Circuits and Systems, 2016, 7, 1404-1420
Published Online June 2016 in SciRes. http://www.scirp.org/journal/cs
http://dx.doi.org/10.4236/cs.2016.78123

Scientific Research Publishing

SAC-TA: A Secure Area Based Clustering for Data Aggregation Using Traffic Analysis in WSN

Mohanbabu Gopalakrishnan*, Gopi Saminathan Arumugam*, Karthigai Lakshmi Shanmuga Vel

Department of Electronics and Communication Engineering, SSM Institute of Engineering and Technology, Dindigul, India
Email: *shamyubabu@gmail.com, *agsaminathan@gmail.com

Received 23 March 2016; accepted 2 May 2016; published 9 June 2016

Abstract

Clustering is the most significant task characterized in Wireless Sensor Networks (WSN) by data aggregation through each Cluster Head (CH). This leads to the reduction in the traffic cost. Due to the deployment of the WSN in the remote and hostile environments for the transmission of the sensitive information, the sensor nodes are more prone to the false data injection attacks. To overcome these existing issues and enhance the network security, this paper proposes a Secure Area based Clustering approach for data aggregation using Traffic Analysis (SAC-TA) in WSN. Here, the sensor network is clustered into small clusters, such that each cluster has a CH to manage and gather the information from the normal sensor nodes. The CH is selected based on the predefined time slot, cluster center, and highest residual energy. The gathered data are validated based on the traffic analysis and One-time Key Generation procedures to identify the malicious nodes on the route. It helps to provide a secure data gathering process with improved energy efficiency. The performance of the proposed approach is compared with the existing Secure Data Aggregation Technique (SDAT). The proposed SAC-TA yields lower average energy consumption rate, lower end-to-end delay, higher average residual energy, higher data aggregation accuracy and false data detection rate than the existing technique.

Keywords

Data Aggregation, False Data Injection Attacks, Malicious Nodes, One-Time Key Generation, Secure One-Time (SOT) Key and Wireless Sensor Networks (WSNs)

*Corresponding author.

How to cite this paper: Gopalakrishnan, M., Arumugam, G.S. and Vel, K.L.S. (2016) SAC-TA: A Secure Area Based Clustering for Data Aggregation Using Traffic Analysis in WSN. *Circuits and Systems*, **7**, 1404-1420.
http://dx.doi.org/10.4236/cs.2016.78123

1. Introduction

WSNs are progressively deployed for applications such as battlefield surveillance applications in military, industrial control applications, wildlife habitat monitoring, healthcare monitoring, forest fire prevention, etc. In these applications, the data gathered by sensor nodes from their physical environment require to be assembled at a data sink (base station) for further analysis. An aggregate value is calculated at the data sink by applying the corresponding aggregate operator such as COUNT, MAX, MEDIAN or AVERAGE to the gathered data. **Figure 1** shows the general architecture for data aggregation model. There are two categories that are applied for data aggregation [1]: tree-based data aggregation [2]-[5] and cluster-based data aggregation schemes [6]-[8]. Recently, various data aggregation protocols are developed to reduce the data redundancy among the nodes in the network. It reduces the amount of energy consumed for the collection of data from the nodes. Hence, the communication cost is reduced. In the existing data aggregation process, the nodes are formed as a tree hierarchy fixed at a BS. The non-leaf nodes perform as data aggregators and gather the data from the child nodes of the tree structure before transmitting the collected data to the BS. Based on this process, the data are processed and collected at each hop on the communication path to the BS. Thus, the communication overhead in the network is largely reduced.

The hop-by-hop data aggregation formulates a new gateway to the false data injection attacks. The sensor nodes are deployed in the unattended and open environments. The malicious nodes can physically attack the nodes and retrieve the confidential information from the compromised nodes. It can also reprogram the compromised nodes into the malicious sensor. The compromised node reports a false aggregation result to their parent node in the tree structure, which causes huge variation in the final aggregation result from the actual measurement result. This attack has become more vulnerable during the collision of multiple compromised nodes. Data encryption is an essential factor in WSN when this type of sensor is subjected to different types of attacks. Without encryption, the malicious nodes can monitor and inject the false data into the sensor network. During the encryption process, the nodes should encrypt the data packets on the hop-by-hop basis. An aggregator node possesses the keys for all the sending nodes. It gathers all the received data and finally decrypts the collected data for transmitting the original data to the base station. Encryption technique can solve the security challenges in WSN, but the aggregation of data decides the overall network performance.

Our previous works [9] [10] also studied about the data aggregation protocol in WSN. But there are some security concerns in these protocols. To overcome the security issues in data aggregation, this paper presents a secure data gathering approach using the area based clustering with traffic flow and energy analysis. The CH is selected based on the cell center and energy among the nodes, which improves the network lifetime. The traffic

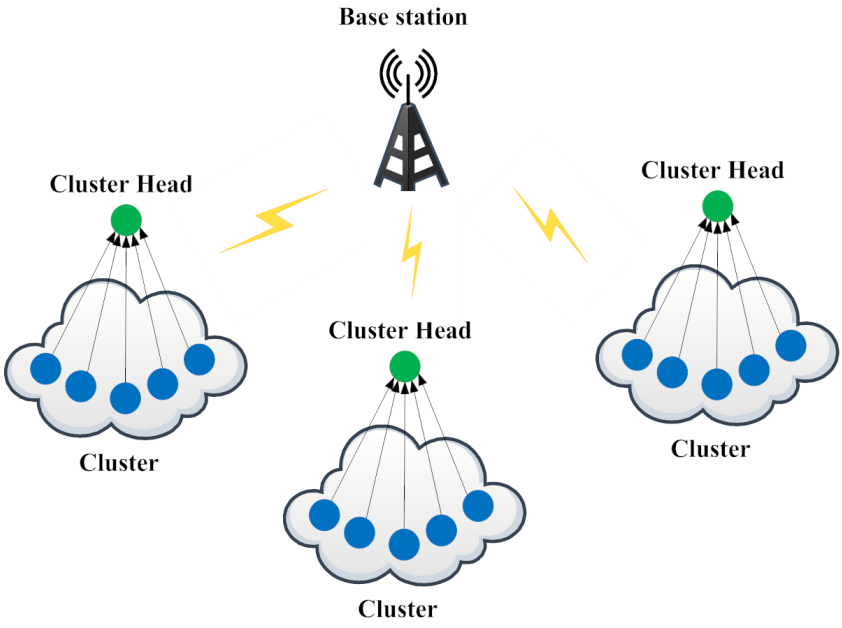

Figure 1. General architecture for data aggregation model in WSN.

behavior between the nodes is analyzed at the time of data transmission. If the traffic flow is dramatically increased, then it is recognized as the route including the malicious nodes that might include the false data. Then, it needs to be identified and eliminated from the path for secure data aggregation. Hence, this paper introduces a one-time key generation step to isolate the malicious nodes and add it into the block list.

The major contributions are described as follows:

1) Identification of false injection attack based on the traffic analysis at the time of route discovery process.
2) The one-time key generation step eliminates the malicious nodes from the network.
3) The efficiency of the proposed data aggregation scheme is evaluated by comparing it with the existing secure data aggregation scheme.

The remaining sections of the paper are structured as follows. Section 2 presents a brief outline about the conventional research works relevant to the secure data aggregation protocols. Section 3 describes the proposed SAC-TA scheme. Section 4 presents the results and comparative discussion with the proposed scheme and existing methods. Section 5 discusses the conclusion and future scope of the proposed work.

2. Related Works

This section describes the various research works related to the secure data aggregation techniques. *Mantri et al.* presented a bandwidth efficient cluster-based data aggregation method. It deliberates the network with heterogeneous nodes based on the energy and dynamic sink to gather the data packets. The optimality was attained by inter-cluster and intra-cluster aggregation on the randomly disseminated nodes with the adaptable data generation rate [11]. *Rout et al.* formulated an analytical approach for adaptive data aggregation. The network coding was used to improve the energy in a cluster based duty cycled WSN. The traffic within the cluster was reduced, and the energy efficiency of the bottleneck region was improved [12]. *Govind et al.* proposed an energy and trust-aware mobile agent migration protocol for data aggregation. A framework was introduced for trust validation to detect the malicious nodes [13].

Roy et al. discussed a secure data aggregation for WSN and formulated a synopsis diffusion approach in which the compromised nodes contribute the false sub-aggregate values. A lightweight verification algorithm was used to determine the aggregation that includes any false contribution [14]. *Li et al.* introduced an energy-efficient and secure data aggregation protocol. Accurate aggregation of the data received from the nodes was achieved and overhead on the sensors was reduced [15]. *Licheng et al.* proposed a discrete logarithm-based method to realize the fully adaptive or multiplicative homomorphism and secure data aggregation [16]. *Groat et al.* introduced a k-distinguishable privacy-preserving data aggregation in WSN. The sensitive measurements were complicated by hiding them among the set of disguise values. It was used to hide the wide range of aggregation functions [17]. *Huang et al.* proposed a secure data aggregation method to eliminate the redundant sensor readings without using the encryption. Also, it maintained the data privacy and security during transmission. This scheme was resilient to chosen plaintext attacks, man in the middle attacks and cipher-text only attacks [18].

Chein et al. implemented a secret data aggregation for data integrity in WSN. The base station can recover all sensing data, even it can be aggregated, and this property was called as recoverable [19]. *Ozdemir et al.* formulated a hierarchical secret data aggregation for WSN. The aggregation of encrypted packets with diverse encryption keys was permitted [20]. *Liu et al.* introduced a high energy-efficient and secure data aggregation for WSN. The communication overhead was reduced and the data accuracy was improved [21]. *Suat and Hasan et al.* proposed a combined approach for the detection of false data along with aggregation and transmission of confidential data in WSN. Also, the small-size Message Authentication Codes (MACs) were computed for the verification of data integrity at their pair levels. Data integrity verification on the encrypted data was performed rather than on the plain text, to enable the confidential data transmission. The simulation results have shown the reduction in the transmitted data rate by using the data aggregation and early false data detection [22].

Aldar et al. designed a security and privacy-preserving framework for data aggregation in WSN. The security model was sufficiently required for covering the most application scenarios and construction of data aggregation [23]. *Lu et al.* presented an optimal allocation strategy for data aggregation in WSN. A distributed algorithm was designed for the joint rate control and scheduling, based on the decomposition. The near optimal performance was achieved by using our approximate solution [24]. *Li et al.* designed a secure and energy-efficient data aggregation protocol with the identification of malicious data aggregator. Here, all the aggregation results were signed with the private keys of the data aggregators. Hence, it cannot be modified by others. Additionally, nodes located on each link utilize the pairwise shared key for secure transmission [25]. *Piyi et al.* introduced an effi-

cient data aggregation method with the constant communication overhead. The efficient actions against the passive and active privacy attacks were facilitated. The proposed method was proven to be robust for data loss and reduced transmission cost which was suited to be applied for large networks [26]. *Arumugam and Ponnuchamy* introduced an energy-efficient LEACH protocol for data gathering. The proposed protocol provided better packet delivery ratio and improved the network lifetime [27]. The limitations of the existing secure data aggregation technique and benefits of the proposed SAC-TA approach are shown in **Table 1**.

3. SAC-TA: A Secure Area Based Clustering for Data Aggregation Using Traffic Analysis

In this section, we proposed a novel secure area based clustering for data aggregation in WSN. In many applications, the physical occurrence is sensed by the sensors and reported the sensed information to the Base Station (BS). Data aggregation is used for solving the disintegration and overlapping problems occur during the data-centric routing in WSN. The data having same attribute is aggregated, when the data reach the same routing node located on the path back to the BS. The security issues, data confidentiality, and integrity are the vital factors for the data aggregation process, during the deployment of the sensor network in a hostile environment. **Figure 2** shows the overall flow of the proposed approach.

To reduce the energy utilization, the application should incorporate an in-network aggregation before it reaches the BS. The compromised nodes can perform malicious actions that affect the aggregation results. Before the detection of the malicious nodes, a secure aggregation protocol will safeguard the data packets and forward the data in a secured route. The sensor nodes are separated into different clusters, where each cluster has a CH. The CH acts as an aggregator for collecting the data received from the sensor nodes.

3.1. Network Model

Let us assume a WSN of "N" sensor nodes that are randomly distributed over the area M * M. The proposed network model is based on some assumptions.

- The sink with infinite energy level is located within the monitoring area.
- The sensor nodes are deployed within the specified area.
- The nodes can dynamically adjust the communication power according to the distance between the sink and other nodes.
- The communication between the nodes is reliable and symmetric.

3.2. Area Based Clustering

The area based clustering method uses the location information for each sensor nodes. The coordinates (x_i, y_i) for each sensor nodes are utilized to calculate the distance between two sensor nodes. The nodes are clustered to minimize the residual energy and maximize the lifetime of the node and network. The CH is selected based on

Table 1. Comparison of the limitations of the existing secure data aggregation technique and benefits of the proposed SAC-TA approach.

Existing Secure Data Aggregation Technique	Proposed SAC-TA Approach
1) The cluster head can be easily attacked by the malicious attacker. 2) The base station cannot ensure the correctness of the aggregate data send to it if the cluster head is compromised. 3) The power consumption at the nodes is increased, due to the transmission of several copies of the aggregate data to the base station. 4) Moreover, data aggregation results in the changes in the data received from the sensor nodes. 5) It is a really challenging task to provide authentication of data authentication along with data aggregation. 6) Due to these contradictory objectives, data aggregation and security protocols must be designed together, to enable aggregation of data without sacrificing security.	1) Identification of the false injection attack is performed based on the traffic analysis at the time of route discovery process. 2) One-time key generation step is introduced to eliminate the malicious nodes from the network. 3) It helps to identify the false data on the gathered data to provide the secure data gathering environment. 4) The proposed approach results in the reduction of the energy consumption rate to gather the data from diverse sensor nodes. 5) It automatically improves the residual energy, because half of the sensor nodes are alive at the end of the gathering process. 6) The data aggregation accuracy is improved by using the proposed technique. This results in the improvement of the network performance.

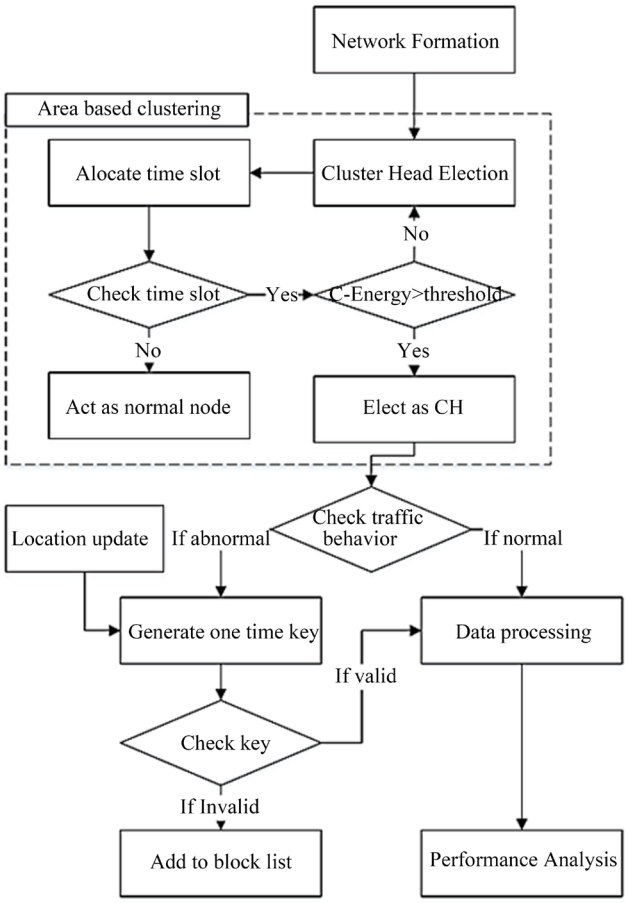

Figure 2. Flow of the proposed SAC-TA approach.

the cluster center and highest residual energy. The cluster center is estimated based on the minimum distance between the cluster node and the centroid. The data gathering is performed between the cluster members to CHs; and CHs to BS. The distance between the reference nodes is calculated using the following equation

$$d = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2} \tag{1}$$

Here, $\left(x_1, y_1\right)$ and $\left(x_2, y_2\right)$ are the coordinates of the reference node.

3.2.1. Node Deployment

The nodes are deployed around the sink using the two-dimensional Gaussian distribution. The deployed nodes are battery powered with the initial energy E_a. σ denotes the standard deviation for the "*x*" and "*y*" dimensions of the node. The Gaussian distribution is defined as

$$f\left(x, y\right) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \tag{2}$$

3.2.2. Cluster Formation

After the deployment of the nodes, the cluster is formed. A non-cluster node considers the size of the cluster "*i*" denoted as $S^k\left(i\right)$ to decide which cluster to join in the k^{th} round. In most of the cases, the non-cluster nodes depend on the signal strength of the CHs to decide about the cluster to join based on the assumption about the uniform distribution of the sensor nodes in the monitoring region. But this is not possible in the practical scenarios. The energy dissipation of the CH of a relatively big cluster for aggregating data from the nodes is highly greater than the CH in a smaller cluster. This leads to the imbalance of energy consumption in the CHs and re-

sults in the adverse impacts in the network lifetime. A new criterion for the cluster formation is introduced in this work, for achieving better load balancing among the CHs. When a non-cluster node decides to join a cluster, it considers the residual energy of the CH and size of the cluster. This is defined as

$$V_a(i,k) = \beta E_i^k + (1-\beta)\left(\frac{N}{C} - S^k(i)\right) \tag{3}$$

where, E_i^k is the energy of the node "*I*" at the k^{th} round; β is a factor that is used for adjusting the impact of the size of the cluster and E_i^k. When a sensor node is located far away from all CHs, the above equation is used to choose its CH having high residual energy and lesser cluster size than the other CHs. This ensures better load balancing to the CHs, to improve the overall lifetime and better performance of the WSN.

3.2.3. Optimal CH Selection

The CH is chosen based on the distance measurement approach. The node located at the center of the cell is selected as the CH, due to the minimum distance between the center node and other cluster member nodes. The residual energy of the node is the main condition for selecting the CH. The nodes forming a cluster are disseminated within a small region, and the sink is located far away from the nodes. The data aggregation results in the energy saving effect, since the nodes require only a minimum amount of energy for transmitting the data directly to the CH, rather than sending the data to the sink. So, the nodes located closer within a cluster are inferred due to the minimum amount of energy consumption. Hence, selection of the CH is performed based on the residual energy amount of the sensor nodes.

Let us consider a WSN with "*N*" number of sensor nodes. $D^k(i)$ is defined as the concentration degree of the node "*i*", for sensing the number of sensor nodes during the k^{th} round.

$W(a,k)$ is defined as the selection weight of the node "*a*" in the k^{th} round.

$$\gamma = \frac{1}{1+\rho}, \quad \text{where } \rho = \frac{E_a^k}{E_a} \tag{4}$$

$$W(a,k) = \gamma \frac{E_a^k}{E^k} + (1-\gamma)\frac{D^k(a)}{\left(\frac{N}{C}\right)} \tag{5}$$

where "*C*" is the number of clusters; "E_a" is the initial energy of the node "*a*"; "E^k" is the average residual energy of the network in the "k^{th}" round; ρ is the residual energy of the node "*a*" in the k^{th} round; γ is an adaptive factor to adjust the influence of the residual energy of the node and concentration degree to the selection weight. The adaptive factor increases gradually with the reduction in the residual energy, to adapt to the decrease in the number of effective nodes in the WSN.

The CH aggregates and compresses data before transmitting the data to the sink. The optimal probability of a node being selected as a CH is defined as the function of the spatial density, during the uniform distribution of the nodes over the sensor field. Optimal clustering is achieved, when the energy consumption across all the sensor nodes is low. Let us consider that the distance between the nodes to the sink or CH is $\le D_0$. Thus, the energy dissipation in the CH, during transmission of "B" bit message over a distance "*D*" in a round is given as

$$E_{ch} = \left(\frac{n}{k} - 1\right) * B * E_D + \frac{n}{k} * B * E_P + B * E_D + B * \varepsilon_{fs} * D_{BS}^2 \tag{6}$$

where, E_D is the energy dissipated per bit; "*D*" is the distance between the sender and receiver nodes; "*k*" is the number of clusters; ε_{fs} is the energy dissipation amount of the transmitter amplifier circuit; E_P is the processing cost of a bit report to the sink or base station; and D_{BS} is the average distance between the sink and CH. The energy used in a non-CH node is equal to

$$E_{nonch} = B * E_D + B * \varepsilon_{fs} * D_{CH}^2 \tag{7}$$

where D_{CH} is the average distance between the cluster member and its *CH*. If the nodes are uniformly distributed, the D_{CH}^2 is defined as

$$D_{CH}^2 = \int_0^{x\max} \int_0^{y\max} \left(\left(x^2 + y^2 \right) * \rho(x,y) \right) \mathrm{d}x\mathrm{d}y = \frac{M^2}{2\pi k} \tag{8}$$

where, $\rho(x,y)$ is the node distribution and "M^2" is the area in which the nodes are distributed. The total energy dissipation in the network is

$$E_T = B * \left(2 * n * E_D + n * E_P + \left(k * D_{BS}^2 + n * D_{CH}^2 \right) \right) \tag{9}$$

The optimal number of *CH*s is estimated by differentiating the total energy dissipation with respect to "*k*" and equating it to zero. The probability of a node to become a *CH* is computed as

$$P_{CH} = \frac{\sqrt{\frac{n}{2\pi}} \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}} \frac{M}{D_{BS}^2}}}{n} \tag{10}$$

where, ε_{mp} is the energy dissipated by the transmitter amplifier circuit. If the residual energy level of the node is greater than the threshold value, optimal CH selection is performed. Otherwise, the residual energy calculation process is repeated. After the selection of the CH, data aggregation is performed and data is forwarded to the BS. If the cluster construction is not in the optimal manner, there is an exponential increase in the total energy consumption of the network, when the number of constructed clusters is greater than or less than the optimal number of clusters.

The CH is used for authenticating the nodes located inside the cluster. The CH validates the sensor nodes in the cluster, shares the secret key to the nodes, validates the data received from the nodes and forwards the data to the BS. The CH forwards the data received from the cluster member nodes after validating the signature of the data. If the sensor node is compromised by the attacker, the data received from that node is sent to the CH. The CH verifies the signature of the data received from that node. If there is a mismatch in the data signature, then the CH detects the node as a compromised node and stops the data communication with that node. After the successful transmission of data from the CH to the sink, the selected CH will lose energy for transmission. Again, a new CH having highest residual energy and minimum distance with other cluster member nodes is chosen. **Figure 3** shows the flow diagram of the optimal CH selection process.

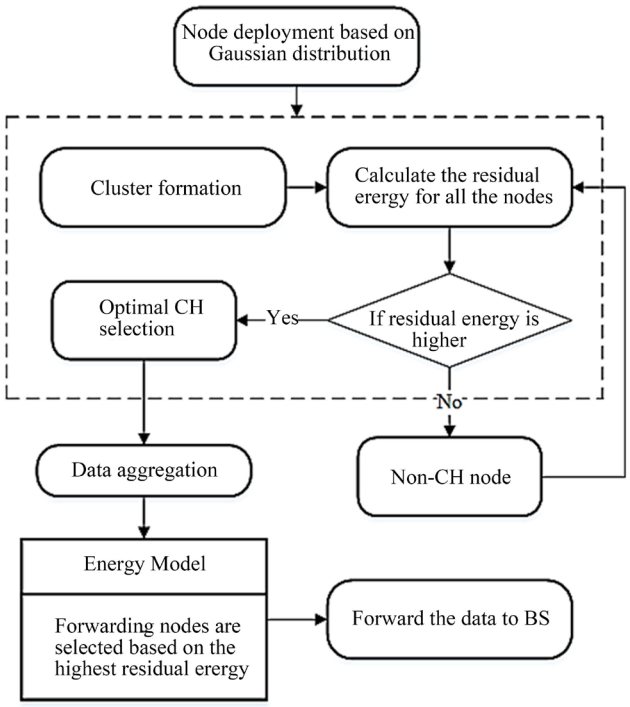

Figure 3. Flow diagram of the optimal CH selection process.

CH Election Procedure

Assume that $X = \{x_1, x_2, \cdots, x_n\}$ be the set of nodes and $C = \{c_1, c_2, \cdots, c_n\}$ be the set of clusters.

Step 1: Estimate the distance between each node with the cluster center.
Step 2: Allocate the time slot *t*.
Step 3: Estimate the maximum residual energy for each node.
Step 4: Select the node having maximum residual energy and located closer to the cluster center and choose that node as the CH.
Step 5: Repeat the above (1-3) steps until all clusters have CHs.

Each node maintains the energy and distance between the cluster centers in the routing table. For the predefined time slot *t*, the CH is selected and data aggregation process takes place. Through the periodic reassignment of the CH role to various nodes, the problem of single point of failure during the depletion of node energy is prevented. It is a really critical task to avoid the failure of the nodes caused by the early depletion of energy, to ensure high network lifetime. After the selection of the CH, the traffic behavior is analyzed. If the traffic is found as normal, the data transmission process is performed. During the detection of the abnormal traffic, the one-time key is generated and checked. If the key is valid, the data transmission is executed. Otherwise, the node is added to the block list.

3.3. Traffic Analysis

The amount of traffic between the sensor nodes is estimated for a specific period. During this period, the amount of traffic emitted from all other regions and the average size of the sent packets are recorded. If the difference between the estimated and actual values of the traffic is more than a threshold value, the probability of occurrence of attack is detected. This includes three main steps:

1) Assess the amount of traffic generated by the nodes located adjacent to the suspected node.
2) Estimate the traffic amount and compare it with the actual values.
3) Comparing the average size of the sent packets with the pre-stored information.

If the increase in the generated traffic occurs as a result of detection of an event in the area covered by the sensor node, the neighbor nodes should have experience the same increase. Therefore, the normal and misbehavior of the nodes is properly differentiated. If the adjacent nodes do not show noticeable increase in the traffic volume, the possibility of occurrence of an attack is detected more precisely. The packet size comparison is performed based on the concept that if the attacker tries to improve the packet sending rate, power consumption of the nodes will be high. Hence, the misbehaving node is identified based on the traffic analysis.

3.4. One-Time Key Generation

In the symmetric key generation process, a static secret key is generated by using the cipher text. During the symmetric key cryptography, if there is N number of nodes, then there is a need to generate $(N-1)$ keys in the network. If the value of N is large, it requires more memory space for storing the large key values. The public key cryptography system requires huge computation power, but the sensor has less processing power. The public key cryptography system is not efficient in the WSN applications. Hence, the dynamic key generation is mostly preferred for the security purpose in WSN, since the dynamic keys are single-time usable symmetric cryptographic keys for forming a sequence of keys. The probability of breaking a dynamic key is low. The dynamic key management does not involve a central key controller such as BS or third party in the rekeying process of the nodes. The key management is handled by the dynamically assigned key controllers. The cryptographic keys are provided securely, while preventing the activities of the attacker nodes inside a network. During the detection of a compromised sensor node, the current secret key of the compromised node is canceled and a new key is generated. This new key is distributed to its associated nodes, except the compromised node. Moreover, it is highly desirable for a dynamic key management scheme to maintain the security of the keys and avoid collusion between the compromised nodes and newly joined nodes. The dynamic key management schemes prevent a single point of failure and ensure high network scalability. However, these schemes are highly prone to the design errors because the compromised nodes can participate in the node removal process.

3.4.1. Key Generation

Several security parameters are defined during the construction of the Secure One-Time (SOT) key [28]. For

signing B-bit messages, initially 'n' and 'k' are selected such that $\binom{n}{k} \geq 2^B$ and the security parameter "p" and one-way hash function "H_F" that operates on p-bit strings is chosen. The p-bit string $(s_1, s_2, \cdots, s_n)$ are generated randomly to generate the public key. Let $O_k = H_F(S_j)$ for $1 \leq j \leq n$. The public key is $PK = (k, O_1, O_2, \cdots, O_n)$ and private key is $SK = (k, s_1, s_2, \cdots, s_n)$.

3.4.2. Signature Generation

Let $H = \text{hash}(M)$ for signing a message "M" with the secret key. Then, "H" is split into "k" substrings $H = h_1, h_2, \cdots, h_k$. Finally, each H_z is interpreted as an integer for $1 \leq z \leq k$. The resulting signature is $\rho = (G_{i1}, G_{i2}, \cdots, G_{ik})$.

3.4.3. Signature Verification

The signature verification is similar as the signature generation process. If the verifier has the message "M", signature $\rho = (G'_{i1}, G'_{i2}, \cdots, G'_{ik})$ and public key $PK = (k, O_1, O_2, \cdots, O_n)$. The signature is accepted only if $H_F(S'_z) = H_{iz}$, for each "z". Otherwise the signature is rejected.

In this scheme, the public key component can be used for multiple times. Signature generation requires only a single call to the hash function. But, the verification process requires "k" calls to the hash function. The main advantage of our scheme is the smaller signature size.

3.4.4. Key Generation Procedure

1) Each node chooses "n" number of secret key components $S_k (k = 1, \cdots, n)$ at random.
2) Each node creates a "m" number of hash chain of length "l".
3) The public key components are obtained through an one-way function $O_k = H(S_k)$. H denotes the hash function.

3.4.5. Public Key Handling

Each node generates a set of one-time keys that are distributed locally among the nodes. Since the one-time keys are valid once or used for a limited period of time, there is a need to update the one-time key for the nodes. Initially, the public key called as Initial key is distributed safely to the nodes during the system setup. This guarantees that each neighboring node holds the authentic copy of the public key. When, a new node enters into the network, it receives the public key and broadcasts to its neighbors. Hence, successive one-time public keys are distributed efficiently through the periodic broadcasting of the Hello message.

The first secret key SK_1 is defined as

$$\left\langle k, H^m(S_1), H^m(S_2), H^m(S_3), \cdots, H^m(S_n) \right\rangle \tag{11}$$

and the corresponding public key is

$$\left\langle k, H^{m+1}(S_1), H^{m+1}(S_2), H^{m+1}(S_3), \cdots, H^{m+1}(S_n) \right\rangle \tag{12}$$

3.4.6. System Setup

If a node enters into the network, it is notified about the security parameters in the network. Then, the node selects its secret key components and creates a hash chain (**Figure 4**).

3.4.7. Route Discovery

If the source node (I) initiates a route discovery to a certain destination node, it simply generates a signature over the Route Request (RREQ). When the CH1 receives the RREQ, the signature of the source node is initially verified. If the signature is found to be correct, the CH1 hashes the received message and generates its own signature by using the SOT key generation scheme. The whole message is retransmitted to the next CH2. Once the CH2 receives this double signed RREQ, it initially verifies the previous hop using the public key of CH1 received through the Hello messages. If the one-time signature is found correct, CH2 hashes the signature again, creates a signature over the hash for replacing the signature of CH1. Then, this new message is broadcasted to the BS, only if both the signatures are correct. These operations are repeated until the RREQ reaches the desti-

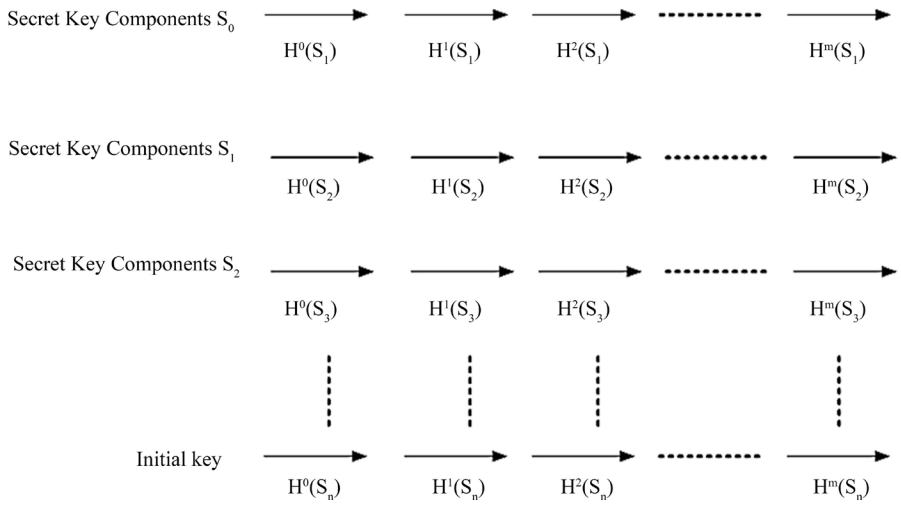

Figure 4. Hash chain of secret key components.

nation node (BS). When RREQ reaches the BS, the BS performs same verification operations as each CH. Then, a route reply (RREP) is generated and signed in a same manner as RREQ. Each CH will transmit the RREP to the source node through the reverse route and same operations are performed along the route.

In our proposed work, the SOT key is generated dynamically based on the source ID, random number and position coordinates of the nodes. The main idea to generate the one-time key is to avoid distribution of the long-term shared cryptographic keys. Due to the randomness, the one-time key is unbreakable by the compromised nodes. Hence, the secret key is generated each time during transmission of packet between the nodes. The one-time key generation is used to prevent the compromised node or third party from extracting the key. Even if the source node is a compromised node, the one-time key is not hacked by the source node. The security of the one-time key depends on the randomness property.

As the sensor nodes report their data, the direction of the data movement is habitually towards the BS. This asymmetric communication design can assist a malicious node in tracking down the location of the BS. It can result in malicious launching the serious attacks on the BS and eventually settling down the entire network. There are many ways to track the location of BS:

1) If the malicious node can understand the information of the packet being transmitted, then the malicious nodes can correlate the packets that are transmitted towards the BS. It will permit the malicious node to follow the direction of these packets towards the locality of the BS, which leads to discovery/jamming and destruction of the BS.
2) If there is a time correlation between when a node receives or forwards a packet, a malicious node can use the time correlation to estimate the direction towards the BS.

Hence, it is necessary to block the malicious nodes from the data aggregation and data transmission process. The misbehavior of the nodes is checked by analyzing the traffic flow between the nodes. Before the key generation process, the location of the node is updated. The key generation step is initiated, if the traffic flow is found to be abnormal. If the one-time key of the encrypted packet is invalid, the current packet is dropped, and the node is added to the block list. Data processing is performed, only if the one-time key is found to be valid. After the BS receives the aggregated data from all the child nodes, it decrypts and validates the signature. It estimates the final aggregation result just like the operation of the normal intermediate node. The final aggregation result stored in the BS. The following algorithm is proposed to block the malicious nodes from the network.

The SOT key is generated based on the key generation scheme adopted in [28]. Here, ID_S denotes the source ID, RP is the random prime number generated for CH, (Pos_x, Pos_y) is the node coordinates and OT_K_i denotes the SOT key generated for the corresponding nodes. The Exclusive-OR operation is performed in the SOT key generation process. **Figure 5** shows the flowchart of the onetime key generation algorithm.

4. Performance Analysis

In this section, the performance of the proposed Secure Area Based Clustering for Data Aggregation Using

One-time Key Generation Algorithm

Input: ID_S, Pos_x, Pos_y
Output: Block Misbehaving Nodes
Begin
//Check misbehavior analysis
Analyze the traffic flow
If traffic flow is abnormal
//Initiate the key generation step
$OT\_K_i \rightarrow ID_S \oplus RP \oplus Pos_x \oplus Pos_y$
// Check authentication
Receive the encrypted packet
If OT_K is invalid
Drop the current packet
Add to block list
Else
Perform data processing
End if
End if
End

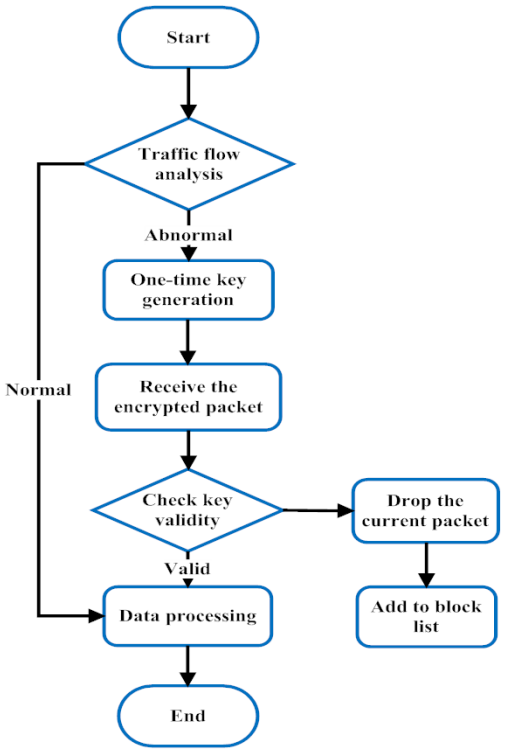

Figure 5. Flowchart for the one-time key generation algorithm.

Table 2. Simulation parameters.

Parameters	Values
Simulation Area	500 × 500 m
Total Number of Nodes	200
Traffic Sources	Constant Bit Rate
Simulation Time	100 s
Total Number of Packets Sent	600

Traffic Analysis (SAC-TA) is analyzed and compared with the existing Secure Data Aggregation Technique (SDAT) [29]. For analysis, there are 200 sensor nodes are deployed, and the results are taken for the simulation time up to 100 s and tested in NS2 tool. **Table 2** presents the simulation parameters.

4.1. Energy Analysis for Data Aggregation

The energy consumption of the sensor network over a period can be estimated as follows:

- Energy spent for sensing the channel.
- Energy spent to send the packets from the sensor nodes to the CHs.
- Energy spent to receive the gathered data from the CHs to the BS.
- Average energy consumption and remaining energy for the entire data gathering process.

Scenario 1: There are totally 600 packets are gathered by the CHs among various sensor nodes. The CHs collect the packets and validate it to analyze the gathered information containing any malicious information. **Figure 6(a)** shows the average energy consumption for a number of packets to be gathered from the sensor nodes to the corresponding CHs. The results are compared with the existing Secure Data Aggregation Technique (SDAT). The proposed SAC-TC results in the lesser energy consumption than the existing SDAT.

Scenario 2: The BS collected the gathered data from the CHs. If any misbehaving activities are identified, then a one-time key is generated to identify whether the packet is included with the false injected data or it includes the secured gathered data. After the process gets completed, the remaining energy for the corresponding packet reception is calculated to determine the energy utilization rate of each node. The energy efficiency of the node is high if the remaining energy value is high.

Figure 6(b) depicts the results for the average remaining energy for the proposed SAC-TA with the existing

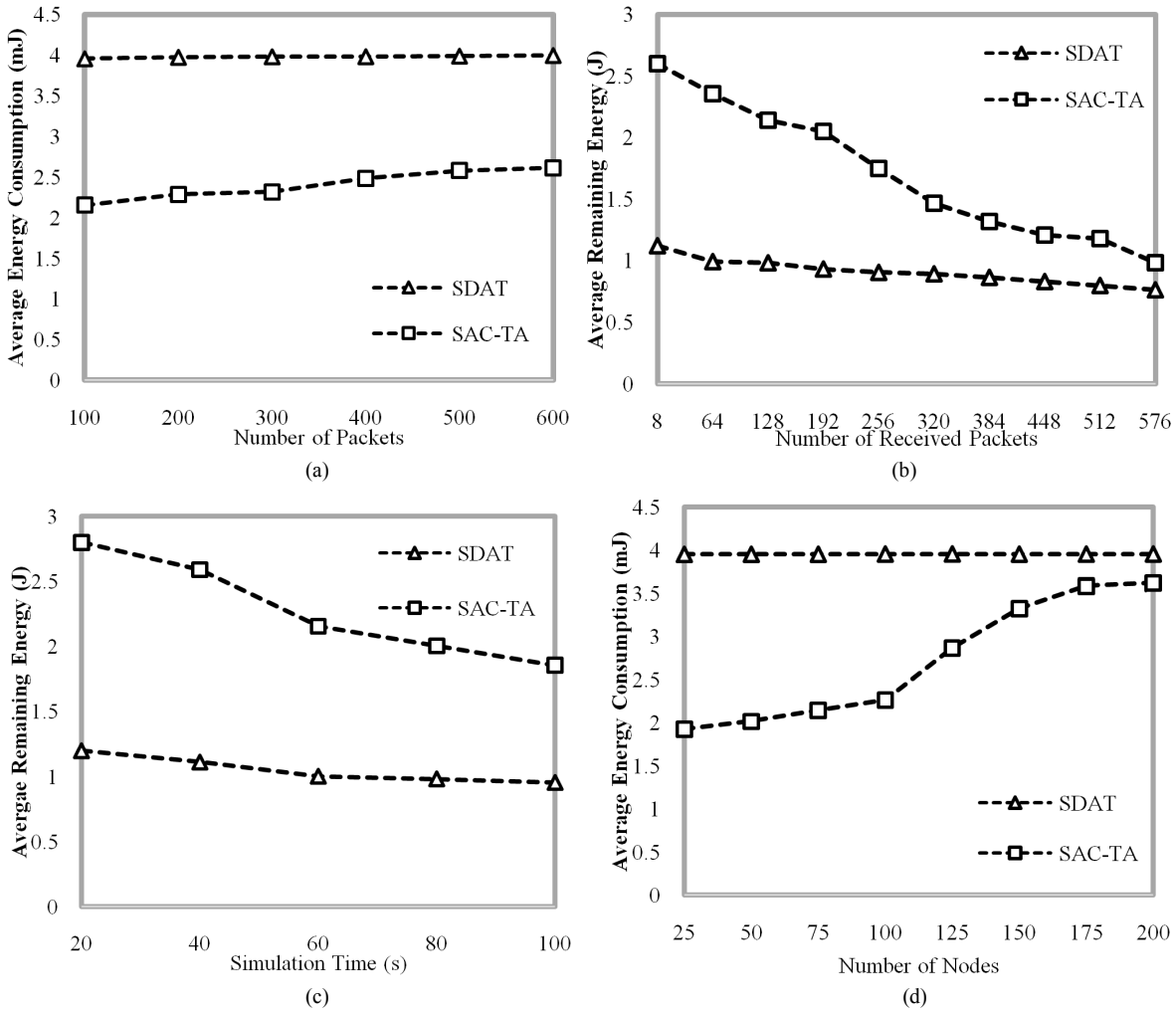

Figure 6. Energy Analysis for the proposed SAC-TA and the existing SDAT (a) Average energy consumption for sending the packets to CHs, (b) Average remaining energy after receiving the packets, (c) Average remaining energy across the simulation time (100 s) and (d) Average energy consumption across varying the number of nodes.

SDAT approaches. The proposed approach results higher remaining energy than the existing SDAT.

Scenario 3: The process is evaluated for simulation time up to 100 s. The remaining energy is noted for every 20 seconds, and it examined with the existing method. We are using the highest energy nodes as the CHs to aggregate the packets. Hence, the remaining energy is obviously high, which helps to increase the network lifetime. **Figure 6(c)** shows the average remaining energy for the proposed and existing method. The proposed SAC-TA results in the higher residual energy than the existing method SDAT.

Scenario 4: In our analysis, 200 sensor nodes are taken for data aggregation process. The average energy consumption is examined for varying the nodes. **Figure 6(d)** shows the comparison result for the proposed SAC-TA with the existing SDAT regarding the variation in the number of nodes up to 200. The results show that the proposed approach takes lesser energy consumption than the existing method.

4.2. Time Taken for Secure Data Aggregation

The time taken to complete the secure data aggregation refers to the time taken to transfer a packet across a sensor network from the sensor to CHs and CHs to BS. The equation to estimate the End-to-End (E2E) delay is defined as follows,

$$\text{E2E-delay} = N\left[d_t + d_{pb} + d_{pr}\right] \tag{13}$$

Here, d_t denotes the transmission delay; d_{pb} is the propagation delay; d_{pr} denotes the processing delay; and "N" denotes the number of links. The CH is selected based on the predefined time slot, to improve the system performance without causing any unwanted link failure. Hence, it automatically reduces the E2E delay for data aggregation.

Figure 7 shows the E2E delay for the proposed SAC-TA and existing SDAT which is quite lesser than the existing method.

4.3. False Data Detection

The malicious node hacks the network and stole the gathered data. The false data is injected by the malicious nodes to intrude the network performance. The malicious node needs to be isolated and removed from the network. Here, the traffic flow is analyzed and the abnormal behavior of the sensor nodes is identified. A one-time key generation procedure is introduced to validate the encrypted information about the gathered data. It helps to identify the false data on the gathered data to provide the secure data gathering environment.

Figure 8 shows the false data detection rate for the proposed SAC-TA with the existing method. The proposed approach results in better detection rate than the existing method for varying the data transmission.

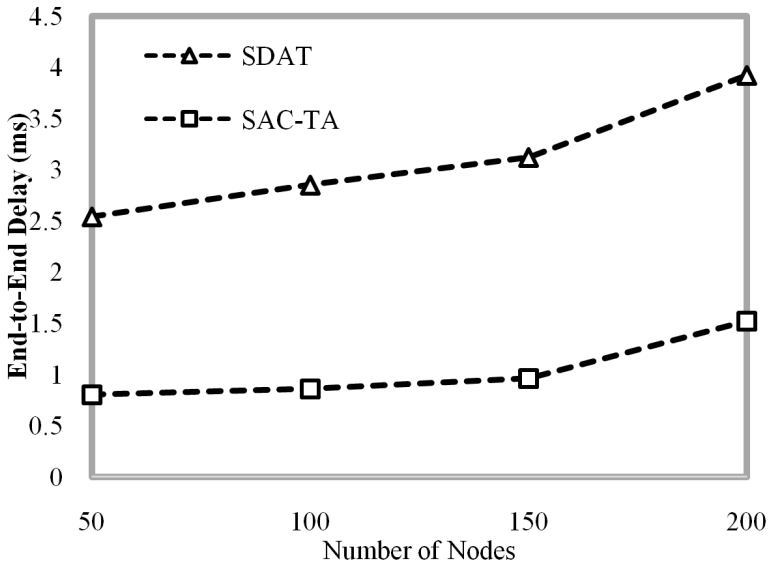

Figure 7. Comparison of end-to-end delay for SCA-TA and SDAT.

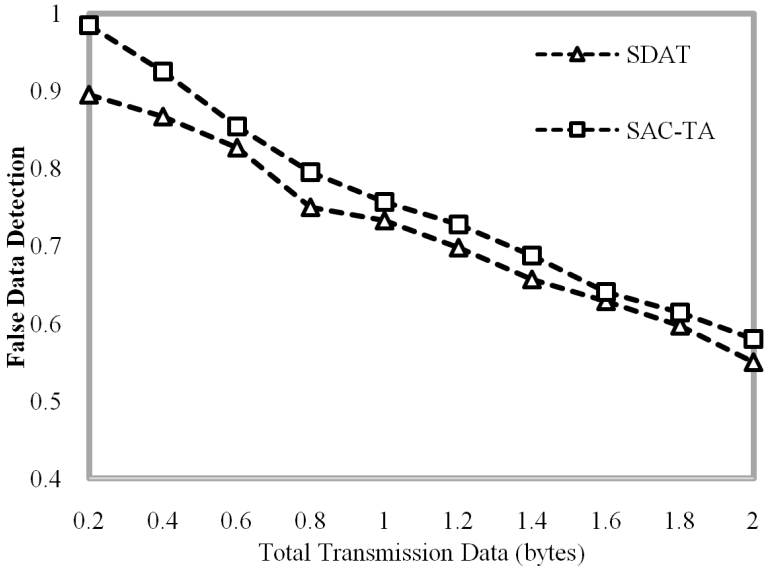

Figure 8. False data detection across the transmission data for SAC-TA and SDAT.

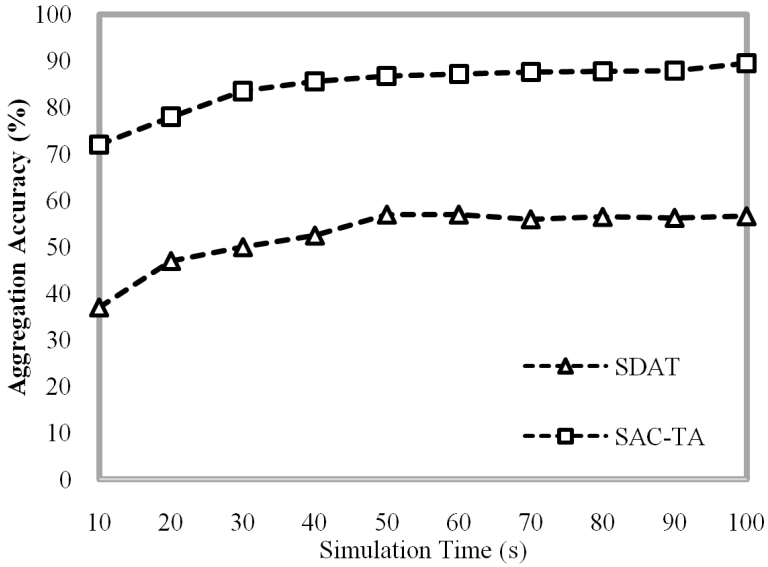

Figure 9. Aggregation accuracy for SAC-TA and SDAT.

4.4. Data Aggregation Accuracy

The data aggregation accuracy is an important factor to predict the successful system performance. The accuracy measure is taken for varying the simulation time up to 100 s. It shows that the process gets completed at the final stage and results in the accuracy of about 89.5% to receive the gathered data at the BS. Totally 600 packets are transmitted by the sensor nodes to CHs, and approximately 576 packets are gathered and collected by the BS. **Figure 9** shows the aggregation accuracy for the proposed approach with the existing method. The comparative analysis of the resultant aggregation accuracy values for the proposed SAC-TA approach and existing SDAT method are shown in **Table 3**. It shows the proposed SAC-TA approach achieves better accuracy rate than the existing SDAT method.

4.5. Alive Nodes Analysis

The number of alive nodes is calculated for each round to find the energy efficiency of the network. After com-

Table 3. Comparative analysis for aggregation accuracy.

Simulation time (s)	SAC-TA	SDAT
10	72	37
20	78	47
30	83.5	50
40	85.6	52.5
50	86.7	57
60	87.23	57
70	87.56	56
80	87.76	56.5
90	87.9	56.3
100	89.5	56.65

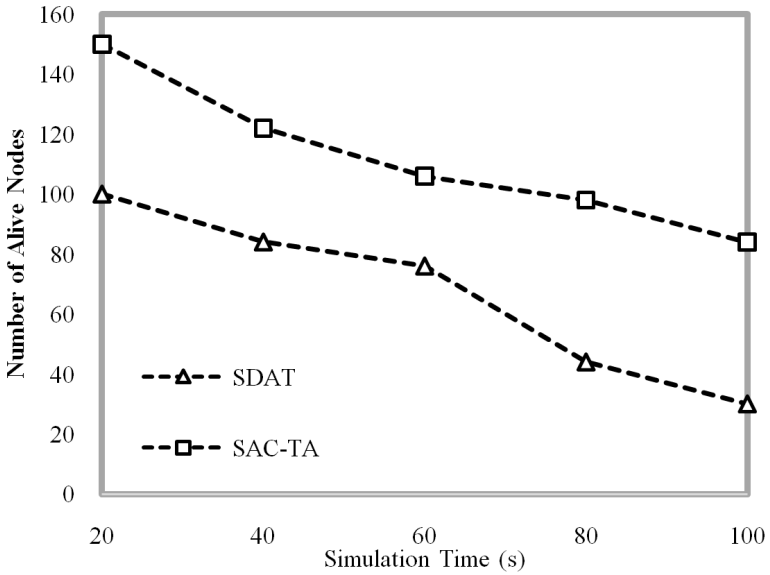

Figure 10. Number of alive nodes after data aggregation completion.

pletion, the proposed approach contains approximately 84 alive nodes, whereas the existing holds only 30 alive nodes. **Figure 10** shows the comparison of alive nodes for proposed and existing methods. The proposed approach holds higher alive nodes than the existing method. The proposed SAC-TA method improves the network lifetime. Hence, it is clearly evident that the proposed approach achieves better performance than the existing methods.

5. Conclusion and Future Work

In this paper, a SAC-TA approach is presented for data aggregation in WSN. The false injection attack is identified based on the traffic analysis at the time of route discovery process. One-time key generation algorithm is introduced to eliminate the malicious nodes from the network. The efficiency of the proposed data aggregation scheme is evaluated by comparing it with the existing secure data aggregation scheme (SDAT). It results in the lower energy consumption to gather the data from diverse sensor nodes. It automatically improves the residual energy, because half of the sensor nodes are alive at the end of the gathering process. The proposed approach also achieves low end-to-end delay and better false detection rate and aggregation accuracy, when compared to the existing method. In the future work, the behavior of the WSN on internal attackers is investigated, and the SAC-TA scheme is extended to resist such attacks. This will guarantee the secure data aggregation under the presence of attacks.

References

[1] Wang, W., Wang, B., Liu, Z., Guo, L. and Xiong, W. (2011) A Cluster-Based and Tree-Based Power Efficient Data Collection and Aggregation Protocol for Wireless Sensor Networks. *Information Technology Journal*, **10**, 557-564. http://dx.doi.org/10.3923/itj.2011.557.564

[2] Hakoura, B. and Rabbat, M.G. (2012) Data Aggregation in Wireless Sensor Networks: A comparison of Collection Tree Protocols and Gossip Algorithms. *The* 25*th IEEE Canadian Conference on Electrical & Computer Engineering* (*CCECE*), Montreal, 29 April-2 May 2012, 1-4. http://dx.doi.org/10.1109/CCECE.2012.6335001

[3] Incel, O.D., Ghosh, A. and Krishnamachari, B. (2011) Scheduling Algorithms for Tree-Based Data Collection in Wireless Sensor Networks. Theoretical Aspects of Distributed Computing in Sensor Networks, 407-445. http://dx.doi.org/10.1007/978-3-642-14849-1_14

[4] Hoang, D.C., Kumar, R. and Panda, S.K. (2012) Optimal Data Aggregation Tree in Wireless Sensor Networks Based on Intelligent Water Drops Algorithm. *IET Wireless Sensor Systems*, **2**, 282-292.

[5] Liu, G., Huang, L., Xu, H., Xu, X. and Wang, Y. (2013) Energy-Efficient Tree-Based Cooperative Data Aggregation for Wireless Sensor Networks. *International Journal of Sensor Networks*, **13**, 65-75. http://dx.doi.org/10.1504/IJSNET.2013.053720

[6] Ma, Y., Guo, Y., Tian, X. and Ghanem, M. (2011) Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks. *IEEE Sensors Journal*, **11**, 641-648. http://dx.doi.org/10.1109/JSEN.2010.2056916

[7] Yuea, J., Zhang, W., Xiao, W., Tang, D. and Tang, J. (2012) Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks. *Procedia Engineering*, **29**, 2009-2015. http://dx.doi.org/10.1016/j.proeng.2012.01.253

[8] Jung, W.-S., Lim, K.-W., Ko, Y.-B. and Park, S.-J. (2011) Efficient Clustering-Based Data Aggregation Techniques for Wireless Sensor Networks. *Wireless Networks*, **17**, 1387-1400. http://dx.doi.org/10.1007/s11276-011-0355-6

[9] Saminathan, A.G. and Karthik, S. (2013) Development of an Energy-Efficient, Secure and Reliable Wireless Sensor Networks Routing Protocol Based on Data Aggregation and User Authentication. *American Journal of Applied Sciences*, **10**, 832-843. http://dx.doi.org/10.3844/ajassp.2013.832.843

[10] Saminathan, A.G., Karthik, S. and Post, S. (2013) DAO-LEACH: An Approach for Energy Efficient Routing Based on Data Aggregation and Optimal Clustering in WSN. *Life Science Journal*, **10**, 380-389.

[11] Mantri, D.S., Prasad, N.R. and Prasad, R. (2014) Bandwidth Efficient Cluster-Based Data Aggregation for Wireless Sensor Network. *Computers & Electrical Engineering*, **41**, 256-264.

[12] Rout, R.R. and Ghosh, S.K. (2014) Adaptive Data Aggregation and Energy Efficiency Using Network Coding in a Clustered Wireless Sensor Network: An Analytical Approach. *Computer Communications*, **40**, 65-75. http://dx.doi.org/10.1016/j.comcom.2013.11.003

[13] Govind, P., Gupta, M.M. and Garg, K. (2014) Energy and Trust Aware Mobile Agent Migration Protocol for Data Aggregation in Wireless Sensor Networks. *Journal of Network and Computer Applications*, **41**, 300-311. http://dx.doi.org/10.1016/j.jnca.2014.01.003

[14] Roy, S., Conti, M., Setia, S. and Jajodia, S. (2012) Secure Data Aggregation in Wireless Sensor Networks. *IEEE Transactions on Information Forensics and Security*, **7**, 1040-1052. http://dx.doi.org/10.1109/TIFS.2012.2189568

[15] Li, H., Lin, K. and Li, K. (2011) Energy-Efficient and High-Accuracy Secure Data Aggregation in Wireless Sensor Networks. *Computer Communications*, **34**, 591-597. http://dx.doi.org/10.1016/j.comcom.2010.02.026

[16] Wang, L.C., Wang, L.H., Pan, Y., Zhang, Z.H. and Yang, Y.X. (2011) Discrete Logarithm Based Additively Homomorphic Encryption and Secure Data Aggregation. *Information Sciences*, **181**, 3308-3322. http://dx.doi.org/10.1016/j.ins.2011.04.002

[17] Groat, M.M., He, W. and Forrest, S. (2011) KIPDA: K-Indistinguishable Privacy-Preserving Data Aggregation in Wireless Sensor Networks. *Proceedings IEEE INFOCOM*, Shanghai, 10-15 April 2011, 2024-2032. http://dx.doi.org/10.1109/infcom.2011.5935010

[18] Huang, S.-I., Shieh, S. and Tygar, J.D. (2010) Secure Encrypted-Data Aggregation for Wireless Sensor Networks. *Wireless Networks*, **16**, 915-927. http://dx.doi.org/10.1007/s11276-009-0177-y

[19] Chien-Ming, C., Yue-Hsun, L., Ya-Ching, L. and Hung-Min, S. (2012) RCDA: Recoverable Concealed Data Aggregation for Data Integrity in Wireless Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, **23**, 727-734. http://dx.doi.org/10.1109/TPDS.2011.219

[20] Ozdemir, S. and Xiao, Y. (2011) Integrity Protecting Hierarchical Concealed Data Aggregation for Wireless Sensor Networks. *Computer Networks*, **55**, 1735-1746. http://dx.doi.org/10.1016/j.comnet.2011.01.006

[21] Liu, C.-X., Liu, Y., Zhang, Z.-J. and Cheng, Z.-Y. (2013) High Energy-Efficient and Privacy-Preserving Secure Data

Aggregation for Wireless Sensor Networks. *International Journal of Communication Systems*, **26**, 380-394. http://dx.doi.org/10.1002/dac.2412

[22] Ozdemir, S. and Cam, H. (2010) Integration of False Data Detection with Data Aggregation and Confidential Transmission in Wireless Sensor Networks. *IEEE/ACM Transactions on Networking*, **18**, 736-749. http://dx.doi.org/10.1109/TNET.2009.2032910

[23] Chan, A.C.-F. and Castelluccia, C. (2011) Security Framework for Privacy-Preserving Data Aggregation in Wireless Sensor Networks. *ACM Transactions on Sensor Networks*, **7**, 1-45. http://dx.doi.org/10.1145/1921621.1921623

[24] Su, L., Gao, Y., Yang, Y. and Cao, G. (2011) Towards Optimal Rate Allocation for Data Aggregation in Wireless Sensor Networks. *Proceedings of the* 12*th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Paris, 16-20 May 2011, No. 19. http://dx.doi.org/10.1145/2107502.2107528

[25] Li, H., Li, K., Qu, W. and Stojmenovic, I. (2014) Secure and Energy-Efficient Data Aggregation with Malicious Aggregator Identification in Wireless Sensor Networks. *Future Generation Computer Systems*, **37**, 108-116. http://dx.doi.org/10.1016/j.future.2013.12.021

[26] Pi, Y.Y., Zhen, F.C., Xiao, L.D. and Zia, T.A. (2011) An Efficient Privacy Preserving Data Aggregation Scheme with Constant Communication Overheads for Wireless Sensor Networks. *IEEE Communications Letters*, **15**, 1205-1207. http://dx.doi.org/10.1109/LCOMM.2011.092911.111598

[27] Arumugam, G.S. and Ponnuchamy, T. (2015) EE-LEACH: Development of Energy-Efficient LEACH Protocol for Data Gathering in WSN. *EURASIP Journal on Wireless Communications and Networking*, **2015**, 1-9. http://dx.doi.org/10.1186/s13638-015-0306-5

[28] Xu, S., Mu, Y. and Susilo, W. (2006) Authenticated Aodv Routing Protocol Using One-Time Signature and Transitive Signature Schemes. *Journal of networks*, **1**, 47-53. http://dx.doi.org/10.4304/jnw.1.1.47-53

[29] Rezvani, M., Ignatovic, A., Bertino, E. and Jha, S. (2013) Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks. *IEEE Transaction on Dependable and Secure Computing* (*TDSC*), **12**, 98-110. http://dx.doi.org/10.1109/TDSC.2014.2316816