SDA-seech: Secure data aggregation using seech algorithm in wireless sensor networks

K. Hemapriya  
Research Scholar, Department of Computer Applications, Tiruppur Kumaran College for Women, Tirupur, India

Dr. C. Kalaiselvi  
Associate Professor & Head, Department of Computer Applications, Tiruppur Kumaran College For Women, Tirupur, India

Abstract---Wireless sensor networks (WSNs) have grown in popularity as a network method. However, owing to the broadcast communication method and unsupervised deployment nature of WSNs, the Sensor Nodes (SN) often becomes targets of attackers. Due to the few processing and energy resources, basic data aggregation methods such as averaging are required. While it may prevent sensitive data from being hacked, it requires regular message exchanges across a network, imposing a significant burden on SNs with limited resources. This article presents the SDA-SEECH technique for securely transmitting data without collusion attacks to address these concerns. The recommended system has improved collusion resistance, accuracy, and convergence speed. The SEECH algorithm is used to perform energy-efficient clustering. The optimal path was determined using the SEECH Route and HOP-TIME algorithms. Theoretical study and simulation measurement demonstrates that the proposed aggregation methods outperform existing protocols regarding privacy protection and communication efficiency.

Keywords--- SEECH, SDA-SEECH, HOP TIME, WSN, aggregation, secure transmission.

Introduction

Wireless sensor networks (WSNs) are rapidly being employed in various applications, including monitoring natural habitats, detecting forest fires, and conducting military surveillance. Once deployed in the area of interest, SNs form a multi-hop network with the Base Station (BS) as the central control point. Typically, a SN's compute capabilities and energy reserves are highly restricted...
A simple way to gather sensed data from the network is to enable each SN’s reading to be transmitted to the BS, maybe through other intermediary nodes, before processing the data. This strategy, however, is excessively costly in terms of transmission overhead (or energy spent) [3].

By measuring members’ trustworthiness in distributed systems ranging from WSNs and e-commerce infrastructure to social networks, trust and reputation systems contribute significantly to a broad variety of distributed systems [4]. At every point in time, a trustworthiness rating accumulates the participants’ actions up to that point. It must be resilient to a variety of errors and hateful conduct [5]. There are several reasons for attackers manipulating participants’ trust and reputation scores in a distributed system, and such manipulation can significantly impair the system’s performance. Malicious attackers’ primary aim is trust and reputation aggregation methods [6] [8].

Recently, trust and reputation have been proposed as secure techniques for WSNs [10] [11]. Although sensor networks are being used in many application areas, determining the integrity of data given by dispersed sensors has remained a difficult problem. Sensors installed in hostile settings may be compromised by attackers attempting to insert fake data into the system through compromising node attacks. Assessing the data’s reliability becomes a difficult challenge in this setting [13] [15]. As the computational power of very low power processors increases dramatically, primarily due to mobile computing demands, and the cost of such technology decreases, WSNs will be able to afford hardware capable of implementing more sophisticated data aggregation and trust assessment algorithms; an example is the recent emergence of multi-core and multi-processor systems in SNs [16].

Computing aggregates in-network (i.e., combining incomplete results at intermediate nodes during message routing) considerably decreases communication and energy consumption in big WSNs [19]. Numerous data gathering techniques for WSNs use a technique that first constructs a spanning tree rooted at the BS and then aggregates data inside the network along the tree. Count and Sum are two significant aggregates that the research community considers. Notably, these aggregates are easily generalized to the predicates Count (e.g., the number of sensors with a reading greater than 100 units) and Sum. Additionally, the average may be calculated using the Count and Sum. Additionally, the Sum technique may be extended to calculate any order’s Standard Deviation and Statistical Moment. The rest of this paper is constructed as follows. Section II presents related authors work on SDA. The proposed model is described in section III. Then, in section IV described with Results and Discussion with comparative analysis. Finally conclusion presented in the section V.

**Background Study**

K. Shanmugam et al. [2] presented a method for cluster-based WSNs to transmit data securely (CWSNs). Clusters are established dynamically and intermittently, avoiding the security concerns associated with in-network aggregation techniques used to calculate aggregates such as count and sum predicates. Secure and
efficient data transmission in WSNs employing the SET protocol, in particular, demonstrated that a false sub-aggregate attack initiated by a few compromised nodes might introduce arbitrarily large errors into the aggregate estimate of the BS. An attack-resistant computing technique would ensure that the aggregate is successfully computed even in the face of an assault. The proposed SET-IBS and SET-DTA protocols outperform current secure protocols for CWSNs in terms of performance, satisfying CWSN security criteria and resolving the orphan node issue of insecure transmission.

Suraj et al. [3] proposed a novel approach to SDA by utilizing the Efficient Distributed Trust Model (EDTM) to avoid data falsification and reduce energy consumption by avoiding multiple redundant data transmissions from multiple sensors via synopsis diffusion during the aggregation process. This results in a shorter average sensor lifetime and lower overall bandwidth utilization. The suggested system securely aggregates data from many SNs to the BS in a dispersed fashion, allowing neighbor nodes to interact with one another. Using EDTM, hostile nodes are recognized and avoided depending on the SN’s trustworthiness. As a result, data is aggregated securely utilizing synopsis diffusion.

Hevin Rajesh. D et al. [4] suggested a strategy for dealing with security threats and preventing a compromised cluster member or cluster head from gaining direct access to data during data forwarding in WSNs. The suggested approach enhances the security of data aggregation in WSNs by using a protected cluster key distribution mechanism. The suggested technique employs simulated annealing to achieve energy-efficient data aggregation. Additionally, the phase of key refreshing mitigates the security risk posed by a hacked node. Because each cluster member only owns a portion of the secret cluster key, and the cluster key is concealed from each member. The proposed scheme’s advantages include eliminating incorrect data as soon as feasible and a reduced computation and communication overhead for authentication compared to existing relevant techniques. The suggested architecture significantly exceeds the current network’s lifespan, node energy consumption, PDR, latency, and filtering efficiency. This technique is well-suited for transmitting sensitive data in densely populated large-scale WSNs.

Limin Shen, Jianfeng Ma, and colleagues [7] developed an identity-based aggregate signature system for WSNs with a designated verifier. This approach may help WSNs maintain data integrity while reducing bandwidth and storage costs according to the benefit of aggregate signatures. Additionally, the security of this identity-based aggregate signature technique is rigorously shown using the random oracle model’s computational Diffie-Hellman assumption. Due to the limited computational, memory, and battery capacity available to SNs, safe and energy-efficient data aggregation techniques should be devised in WSNs to lower the energy cost of data collecting, processing, and transmission.

Di Tang, Tongtong Li, and colleagues [9] suggested a unique secure, and efficient Cost-Aware SEcure Routing (CASER) protocol to resolve these two opposing difficulties by adjusting two parameters: energy balance control (EBC) and probabilistic based random walking. The energy usage is very disproportional to
the uniform energy deployment for the current network structure, significantly reducing the sensor networks' lifespan. This challenge will be solved by an efficient non-uniform energy deployment technique that optimizes the lifespan and message delivery ratio while maintaining the same energy resource and security requirements. Increase the lifespan, and the total number of messages sent more than fourfold under the same assumption for non-uniform energy deployment. The suggested CASER protocol can achieve a high message delivery ratio while avoiding routing traceback attacks. A cost-effective and safe cost-aware secure routing (CASER) protocol for WSNs balances energy usage and extends network lifespan. CASER's message forwarding capability enables it to support different routing schemes, extending the message's life while boosting routing security.

A. Selcuk Uluagac et al. [12] introduced the Secure SOURCE-Based Loose Synchronization (SOBAS) protocol for securely synchronizing network events without transmitting explicit synchronization control messages. In SOBAS, nodes encrypt each message using their local time values as a one-time dynamic key. Thus, SOBAS implements an excellent dynamic en-route filtering method, filtering harmful data from the network. It can synchronize sink events as rapidly, precisely, and secretly possible. SOBAS decreases the number of control messages required for a WSN to function, resulting in significant energy savings and a decreased potential for malevolent nodes to eavesdrop, intercept, or become aware of the network's existence. H. R. Roopashree1 et al. [14] presented SARDS (Secured Anonymous Routing with Digital Signature), a system that verifies the routing information transmitted between sensors in a WSN. SARDS is based on elliptical curve cryptography and authenticates all communication nodes in the network. Additionally, the system provides a twofold layer of security by using a new signature-based public-key encryption strategy. The study's findings indicate that SARDS outperforms current secure and energy-efficient routing systems in terms of performance. SARDS offers a twofold layer of security by generating and authenticating signatures using two distinct mathematical formulations.

Triana Mugia Rahayu et al. [17] stated that for a safe data aggregation protocol to be successful, it must include a secure underlying routing protocol. Energy-aware protocols based on LEACH and ESPDA seek to solve this problem by combining a safe routing mechanism with a SDA approach. It assesses its security efficacy and energy efficiency, constantly bearing in mind that they are always trade-offs. It offers a secure routing protocol that takes safe data aggregation into account to address this issue. It combines a security mechanism based on LEACH with ESPDA, a SDA system optimized for clustered WSNs. Security, communication, and computational complexity analyses have been performed on the protocol design. Security and energy efficiency are always trade-offs. Keeping the distance between them as small as possible is a difficult problem for WSN protocol design. Compared to the other protocols under examination, it devotes a large portion of its energy-efficient characteristics to enabling security. We want to emphasize the critical nature of including safe routing and data aggregation in designing a protocol for WSNs to build energy-aware but secure protocols.

U. Korupolu1 et al. [18] introduced a novel approach for data aggregation dubbed Redundancy Elimination for Accurate Data Aggregation (READA). The READA
technique will modify the aggregate in two ways: via a monitoring system and event detection. Two ways will be employed in this case: compression and grouping. A data grouping strategy is to be employed for prediction. SNs with the same profiles and logically - small-scale characteristics are combined to produce new SNs. Thus, the number of nodes may be decreased linearly over time. The use of grouping expressions constructs the elevator.

**Proposed Model**

Data aggregation is presenting all data in a combination and summary. The WSN is a collection of SNs tied together to perform a single program's operation. These small and lightweight sensors are randomly put in undisturbed areas to gather information on the movement.

**Secure Data Aggregation Techniques**

The BS developed a tree-based verification technique for determining if the final aggregate, Count, or Sum, is collusive. We cannot use this concept to check a synopsis since the calculation of the synopsis is duplicate-insensitive. A verification technique was developed for calculating Count and Sum using the synopsis diffusion approach. Our technique is similar, except our system takes a unique way to reduce communication costs. Additionally, we do an in-depth theoretical study to determine the optimal trade-off between security and communication overhead. Many innovative methods for "safe outsourced aggregation" have been developed; nevertheless, these algorithms are not optimized for WSNs.

While algorithms and our verification procedure protect the BS from accepting a bogus aggregate, they do not ensure that the aggregate is successfully computed in the case of an assault. Additionally, several researchers developed attack-resistant computing methods that enable the BS to filter out the aggregated contributions from compromised nodes. We developed the first attack-resistant hierarchical data aggregation algorithm. However, this strategy is safe when just one rogue node is present. In the case of a few hacked nodes, the authentication step of SDAP may be utilized to calculate Count and Sum at a high cost. Recent work has presented an attack-resistant aggregation approach for calculating Count and Sum using a sampling strategy. This algorithm can produce a close match to the target aggregate despite adversarial interference. We earlier provided a synopsis diffusion framework with an attack-resistant aggregation technique. Compared to all these attack-resistant systems, the verification methodology we present in this study has a relatively low overhead. We emphasize that attack-resistant computing is a broader issue than verification. Our prior study addressed a more broad issue than the one covered in this paper, which may generate some reader concerns. We emphasize that, although our earlier approach covers a broader issue, it is inefficient and lacks a lightweight verification technique.
Threat Model

By itself, the outline dissemination structure is devoid of security safeguards. As a result, it is vulnerable to various assaults by unauthorized or compromised nodes. We can supplement the aggregation framework with industry-standard authentication and encryption methods to prevent unauthorized nodes from interfering with (or spying on) interactions between honest nodes. As a result, we see no need to contemplate attacks from unauthorized nodes. Cryptographic techniques are incapable of preventing assaults by compromised nodes since the adversary may access cryptographic keys from compromised nodes.

Violating data privacy

A hacked node that also functions as an in-network data aggregator may leak (to the attacker) sensor readings (and sub aggregates) received from child nodes. Numerous academics presented methods that preserve privacy.

The collusion of the local value

A hacked node may impact the aggregate value by colluding with its sensor readings. There are three situations. The instance I If an honest node’s local value may be any value (i.e., not constrained by the application domain), then a hacked node can claim to perceive any value. There is no method to detect the Collusion local value attack in this instance. If an honest node's local value is bounded and the compromised node falsifies the local value inside the limit, there is no way to detect such an attack. Case (iii): An honest node's local value is finite, whereas a compromised node's local value is inflated beyond the limit.

The collusion of the sub-aggregate

A hacked node is capable of colluding with the sub-aggregate, which is meant to calculate based on messages received from child nodes. It is difficult to defend against this onslaught.

SEEECH Algorithm

The SEEECH protocol (Scalable Energy Efficient Clustering Hierarchy): The SEEECH protocol operates once nodes are distributed over the network environment. This protocol begins with the setup phase.

While LEN>0 do
  If |R_{CHC}|>K_{CH}, then
    Delete a member of R_{CHC} with the smallest residual energy;
  End if
  LEN←LEN -1;
End while

While |R_{CHC}|>K_{CH} do
  For i=1 to |R_{CHC}| do
    Score=d_{ij};
  End
Find $i$ so that $\max \text{score}_i = \max \{\text{score}_1, \text{score}_2, \ldots, \text{score}_{|\text{RCH}|}\};$
Delete $i^{th}$ member from $\text{R}_{\text{CHC}};$
Score←0;
End while
$\text{R}_{\text{CH}} \leftarrow \text{R}_{\text{CHC}};$
End

**Hoptime Algorithm**

Begin
Initialize SEECH with random population $K$ and number of cluster centers $D.$
Define the light absorption coefficient $Y$ ($Y = 1$)
If[$t < \text{max(iteration)}$]
the objective function using LEN
If[$i_j > i_i$]
Move SEECH $i$ toward $j$ based on the locations.
End if
End if
Find the current best and global best values of SEECH.
Initialize the center of the cluster by global best.
Update the cluster location
Stop until the maximum iteration is reached.
End

**Sceeh_route algorithm**

- Step 1: Initialize the number of population (nodes) and nodes.
- Step 2: Evaluate the fitness function of each node using SEECH and rank them according to the fitness values.
- Step 3: Apply coded operators to the top $n$ nodes and update new $n$ nodes.
- Step 4: From the initial population, selects the $n$ best nodes according to their fitness.
- Step 5: Apply crossover to update the best $n$ individuals using HOP-TIME
- Step 6: Apply mutation to the best $n$ SNs.
- Step 7: Apply SN operators for updating the $n$ nodes with the worst fitness.
- Step 8: Update the new positions of the nodes.
- Step 9: Extract the best harmony (node) as CH.
- Step 10: Repeat until the maximum iteration is reached

**Detecting Collusion Sub-aggregate Attack:**

The sink estimates the aggregate using the lowest-order bit $r$, which is always 0 in the final fused synopsis; a compromised node would need to modify its fused synopsis so that the value of $r$ is affected. It may easily achieve this by simply putting ones into one or more bits in locations $j$, where $r \ j \ k$, in the fused description it broadcasts to its parents. Not that the compromised node need not know the real value of $r$; it may just change some higher-order bits to 1 in the hope that this affects the value of $r$ calculated by the sink. Because the synopsis fusion function is a bitwise Boolean OR, the resultant synopsis calculated at the
sink will include contributions from the compromised node. Let \( r' \) be the lowest-order bit in the corrupted synopsis 0, and \( r \) be the lowest-order bit in the proper synopsis that is 0. The sink's estimate of the aggregate will thus be a factor of two bigger than the accurate estimate. It's simple to understand how the hacked node might introduce a significant amount of inaccuracy into the final estimation of the sink using the aforementioned approach. Observe that even a single node may successfully launch this attack due to the synopsis dispersion approach's multi-path routing, which increases the likelihood that the Collusion synopsis will be relayed to the BS. If \( p \) denotes the packet loss rate and each node in the aggregation hierarchy has parents, then the chance of success for this attack is \((1-p)h\), where the compromised node is \( h \) hops distant from the sink. For instance, if \( p = 0.2, \alpha = 3, \) and \( h = 5 \), the likelihood of a successful assault is 96 percent.

**Robust data Aggregation and Encoding & Decoding**

A unique technique for evaluating the bias and variance of noise in sensor measurements. The variance and bias of sensor noise may be considered the distances between the sensor readings and the signal's real value. The distance metrics used to evaluate the bias and variance of sensors also make sense in the presence of non-stochastic errors. The heuristic point removes the "systematic component" of the error by subtracting a quantity that, in the case of a stochastic error, corresponds to an estimate of bias; this enables us to estimate the variability surrounding such a systematic component of the error, which corresponds to variance in the case of stochastic errors. An estimating technique equivalent to the MLE formula for zero-mean normally distributed errors but using estimated rather than real variances. As a result, the predicted value \( r_t \) of the measurements equals the real value of the quantity measured and serves as the only parameter in the likelihood function. The attacker uses the IF algorithms' flaw, which stems from an incorrect assumption about the initial trustworthiness of sensors. To remedy this problem, we propose to use robust data aggregation as the starting reputation for these algorithms. Additionally, the starting weights for all SNs may be calculated based on the sensor data's distance from such an initial reputation.

**Correlated noise**

Our first variance estimate algorithms assumed that sensor errors are uncorrelated. Thus, we examined how our technique performs when the noise gets correlated and compared it to current methods in the same situation. Thus, we assume in this case that sensor errors are no longer uncorrelated.

**Energy Consumption**

The models of energy consumption for a link between two nodes are a transmission of a unit message; the model of the minimum energy needed for a link between nodes \( v_i \) and \( v_j \) is \( P_{ij} = k_1 (r_{ij})^\beta + k_2 \), where \( r_{ij} \) Euclidean distance between \( v_i \) and \( v_j \), \( k_1 \) is a constant dependent on the properties of the antenna, \( \beta \) is the path loss exponent that depends on the propagation losses in the medium, and \( k_2 \) is a constant that accounts for the overheads of electronics and digital processing. Note that we assume that each multicast session only multicasts a
unit-length message. The transmission power of each node determines the network's connection. Each node can dynamically adjust its transmission power. Each multicast tree in which a node participates might employ a variable power level. Each node is equipped with omnidirectional antennas. Each network node vi has two coverage areas: (1) control coverage area (CRi); and (2) data coverage area (DRi), where DRi ⊆ CRi. These coverage zones are determined by the transmission power that node vi selects for transmitting its control and data packets.

**Crossover Scheme**

The crossover operator generates a child $T_c$ by identifying the same links between $T_a$ and $T_b$, and retaining these common links in $T_c$. According to the fitness function definition, the “better” individual has a higher probability of being selected as a parent. Thus, the common links between two parents are more likely to represent the “good” traits. However, retaining these common links in $T_c$ may generate some separate sub-trees. As a result, connections must be chosen to join these sub-trees to form a multicast tree. The following procedure is used to link distinct sub-trees. To begin, two distinct sub-trees are randomly chosen from these sub-trees. The chosen sub-trees are then joined through the route with the least latency to construct a new sub-tree. This technique is repeated until a multicast tree is formed. We add two nodes to determine the route with the least latency between two subtrees. A single node is linked to all nodes in a sub-tree through zero-delay connections. Similarly, the other node is linked through zero-delay connections to all the nodes in the other sub-tree.

**Results and Discussion**

As a consequence of this technique, several successful deliveries are achieved, leading to a rise in PDR. Increased data rate lowers the time necessary for a packet to reach its destination, resulting in a decreased probability of collision and an increase in the number of successful packets. However, it is crucial to realize that the linear connection between PDR and data rate does not hold for ever-increasing data rates because of the restricted bandwidth of acoustic transmission. When PDR reaches a certain level, it cannot be further raised.

**Performance metrics**

The following performance measures are used to assess the network’s security, energy efficiency, and dependability.

- Malicious node detection is the percentage of accurately recognized harmful nodes to the total number of malicious nodes injected.
- Energy consumption: The amount of energy used by nodes in a network to transport data.
- PDR: Packet delivery ratio is the percentage of successful data packet receipt to total data packet transmission.
Simulation Results

We simulate our proposed work in NS2 Simulator. We compare our SDA-SEECH Model with the Coordinated Multi-point (CoMP) transmission and Weighting Depth and Robust Data Aggregation (RDA) method. The Network size is 550 x 480 m. The aggregation of the SDA-SEECH method depends on the energy of the current forwarder as well as aggregation in the next expected forwarding region; and the Shortest Path Index (SPI), which is calculated based on several hops to the sink and average Depth of neighbors in the next expected hop and to solving the aggregation problem and improve the Packet Delivery Ratio (PDR).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation Time</td>
<td>900(s)</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>0 to 89</td>
</tr>
<tr>
<td>Data Rate</td>
<td>1Mbps</td>
</tr>
<tr>
<td>Routing Protocol</td>
<td>SEECH AODV</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2 Mb</td>
</tr>
<tr>
<td>Simulation Area</td>
<td>1300 x 2250 m</td>
</tr>
<tr>
<td>Transmission Range</td>
<td>250m</td>
</tr>
<tr>
<td>Threshold</td>
<td>100dbm</td>
</tr>
<tr>
<td>MAC</td>
<td>802.11</td>
</tr>
<tr>
<td>Power monitor threshold</td>
<td>120dbm</td>
</tr>
</tbody>
</table>

Figure 4 illustrates the data transmission delay. The CoMP and RDA methods are used for a high transmission delay. The SDA-SEECH method has less delay in transmission. The X-axis represents the sink node, and the Y-axis represents the delay in time.
Figure 3 illustrates the data flow level by packet transmission: the CoMP and RDA methods are used as low data flow levels. The SDA-SEECH method has a high data flow level by comparing the existing methods. The X-axis represents the data flow in seconds, and the Y-axis represents the packets.

Figure 2 illustrates the time synchronization with energy consumption. The SDA-SEECH method energy consumption is very less. The CoMP and RDA methods are high energy consumption of active nodes. The X-axis represents the time in seconds, and the Y-axis represents the energy level.
Figure 6 illustrates the routing with throughput. The accuracy of SDA-SEECH is increasing the message communication. It shows the throughput comparison; the SDA-SEECH has a better throughput than CoMP and RDA methods. In X-axis represents the time, and Y-axis represents the throughput levels.

Figure 5 illustrates the routing with bandwidth frequency: the CoMP and RDA methods used high bandwidth frequency levels. The SDA-SEECH method has less usage in bandwidth. The X-axis represents the average energy, and the Y-axis represents the bandwidth frequency.
Conclusion

This article presented the SDA-SEECH technique for securely transmitting data without collusion attacks. Additionally, we presented a modification to data aggregation utilizing the SEECH, HOP-TIME, and SEECH-ROUTE algorithms by giving an initial estimate of the SNs’ trustworthiness, making the algorithm’s collusion robust and more accurate, allowing for quicker convergence. The proposed SDA-SEECH approach compares the simulation results with CoMP and RDA. Compared to other available approaches, SDA-SEECH outperforms them in terms of PDR and Throughput. Our future work will focus on implementing the suggested algorithm in real-world applications, particularly in home automation systems that operate with the assistance of sensors. Additionally, the proposed algorithm’s performance will be evaluated using other performance measures such as end-to-end latency and network throughput.

References


