to itself. Even the intermediate aggregated data should be kept private from other sensors in the network. Disclosure of any such data can leak private information. In order to preserving privacy, the links between sensors must be kept secret to prevent the outsiders from eavesdropping.

(c) Post-Aggregation at Query Server - Now it's time for the cluster heads to pass on their intermediate aggregated values to the query server. The cluster heads form an aggregation tree rooted at the query server. When the server receives all these intermediary sums, he just adds up the values. Thus the query server is reported with the sum of raw data of sensors while preserving each individual sensor's privacy.

(5) Attack Resilient Hierarchical Data Aggregation

Algorithms for resilient hierarchical data aggregation despite the presence of compromised nodes in the aggregation hierarchy are presented in [5]. The author showed that a compromised node can launch several simple attacks on the existing aggregation algorithms, which could significantly deviate the estimate of the aggregate. They also proposed modifications to the aggregation algorithms that guard against attacks. Analysis and simulation results show that proposed approach is scalable and efficient.

(6)Energy Efficient and Secure Pattern Based Aggregation

In order to perform data aggregation, generally the data transmitted by the sensor nodes should be decrypted at the cluster-head. The aggregated data is then encrypted before being transmitted to the base station. This technique is vulnerable from security perspective because decryption of data requires the cluster-head to obtain the symmetric key. In ESPDA as proposed in [6], since cluster-head does not decrypt the data the protocol is more secure. By implementing ESPDA this intermediate process is eliminated, which reduces the overhead of the cluster-heads and thus contributing to energy efficiency. The data aggregation is done before the actual data is transmitted by the sensor nodes. The sensor nodes have a unique secret built in key. The base station, periodically broadcasts a session key to maintain data freshness. The sensor node computes a node-specific-secret- key (NSSK) using the session key and the built in key. This NSSK is used to encrypt and decrypt all the consequent data transmission during that session. The base station has the knowledge about all the unique built in keys of the sensor nodes, which is used to compute NSSK at the base station for decryption.

(7) Trust Based Framework

The trust based approach for secure data aggregation in wireless sensor network was proposed in [7].The reputation of each individual sensor node is evaluated by using an information theoretic concept, Kullback-Leibler (KL) distance, to identify the compromised nodes through an unsupervised learning algorithm. With the help of the powerful Josang's belief model, the uncertainty existing in the sensory data and aggregation results is explicitly represented and quantified. Compared with the conventional schemes that are based on cryptography schemes, the proposed framework can effectively block the false data in the presence of multiple compromised nodes that would bypass outlier detection. Simulation results demonstrate that the trust based framework provides a powerful mechanism for detecting compromised nodes and reasoning about the uncertainty in the network.

5. Conclusions

Security is very crucial while aggregating the data in wireless sensor network because sensor nodes are deployed in remote and harsh environment. We have discussed many existing methods for secure data aggregation based on our literature review.

References

[1] Yi Yang, Xinran Wang, Sencun Zhu, and Guohong Cao "SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks" Journal ACM Transactions on Information and System Security, Volume 11, Issue 4, July 2008, pp. 356-367J.

[2] Haowen Chan, Adrian Perrig, Dawn Song "Secure Hierarchical in Network aggregation in Sensor Networks" Proceedings of the 13th ACM conference on Computer and communications security CCS'06, October 30–November 3, 2006, Alexandria, Virginia, USA, ACM, pp.1-10

[3] Amar Rasheed, Rabi Mahapatra "Secure Data Collection Scheme in Wireless Sensor Network with Mobile Sink" Proceedings of the 2008, Seventh IEEE International Symposium on Network Computing and Applications, ISBN: 978-0-7695-3192-2, pp 332-336

[4] Farzana Farzana Rahman, Md. Endadul Hoque, Sheikh Iqbal Ahamed "Preserving privacy in wireless sensor networks using reliable data aggregation" ACM SIGAPP Applied Computing Review, Volume 11 Issue 3, August 2011, pp. 52-62

[5] Sankardas Roy, Sanjeev Setia, Sushil Jajodia "Attack resilient hierarchical data aggregation in sensor networks" Proceedings of the fourth ACM

workshop on Security of ad hoc and sensor networks, NY, USA, 2006, ISBN: 1-59593-554-1, pp. 71-82

[6] H Cam, S Ozdemir, P Nair, D Muthuavinashiappan "ESPDA: Energy efficient and secure pattern based data aggregation for wireless sensor networks" Computer Communications (2006), Volume 29, Issue 4, ISBN: 0780385217, pp. 446-455

[7] Wei Zhang, Sajal K. Das, and Yonghe Liu "A trust based framework for secure Data aggregation in wireless sensor networks" 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks 2006, pp. 60-69.

Author Profile

Patel Swapnil received the B.E. degree in Computer Science & Engineering from Sakalchand Patel College of Engineering Visnagar in 2011 is presently a PG student in Computer Science & Engineering in Merchant Engineering College, Basna.