ENERGY OPTIMIZED SECURE DATA AGGREGATION IN WIRELESS SENSOR NETWORK

Priyanka Gaikwad and Prof. Manisha Dhage

Department of Computer Engineering, Sinhgad college of engineering, Vadgaon bk, Pune, India.

ABSTRACT:
Wireless Sensor Network (WSN) is an evolving technology which is used in many applications including military surveillance, health monitoring, disaster detection and fire monitoring. WSN consists of many tiny sensor nodes which are used to monitor physical and environmental conditions such as temperature, motion and sound. Nodes are deployed in a hostile environment, and so security is the main concern in WSN. There is a lot of energy consumption when communication takes place between sensor nodes. Data aggregation techniques can efficiently help to reduce the communication overhead and consumption of energy by eliminating unnecessary data travelling back to the sink. The proposed project uses an attack resilient algorithm to enable the base station to securely compute the true estimate of aggregate.

Keywords: Data aggregation; security; energy; attack; wireless sensor networks.

INTRODUCTION

The WSN is a highly distributed networks consisting thousands of tiny, light-weight, inexpensive nodes known as sensor nodes and these nodes are deployed in large numbers to monitor the system by the measurement of relative humidity, temperature, pressure [2]. Sensor nodes are deployed and then they collaborate to form an network capable of reporting to a sink. WSNs have limited energy lifetime and batteries of sensor nodes are non-rechargeable. Most of the sensor nodes are placed in hostile environments and then nodes can be made compromised from attacks since any adversary can take control of the deployment area. There are certain limitations to sensor networks which include limited memory resources, limited bandwidth and transmission power and vulnerability of nodes to physical capture [4].

Data gathering is defined as the collection of sensed data from multiple sensor nodes to be transmitted to the base station for further processing [11]. As sensor nodes are resource constrained in terms of energy, it is unsuitable for all the sensor nodes to transmit the data directly to sink. Data gathered from all sensor nodes is redundant and highly correlated. In large sensor networks, the data generated is usually vast for the base station to do processing. Hence, there is a need to combine and summarize data at the sensor node which will ultimately eliminate the redundant packets transmitted to the base station resulting in energy conservation. This can be efficiently achieved by data aggregation. Data aggregation is the process of aggregating the data from sensor nodes to eliminate unnecessary transmission and provide quality information to the base station. It becomes easy for the adversary to change
the aggregated result and inject false data if it makes the node compromised, which will result in incorrect aggregate value [2].

The rest of the paper is organized as follows: In section II, data aggregation is discussed. In section III, the existing systems are discussed and comparison is made. In section IV, the proposed method is presented. In section V, results are shown. Finally, section VI concludes the paper.

[2] DATA AGGREGATION IN WSN

Aggregation can be simply defined as “the ability to summarize information”, quoted by Robbert Van Renesse [10]. Data aggregation usually involves the integration of data from many sensor nodes and the aggregator node transmits the aggregated data to the base station [9]. Nodes are of three types in WSN: sensor nodes, aggregator node which aggregates the data, and a querier (sink) which sends the query [2]. The aggregator node gathers data from sensor nodes; aggregate the collected data using aggregation function such as SUM, COUNT, MIN, MAX [8] and then transmit the result to the querier who generates the query. Communication between sensor nodes, aggregator nodes and the querier consumes a lot of energy. It is much difficult to develop an energy optimized data aggregation algorithms so that lifetime of network is enhanced. [Figure-1] data aggregation process using SUM function.

Figure: 1. Data Aggregation using SUM function.

[2] RELATED WORK

Przydatek et al.’s [1] algorithm discusses the secure information aggregation SIA that considers a WSN in which a large number of sensor node are placed in an area far from a user. Base station is used as an intermediary between the user and the sensor nodes. Sensor nodes send their sensed value to the base station, then the base station performs the aggregation process and forwards the aggregation result to the home server (user). The base
station is the single aggregator node in the network. SIA assures that if the result reported by the aggregator is accepted by the user, then the reported result is much close to the true aggregation value; else, if the reported value is different from the true value, the attack is detected by the user and rejects the reported result. SIA achieves this goal by constructing sampling mechanisms and interactive proofs. Aggregates can be computed using aggregate functions such as SUM, Count, Average, Min, Max.

Madden et al [7], discusses a tree topology used in TAG, so as to order to avoid double-counting sensor readings. TAG operates as follows: users insert aggregation queries from a storage-rich and powered base station. Operators that implement the query are distributed by piggybacking on the existing ad hoc networking protocol into the network. Sensor nodes routes the data at the base station in the routing tree. Based on an aggregation function and value-based partitioning specified in the query, the data is aggregated. TAG consists of two phases:

1) Distribution phase: User poses aggregate queries down into the network
2) Collection phase: The aggregate values are continually routed up from child nodes to parents.

Hu et al.’s algorithm [6] is the first attack-resilient hierarchical data aggregation algorithm that computes secure data aggregation (SDA). Securing the hierarchical data aggregation algorithms is a harder problem because the intermediate aggregator nodes may falsify their sub-aggregates. The proposed scheme offers data freshness, authentication and integrity. In this protocol, each aggregator node \( X \) forwards its inputs to its parent nodes in the aggregation tree so that parent nodes can verify if \( X \) has performed the aggregation correctly. This scheme is secure unless more than one malicious node is present.

Nath et.al [12] proposed synopsis diffusion algorithm that is a novel in-network aggregation framework that gives accurate estimations of duplicate-sensitive aggregates and enables robustness. The approach in synopsis diffusion is to use best effort and multi-path routing schemes together with duplicate-insensitive in-network aggregation schemes. Synopsis is defined as the partial result at a node, a small digest (e.g., bit-vectors, sample, histogram, etc.) of the data. [Figure-2] shows synopsis Diffusion over a ring topology.

- Synopsis Generation: A synopsis generation function takes a sensed value and generates a synopsis representing that data \( SG(\cdot) \).
- Synopsis Fusion: A synopsis fusion function takes two sensed value (synopses) and generates a new synopsis \( SF(\cdot, \cdot) \).
- Synopsis Evaluation: A synopsis evaluation function transforms a synopsis into the final answer \( SE(\cdot) \).

![Figure-2](Synopsis Diffusion over a ring topology [12].)
Roy et al. [13] proposed the attack-resilient hierarchical data aggregation that is designed for resilient hierarchical data aggregation in the presence of compromised nodes. In this algorithm, a portion of the total number of sensor nodes in the network generates a message authentication code (MAC) along with their sensed data as a response to query. These MACs and the sensed data are routed to the base station that is computed at each aggregator node in the hierarchy. The base station estimates the final aggregate value accurately by verifying the MACs, and filters the effect of any falsified sub-aggregate attack contributed by compromised nodes.

Roy et al. [14] proposed the verification algorithm that is designed to validate the computed aggregate by the base station. This algorithm is an aggregate computation and verification algorithm, also known as verification algorithm. The key observation in this algorithm is to minimize the communication overhead. This algorithm is used to verify the correctness of the aggregate of the entire network. There is no need for base station to receive authentication messages from all nodes. The goal of this algorithm is to detect the falsified sub aggregate attack generated by any compromised node. Algorithm involves two phases:

- Query Dissemination: In this phase, base station broadcasts a random number Seed, the aggregate name to compute and the chosen value of “test length”. In this phase, a set of ring is formed around the base station by the nodes based on their distance in hops.
- Aggregation Phase: Each node sends some authentication messages and executes the aggregation phase.

**TABLE I: Comparison of Existing Schemes**

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Based-on</th>
<th>Aggregate considered</th>
<th>No. of compromised nodes</th>
<th>Integrity</th>
<th>Authentication</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIA</td>
<td>Single</td>
<td>Median, Min, Max, Average</td>
<td>≥1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TAG</td>
<td>Hierarchical</td>
<td>Count, Sum</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SDA</td>
<td>Hierarchical</td>
<td>Count, Sum</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SD</td>
<td>Hierarchical</td>
<td>Count, Sum</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Roy et al</td>
<td>Hierarchical</td>
<td>Count, Sum</td>
<td>≥1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Verification</td>
<td>Hierarchical</td>
<td>Count, Sum</td>
<td>≥1</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
[4] PROPOSED METHOD

An algorithm is designed called attack resilient algorithm[15] in which base station computes true estimate of aggregate even in presence of an attack called falsified sub-aggregate attack in an energy optimized form.

MODULES:
1. Initialization Phase:
During the query distribution phase, each node shares a key with BS. Let the key of the node with ID X be denoted as KX. To authenticate a message to BS, a node X sends a MAC (Message Authentication Code) generated using the key KX. Further assumption is that each pair of neighboring nodes has a pair-wise key to authenticate its mutual communication and that BS cannot be compromised and it uses a protocol such as teals to authenticate its broadcast messages to the network nodes.

2. Aggregation Phase1
In phase one, each node forwards one randomly selected MAC for each 1 bit in synopsis. At the end of the phase, BS verifies the received MACs. A restriction is required on the number of MACs that a node can forward. Consider, if node X sends an aggregation message (sensed reading and corresponding MACs) to its parent node Y, Y does not accept more than one MAC for each 1 bit. This assumption can be enforced by employing authentication techniques in the communication procedure among neighboring nodes. If parent node is not malicious, it computes true aggregation value and sends it to the base station. Else, it injects false 1s in synopsis and forwards to BS. At the end, BS verifies the received MACs.

3. Aggregation Phase2
In this module, if parent node is malicious, it injects false 1s in synopsis. Parent node changes zero bit to one bit and one bit to zero bit, and so the aggregation value changes which it sends to base station along with the MAC generated. Base station has the received MAC on basis of which it estimates the false aggregation value. Base station will be able to detect the false value as no node send the MAC for it.

[5] RESULTS
The experiment is run in Network Simulator version 3. [Figure-3] shows that child nodes send their message along with MAC generated to parent node.
Figure: 3 Output Screen

[Figure-4] shows the simulation process in which child nodes send their message along with MAC generated to parent node. Also, Base station accepts message and MAC sent by parent node, the base station sends message to start the phase.

Figure: 4 Simulation Screen

After child node sends message, parent node computes aggregation using aggregate function OR. If parent node is not compromised, it changes bit 0 to 1 and 1 to 0. Parent node sends changed aggregated value and MAC to base station. Base station with the help of MAC detects falsified sub-aggregate value. This is shown in below [Figure-5].
The below graph shows the impact of Compromised Nodes on the Deviation of Aggregate “r”. The maximum deviation in estimate of aggregate depends on how many compromised nodes participate.

**CONCLUSION**

Discussion on security issues of in-network aggregation algorithms to compute aggregates is presented. The attack resilient algorithm will securely compute the true estimate of aggregate even in presence of attack. Energy optimization is achieved as MAC is generated for only 1 bit as well data aggregation is itself an energy optimized process. Thus, true estimate of aggregate even in presence of attack is presented in an energy optimized form.

**ACKNOWLEDGMENT**

I would like to thank my teachers who guided me in project. Also, I would like to thank my parents and friends for their support.
REFERENCES


