Secured Data Aggregation in Wireless Sensor Networks

Mrs. P. Padmaja  
Asst. professor, ECE Dept.,  
VITS, Deshmukhi, Hyderabad,  
Telangana, India.

Dr. G. V. Marutheswar  
Professor, Department of Department of EEE,  
S. V. U. College of Engineering,  
Tirupati, Andhra Pradesh, India.

Abstract  
In recent years Wireless Sensor Network has become potentially most important technology. Improvisation in wireless communications and electronics paved way for the production of low-cost, low-power, multifunctional miniature devices that can be implemented in remote sensing applications. As a result of these factors, it is possible to collect the process and disseminate valuable information that has been gathered from variety of environments using a sensor network comprising a large number of intelligent sensors. A network of energy-constrained sensors deploying over a region is considered, in that each sensor monitors its surrounding area and periodically generates information. The systematic gathering and transmission of sensed data to a base station for further processing is the basic operation in such a network. Sensors have the ability to carry out in-network aggregation or fusion of data packets reroute to the base station when data gathering. In such sensor system, the lifetime is the time in which the information can be gathered from all the sensors to the base station. In data gathering, from agreed energy constraints of the sensors expanding the system lifetime is a major threat. The data aggregator node or the cluster head combine the data to the base station and the malicious attacker may attack this cluster node. The base station cannot ensure the accuracy of the aggregate data sent to it, if a cluster head is compromised. Due to the uncompromised nodes, the existing systems may send several copies of aggregate results to the base station and the power consumption at these nodes is increased. When sensors are deployed at different locations in wider area, it is possible to compromising attacks by adversaries. False data injected in compromised sensors during data aggregation process which results in false decision making at the Base Station (BS). Simple average data aggregation process is suitable only in attacker free environment. So to filter the false data during data aggregation, induced by the attacker. For every round of data agg. reagation need to observe the behavior of nodes. So that it easy to minimize an impact of attacker contribution at the final result. For secure data aggregation process along with trustworthiness estimation using Trust wEighted Secure Data Aggregation algorithm (TESDA). Data aggregation process is optimized by performing aggregation in energy efficient manner through clustering. If the aggregator is compromised, then it affects entire aggregation accuracy. Hence it is necessary to propose a aggregation protocol that is resilient against compromised sensor and compromised aggregator in energy efficient and secure manner.

Keywords: wsn, TEEN, LEACH, base station, cluster head, tesda, sdaf, cluster, aggregation.

Introduction  
WSN applications are classified into four classes, they are event detection, periodic reporting, base station querying, and tracking. These classes are briefly explained as follows:  
i) Event Detection: The objective of sensor networks in this application class is to detect rare events, such as forest fires or intrusions, and to promptly communicate a report of such an event to the sink.  
ii) Periodic Reporting: The objective of the sensor networks in this type of application is to send periodic updates to the sink. Thus, there is regularity in terms of data gathering phases, and there is a steady flow of data from the sensor nodes to the sink. In-network data aggregation is useful in such applications because measurements of neighboring nodes are likely to be correlated, and could be used to reduce the amount of data that needs to be communicated to the sink.  
iii) Base Station Querying: In several application classes, the sink is not interested in data updates from all the nodes in the network. The sink may want updates from different regions at different times. Thus, requiring all the nodes to send their data to the sink at all times increases the energy consumption on communication as well as on computation. In such cases, the sink selectively queries a set of sensor nodes located in the region of interest. This results in a more energy-efficient use of resources.  
iv) Tracking: Tracking WSN applications are interested in detecting, localizing and tracking targets, and conveying the relevant information to the sink, in a timely fashion.

Designing issues of wsn routing protocols  
Due to sensors reduced computing, radio and battery resources, WSN routing protocols are expected to fulfil the following:  
Autonomy: As suming that a dedicated unit controlling radio and routing resources does not come in a WSNs way as it could be easily attacked. As there will be no centralized entity to make routing decisions, routing procedures are transferred to network nodes.
Energy Efficiency: Routing protocols must prolong network life while ensuring good connectivity to ensure intra node communication.

Scalability: As WSNs have 100s of nodes, routing protocols must work with these nodes.

Resilience: Sensors may unexpectedly stop operating for environmental reasons or due to battery consumption.

Device Heterogeneity: Though WSNs civil applications rely on homogeneity nodes, introduction of various sensors could reap benefits.

Mobility Adaptability: WSNs different applications could make nodes cope with own mobility, sink mobility or mobility of event to sense.

**Reliable Data Acquisition**

Wireless Sensor Networks (WSN) owe their genesis to the concept of ubiquitous Computing. Pervasive Computing represents the concept where technology is tightly integrated to all aspects of human life in a completely non-intrusive manner i.e. without the technology becoming the focus of attention. Pervasive computing envisions an intelligent and adaptive environment which continuously eases the interaction between the human being and his environment by sensing his actions and predicting his requirements from the environment around him, thus significantly enhancing the quality of interaction between the human being and the environment. It further envisions that this would be achieved by the presence of large number of tiny computing devices with sensing and radio communication capabilities, densely spread over the environment, gathering information about the environment, gathering information about the actions of the human subject and monitoring the interaction of the human subject with the environment. It also envisions that these computing devices would collaborate among themselves and be aware of the context while interpreting the data gathered by them to decide upon the subsequent action to be performed. Various types of context are defined in including Social context, Location context, Motivational Context, Temporal Context etc. Wireless Sensor Networks have emerged as an acceptable methodology for sensing events and acquire data generally spread over various locations in a geographic area to satisfy the requirement mentioned in the implementation of the event sensed by it. Reliability also includes the ability of the network to be tolerant to faults, within a limit, without comprising on the basic issue as mentioned above.

**Issues of Data Aggregation**

A network of energy-constrained sensors deploying over a region is considered, in that each sensor monitors its surrounding area and periodically generates information. The systematic gathering and transmission of sensed data to a base station for further processing is the basic operation in such a network. Sensors have the ability to carry out in-network aggregation or fusion of data packets reroute to the base station when data gathering. In such sensor system, the lifetime is the time in which the information can be gathered from all the sensors to the base station. In data gathering, from agreed energy constraints of the sensors expanding the system lifetime is a major threat [14]. The data aggregator node or the cluster head combine the data to the base station and the malicious attacker may attack this cluster node. The base station cannot ensure the accuracy of the aggregate data sent to it, if a cluster head is compromised. Due to the uncompromised nodes, the existing systems may send several copies of aggregate results to the base station and the power consumption at these nodes is increased [10].

**Proposed System**

An optimized and secure data aggregation protocol is proposed that is resilient to false data injection attack launched by compromised sensor and aggregator. Proposed protocol performs secure data aggregation process along with trustworthiness estimation. Data aggregation process is optimized by performing aggregation in energy efficient manner through clustering. Sensor network is divided into clusters and each energy efficient Clusterhead (CH) aggregates data collected from its cluster members and transmits to BS. Secure data aggregation is carried out in two phases first at the aggregator to make it resilient against compromised sensors and second at BS to make it resilient against compromised aggregator.

In first approach CH aggregates weighted average of reported value by each sensor in its cluster. Weight parameter is applied to reduce the impact of contribution of compromised sensor in the aggregation result. Trust value of the sensors is transformed into weight. Trust value of the previous round is applied as the weight of the current round. It is measured as the inverse proportion of deviation. Deviation is computed as the difference between aggregated value and the original value reported by the sensor. If the deviation is high then trust value is reduced. When the false data is injected by the compromised sensor, high deviation results in low trust value and weight that reduces the impact of attacker contribution in the final result.

In secondly, BS executes verification mechanism to check the validity of the aggregation result reported by the CH. It selects subset of nodes from each cluster and queries original data from those nodes. The data from those sensors are propagated without aggregation. BS aggregates the received information from the sensors. Then it computes the deviation of aggregated result from the reported value by CH. The trust value for the CH is estimated from the computed deviation. If the CH is compromised its deviation becomes high that results in reduced trust value. Hence the contribution of the compromised CH is reduced at the BS. Proposed protocol is optimized through cluster based data aggregation process and security is enhanced by making the protocol resilient against compromised aggregator along with compromised sensor. Trust worthiness measurement in the proposed work, assist in the secure data aggregation process as well as other network processes such as trust based routing trusted based cluster head selection and so on.

**Optimized Data Aggregation via Energy Efficient Clustering**

Input is Sensors ID and Residual energy and Output is Clusterhead each sensor attaches its ID and residual energy in its hello message. The node that receives the hello message add the sender in its neighbor list. Each node compares the residual energy of all of its neighbors. It selects the neighbor that has high residual energy as its ClusterHead. Cluster
member attaches its CH ID in hello message. On receiving hello message each node checks whether the CH ID mentioned in the hello message and its own ID are same. If it so it adds the corresponding sender its its member list. CH roles are rotated in every round in order to balance the energy consumption. Clustering rounds depends on the clustering period.

<table>
<thead>
<tr>
<th>QUEUE TYPE</th>
<th>Droptail/Priority Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUEUE LENGTH</td>
<td>200 Packets</td>
</tr>
<tr>
<td>ANTENNA TYPE</td>
<td>Omni Antenna</td>
</tr>
<tr>
<td>PROPAGATION TYPE</td>
<td>TwoRayGround</td>
</tr>
<tr>
<td>ROUTING AGENT</td>
<td>TESDA</td>
</tr>
<tr>
<td>TRANSPORT AGENT</td>
<td>UDP</td>
</tr>
<tr>
<td>APPLICATION AGENT</td>
<td>CBR</td>
</tr>
<tr>
<td>INITIAL ENERGY</td>
<td>50 Joules</td>
</tr>
<tr>
<td>TRANSMISSION POWER</td>
<td>0. 3watts</td>
</tr>
<tr>
<td>RECEPTION POWER</td>
<td>0. 1watts</td>
</tr>
<tr>
<td>SIMULATION TIME</td>
<td>30seconds</td>
</tr>
</tbody>
</table>

**Modules**

i. Optimized Data Aggregation via Energy Efficient Clustering
ii. False Data Injection Attack
iii. Secure Data Aggregation-Resilience against Compromised Sensors
iv. Secure Data Aggregation-Resilience against Compromised Aggregator
v. Performance Evaluation

**Optimized Data Aggregation via Energy Efficient Clustering**

**Input:** Sensors ID and Residual energy

**Output:** Clusterhead

Each sensor attaches its ID and residual energy in its hello message. The node that receives the hello message add the sender in its neighbor list. Each node compares the residual energy of all of its neighbors. It selects the neighbor that has high residual energy as its ClusterHead (CH). Cluster member attaches its CH ID in hello message. On receiving hello message each node checks whether the CH ID mentioned in the hello message and its own ID are same. If it so it adds the corresponding sender its its member list. CH roles are rotated in every round in order to balance the energy consumption. Clustering rounds depends on the clustering period.

**False Data Injection Attack**

Input is false data through compromised sensor and output is falsified aggregate Sensed result (X) of every sensor (S) is submitted to CH. It derives the aggregated result (A) by taking weighted average of collected information. Attacker compromises the sensor and alters its sensed value to very low or high to distort the aggregation result. False data from compromised sensor, reduces aggregation result as CH computes aggregation result from the reported value. When the CH submits the falsified aggregate to the base station, it leads to false decision making.

**Secure Data Aggregation-Resilience against Compromised Aggregator**

Input is false data through compromised aggregator, actual data from genuine aggregators output is reduced trust of compromised aggregator. Filtered attacker contribution in aggregated result.

If the attacker compromises the CH, it alters the aggregated result (Z) before submitting it to the Base Station (BS) in order to distort the final aggregation result (BS(A)) at the

---

**Simulation Model**

In WSN, sensors S are located in the network area of X * Y. X is the network width and Y is the network height. Energy model is applied on every sensor node with initial energy, transmission power, reception power, sleep power and idle power. Sensors are formed into k clusters with k CH. Each CH acts as aggregator and aggregates data collected from n sensors and send it to BS. Attackers can compromise sensors as they are deployed in a hostile unattended environment. It is assumed that if the sensor is compromised, all the secret information resided in the sensors can be extracted by the attacker. Hence no cryptographic primitives can be applied for security. Aggregation at the CH and BS is the function of sensed data X of ’n’ sensors in each ‘k’ cluster, and trust value of each sensor T in ‘m’ rounds.

**System Model**

Here the simulation model is shown in table 1.

**Figure 1: Implementation method**

**Simulation Model**

Here the simulation model is shown in table 1.

**Table 1: Proposed simulation model**

<table>
<thead>
<tr>
<th>SIMULATOR</th>
<th>Network Simulator 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF NODES</td>
<td>30, 40, 50, 60</td>
</tr>
<tr>
<td>AREA</td>
<td>500m x 500m</td>
</tr>
<tr>
<td>COMMUNICATION RANGE</td>
<td>250m</td>
</tr>
<tr>
<td>INTERFACE TYPE</td>
<td>Phy/WirelessPhy</td>
</tr>
<tr>
<td>MAC TYPE</td>
<td>802. 11</td>
</tr>
</tbody>
</table>
base station. To overcome this issue BS verifies the trustworthiness of the CH (TCH) through the original sensed information collected from the subset of CH nodes (k). BS aggregates the collected data from subset of CH nodes and finds the deviation (DCH) between reported result by CH. If the deviation is high BS reduces the trust value of the CH as the inverse proportion of the deviation and direct proportion of the non-deviation factor (NDF_CH). Hence the impact of falsified data contributed by the CH is reduced at the base station. Final aggregated result at BS is the trust weighted.

**RESULTS**

By implementing TESDA algorithm for the sixty nodes, first phase of experiment is forming cluster, data aggregation at the cluster head. Optimized data aggregation via energy efficient clustering, false data through compromised sensor, false data through compromised sensor, actual data from genuine sensors, false data through compromised aggregator, actual data from genuine aggregators, Performance Evaluation The TESDA based proposed approach is evaluated and compared with existing approach Secure Data Aggregation in WSN using Filtering (SDAF).

**Energy efficient clustering**

i. Residual energy of the neighbor is accessed, high energy neighbor is selected as CH and it is attached with hello message

ii. On receiving the hello message, the node checks it's ID and the clusterhead id mentioned in the hello message. If it matches it adds the node that sent the hello message to its member list.

iii. Each node is aware of cluster member and it's member list

**Secure Data Aggregation-Resilience against Compromised Aggregator**

**Input:** False data through compromised aggregator, actual data from genuine aggregators

**Output:** Reduced trust of compromised aggregator, Filtered attacker contribution in aggregated result

If the attacker compromises the CH, it alters the aggregated result (Zr) before submitting it to the Base Station (BS) in order to distort the final aggregation result (BS(Ar)) at the base station. To overcome this issue BS verifies the trustworthiness of the CH (TCH) through the original sensed information collected from the subset of CH nodes (k). BS aggregates the collected data from subset of CH nodes and finds the deviation (DCH) between reported result by CH. If the deviation is high BS reduces the trust value of the CH as the inverse proportion of the deviation and direct proportion of the non-deviation factor (NDF_CH). Hence the impact of falsified data contributed by the CH is reduced at the base station. Final aggregated result at BS is the trust weighted summation of data reported by the CHs in the round.

\[
\text{BS Average Data (BSAvg)} = \sum_{i=1}^{k} Z_{i} \quad r=1, 2, \ldots, m
\]

Deviation \( DCH(r) = \text{BSAvg} - Z_{i}(r) \)

Total Deviation \( \text{TDCH}(r) = \sum_{i=1}^{k} DCH_{i}(r) \)

Non Deviation Factor \( NDF_{CH}(r) = \text{TDCH}(r) - DCH_{i}(r) \)

Total Non Deviation \( \text{TNDF}_{CH}(r) = \sum_{i=1}^{k} NDF_{CH}(r) \)

\[
\text{Trust TCH}(r) = \frac{\text{NDF}_{CH}(r)}{\text{TNDF}_{CH}(r)}
\]

Weight \( w_{ch}(r) = TCH(r) \)

Aggregated result \( \text{BSA}_{r} = \sum_{i=1}^{r} w_{ch} Z_{i} \quad r=1, 2, \ldots, m \)

**Data transmission to clusterhead**

i. On sending route request to the BS, sensor assigns destination address as its cluster head id and port as 255 which is for transferring packet to routing agent of clusterhead and stores the original port and address information of BS. It enqueue the packet in buffer and send route request to clusterhead.

ii. On receiving reply from clusterhead, sensor access packet from buffer and send it to clusterhead

iii. On receiving the data each node, clusterhead checks whether it has received a data from all of its members. After it received from all of its members it increments the received data count and stores the received data. If the received data count equals to the member count, aggregation is performed on the received data. Clusterhead assigns destination address as BS which had been stored in step1, enqueue the packet and initiates route request.

iv. On receiving route reply from BS, cluster head access packet from buffer and send it to BS.

![Figure 2: Data Aggregation at CH, Attack detection and Filtering](image)

**Data Aggregation at CH, Attack detection and Filtering**

i. Average data from received data is found.

ii. Deviation of each sensor data from average is calculated

iii. Attacker incurs high deviation
iv. Non deviation degree is calculated from deviation and total deviation
v. Genuine sensor has high non deviation factor whereas attacker has low non deviation factor. Trust value for the corresponding sensor data is calculated based on the non deviation factor that results in a weight for the sensor data
vi. Aggregated temperature is obtained through the summation of weighted data of every sensor data in the cluster.
vii. Deviation percentage is calculated. If it is greater than 40% then the weight for the corresponding sensor is reduced to 1/member_count and the remaining weight is distributed to remaining members. This reduces the attacker’s impact on aggregation result

Performance Evaluation
The TESDA based proposed approach is evaluated and compared with existing approach Secure Data Aggregation in WSN using Filtering (SDAF) [1] for the following parameters using the ns-2 simulation.

Data Aggregation Deviation
It refers to the percentage of the aggregation error. It is calculated as the ratio of deviation to the true value sensed by the sensors.

Network lifetime
It refers to the time till half of the nodes in network remains alive.

Overhead
It refers to the total number of control packets involved for the secure data aggregation process.

Attacker Impact Reduction Ratio
It refers to the ratio of reduced trust of the compromised sensors from the actual trust of the compromised sensors resided in the network.

Energy Consumption
It refers to the total amount of energy required for data aggregation process.

Data aggregation deviation is reduced in proposed protocol TESDA than existing protocol SDAF. Because, proposed protocol finds the aggregated result based on the deviation of the values reported by all the sensors in the cluster. But SDAF simply takes OR operation of received synopsis without considering deviation. With the increasing number of sensors, number of genuine sensors are increased that reduces the aggregation deviation.

References
