IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 18, 2019, accepted October 30, 2019, date of publication November 11, 2019, date of current version November 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952904

Secure and Energy-Efficient Data Aggregation Method Based on an Access Control Model

AHMED ABDULHADI JASIM[1], MOHD YAMANI IDNA BIN IDRIS[1], SAADIAL RAZALLI BIN AZZUHRI[1], NOOR RIYADH ISSA[1], NOORZAILY BIN MOHAMED NOOR[1], JAGADEESH KAKARLA[2], AND IRAJ SADEGH AMIRI[3,4]

[1]Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

[2]Indian Institute of Information Technology Design & Manufacturing (IIITDM), Chennai 600127, India

[3]Computational Optics Research Group, Advanced Institute of Materials Science, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam

[4]Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam

Corresponding author: Mohd Yamani Idna Bin Idris (yamani@um.edu.my)

This work was supported in part by the Faculty of Computer Science and Information Technology University of Malaya RU Grant (GPF005D-2018).

ABSTRACT Wireless sensor networks (WSNs) consist of a large number of sensor nodes that are distributed to capture the information about an area of interest. In WSN, many of the secure data aggregation works are conducted without addressing the authentication process. It is challenging to implement authentication while preserving the energy consumption in the network. The previous research that focus on these issues have several limitations, such as sharing the security key and the key length with a base station node, and not much attention is given to enhance the authentication of the Medium Access Control (MAC) server. This makes the data aggregation network are exposed to malicious activities. This paper presents a new protocol to address the security and energy issue in Wireless Sensor Network (WSN). This newly developed protocol is named Secure and Energy-Efficient Data Aggregation (SEEDA), which is the extension of SDAACA protocol. The proposed protocol aims to enhance authentication by generating a random value and random timestamp with a secret key. The base station node will verify the fake aggregated data when the packets are received using the generated key earlier. Furthermore, the attacks are detected and prevented by utilizing secure node authentication, data fragmentation algorithms, fully homomorphic encryption, and access control model. The secure node authentication algorithm prevents attacks from accessing the network. To avoid network delays, the base station node utilizes the distance information between the participating nodes. To ensure the reliability of our proposed method, we simulate two well-known attacks, called Sybil and sinkhole attacks. Several experimental scenarios are conducted to observe their effect. Evaluation metrics such as malicious activity detection rate, energy consumption, end-to-end delay, and resilience time are measured. The performance of the proposed protocol is compared with SDA, SDAT, SDALFA, EESSDA, SDAACA, and EESDA, which is a widely used protocol in the area of secure data aggregation. The simulation results show that the proposed SEEDA method outperforms the existing scheme with 98.84% malicious nodes detection rate, 3.04 joules for energy consumption, the maximum delay of 0.038 seconds, and the resilient time 0.054, 0.075 seconds when 8%,16% of malicious nodes affecting the network.

INDEX TERMS Secure data aggregation, access control, wireless sensor network, energy consumption, Sybil attack, sinkhole attack.

I. INTRODUCTION

A wireless sensor network (WSN) consists of nodes and sub-nodes that transmit and aggregate data to the base station node. Due to its low-cost implementation, WSNs are employed in various applications such as wildfire tracking, healthcare, disaster management, smart grid, military surveillance, homeland security, and monitoring [1], [2]. However, WSNs are vulnerable to various attacks because of its distributed wireless nature, which results in delays and loss of

The associate editor coordinating the review of this manuscript and approving it for publication was Sedat Akleylek.

data in the network [3], [4]. WSNs have a high data sensitivity, thereby allowing the adversary to discreetly intercept information from the nodes [5]. For example, the adversary can intercept the transmitted packets by disconnecting the link between the source and destination nodes, generate a fake node with similar identity as the authentic node, or change the path of the transmission. Therefore, security implementation in WSN is crucial to preserve the integrity of the network. However, implementing security in WSN can be challenging due to limited energy available as the energy is highly consumed during data transmission [6], [7]. To extend the network lifetime and allocate the energy for implementing the security, the amount of the transmission overhead should be reduced [8], [9]. Therefore, efficient energy management of data aggregation must be considered in designing a secure network to protect it from attacks and prolong the network lifetime [2], [10], [11]. There are two types of attacks considered in this study, namely, Sybil and Sinkhole. These attacks are harmful and hazardous to the WSNs because they serve as gateways to other attacks, such as disrupting the routing, voting, data aggregation, distributed storage, and may cause network misbehavior. The Following describes the Sybil and sinkhole attacks:

A. SYBIL ATTACKS

The Sybil attack occurs when a malicious node claims to have multiple identities either by creating new identities or impersonating the existing identities. For example, a malicious node may impersonate the identities of the neighboring nodes. This malicious node will repeat automatically and make several copies of themselves to disrupt the network operation. Other than that, the Sybil attack can affect the data aggregation in the network by claiming a fake ID. The malicious node can also steal an identity to enable them to join into the network [3], [12], [13].

B. SINKHOLE ATTACKS

The sinkhole attack affects the network layer by using a path or bandwidth among the nodes. The attack attracts the nearby distributed nodes. The neighbor nodes are faked by the adversary. The sinkhole attack able to drop the data packets or forward them to another attack, or tamper it with aggregated data [13], [14].

Generally, aggregation nodes merge the data collected from their child nodes and forward the secure aggregated data to the base station node [15]. Assuming the adversary nodes may be familiar with most of the security techniques in the WSN, they can reach the nodes by utilizing a wireless communication channel. The adversary may also exploit the process in an ad hoc network [16], [17] due to the unavailability of public-key cryptography techniques in typical WSN. Therefore, a secure data aggregation is necessary to have secure access control for successful data aggregation. Hence, the accuracy of data aggregation with access control may improve the quality of service and reduce energy consumption [18]. Thus, data aggregation requires secure access control to preserve data authenticity and integrity. Data aggregation should be obtained with high accuracy and low communication cost without compromising the data privacy.

In this paper, we propose a secure and energy-efficient data aggregation (SEEDA) using an access control model. We put forward the idea of authentication for WSNs security while at the same time reducing the energy consumption in the network. For authentication, we improve MAC by generating random timestamp and random value with a secret key for the verification of the fake aggregated data when the base station received the packets. This allows the detection and prevention of the attacks when a new node attempts to join the network. In addition, the base station utilizes the distance information to detect attacks. This proposed scheme reduces energy consumption by reducing the redundancy of the transmitted data [19], [20]. The proposed protocol is also focusing on protecting the aggregated data from attacks such as Sybil and sinkhole. These attacks attempt to engage all the nodes in the network which also provides a platform for other forms of attacks, such as tricking and alleviating routing information. This scenario will lead to the increase of traffic generation in the network, sending fake routing data to nodes, and increase the redundancy of data transmission in the network. The contributions of this paper are described below:

- We propose a secure and energy-efficient data aggregation (SEEDA) using an access control model.
- Requires MAC address authentication at the base station with a random value, random time and secret key without sharing key using fully homomorphic encryption.
- The cluster head node sends the broadcast query message. This message consists of data such as node ID, and distance, cluster head node ID, data packets.
- The base station nodes verify the distance and timestamp of all nodes with the broadcast query message from the cluster head node.
- The SEEDA protocol calculates the distance between nodes to choose the best path, reduce the energy consumption, and to avoid the delay in the network.

The rest of this study is organized as follows. Section II presents the related work and section III describes the methodology of the proposed approach. Section IV presents the evaluation metrics and experimental setup. Experimental results are discussed in Section VI. Finally, Section VIII summarizes this study and outlined future works.

II. RELATED WORK

This section describes the methods for securing data aggregation in WSN. Reliable security is important in sensor networks due to the distributed nature of the sensor nodes that make it vulnerable to various types of attacks. Sensor nodes are mainly powered by batteries and frequent replacement of batteries for a large number of nodes is impractical. Therefore, algorithms or security protocols should be highly efficient in terms of energy consumption. Another limitation of the WSNs is to preserve and deliver quality data to another wireless device without the interference of the adversary.

Previous researchers implemented encryption and decryption of secret keys to overcome various security issues using key schedule and lightweight cryptographic. Two types of encryption and decryption approaches exist, namely, symmetric and asymmetric key encryptions. The former uses only one public key for data encoding and decoding, whereas the latter uses two different keys; one for encoding and the other for decoding.

In the symmetric key approach, researchers such as [21]–[23] proposed using the public key for secure data aggregation. The public key can deliver the data from sensors to the base station with one key for encryption and decryption in the network, but the key may be discovered by attackers when the data are sent to the network. To overcome these problems, a message authentication code (MAC) was introduced with symmetric encryption for secure data aggregation [24]. This method successfully secures the authentication of the message of all the nodes in the network because the attacker is unable to guess the MAC address of the author's message when the data is sent through the network. However, this method does not address the confidentiality of the authorized user which makes the network vulnerable to the adversary and risk for an interruption in the communication. Additionally, this method increases the communication and computational overhead and energy consumption when transmitting data between nodes.

In asymmetric key approach, the security is enhanced from inside and outside of the network with authentication [25]. An asymmetric homomorphic encryption scheme was proposed to secure recoverable data aggregation and signature scheme for the heterogeneous network [26]. Although this method protects data privacy, data confidentiality, and integrity, it does not address the authentication during the secure data aggregation operation.

In addition, [27] proposed a polynomial with a probabilistic algorithm to reduce the security risk of the sensor under attacks and share the key among nodes. However, this method does not address energy consumption and does not have a security measure to detect the attacks in the network.

The secure data aggregation homomorphic encryption technique (SDAT) was proposed to achieve network integrity and security via secure data aggregation. However, this method focuses on encryption without considering the detection of malicious nodes, thereby making the network highly vulnerable, which will lead to an increase in energy consumption [28]. To overcome this limitation, Synopsis Diffusion Approach (SDA) enables the sink node or the base station to calculate the aggregation value even with false sub-aggregation attacks. The algorithm calculates the true aggregate value by filtering out the compromised aggregation node hierarchy. However, the base station only receives the authentication process from a few nodes in the network, thereby, the nodes may not have the trust of the server. Also, the adversary can join the network by compromising the authentication of the sensor nodes [29]. The Energy Efficient secure highly accurate and Scalable Scheme for Data Aggregation (EESSDA) on the other hand, proposed a protocol to preserve data privacy for all sensor nodes. However, this protocol is vulnerable to attacks as the secret key is not provided during the aggregation process. Furthermore, authentication between nodes was not addressed in this protocol. Consequently, it decreases the security network and increases communication overhead [30].

The Energy Efficient Secure Data Aggregation (EESDA) proposed channel security and slicing technology to reduce energy consumption and communication overhead [31]. However, this protocol does not ensure the security of all the nodes in the networks because the authentication between nodes is not provided, thus, allowing various attacks, such as Sybil and sinkhole, to join the network. Therefore, the protocol consumes substantial energy and generates, delays in the network.

To address the authentication and solve these problems, the Secure Data Aggregation using Iterative Filtering Algorithms (SDALFA) was proposed to make the network robust against attacks, accurate, and to secure information when individual sensor nodes sending false bit data to the aggregation node. The limitation of this method is that it only provide security for the sink nodes and does not support cryptographic methods [32]. Also, data privacy and integrity are not addressed, thus, resulting in high data redundancy during transmission. The integrity, privacy, and authentication were addressed in the Secure Data Aggregation wireless sensor network that uses Access Control (SDAACA) which calculates the false aggregate value and prevents the network from attacks. However, the MAC authentication of this method is not secure and vulnerable to attacks. Also, the keys are shared through the network where the malicious nodes can steal the key and data in the network [2]. The secret sharing based on energy consumption by multi-hop routing protocol was proposed [33]. However, this protocol only focuses on the security network but does not consider the authentication between clusters member with base station nodes.

Secure data aggregation protocol with malicious nodes identification by [1] was proposed to help the sink node to detect malicious nodes wherein each node has a private key. However, this protocol may generate delays when the nodes encrypt the key between them. In addition, this protocol does not detect and prevent attacks during data aggregation which the attack can create delay, traffic, disconnect, send data to others, generate large data, etc. This results in high energy consumption and shortens the network lifetime. Moreover, distributed data aggregation for wireless sensor network was proposed [34]. This method proposed to reduce the delay and to aggregate data without conflicts. The limitation of this method is that they do not address security and authentication, which may lead the attacks can access the network.

Referring to the aforementioned limitations, the majority of the previous approach unable to secure and authenticate the nodes in the network. Securing all the nodes in the network is important to prevent attacks from an adversary. Furthermore, managing efficient energy consumption and redundancy of data during the transmission in the network can be challenging when sending a packet between nodes. Therefore, SEEDA is proposed to address these limitations. SEEDA uses different functions with different types of nodes to enhance the performance in the network. Furthermore, the protocol reduces energy consumption, by increasing accuracy, and providing authentication, hence reducing communication overhead, and data transmission in the network.

III. THE PROPOSED SEEDA PROTOCOL

This section presents the details of our proposed secure and energy-efficient data aggregation (SEEDA) protocol for WSN. This paper follows the same scenario as presented in SDAACA [2], which considers the oil-refinery monitoring process using WSN. The proposed SEEDA protocol aims to enhance the authentication between the nodes in the network. Furthermore, the proposed protocol detects and prevents the malicious node from joining and accessing the network. The base station will perform checks on the fake aggregated data before sending them to the server by comparing the information of the nodes with the broadcasted query message from the cluster head node when a new node joins the network. The cluster head nodes have the information of the cluster member nodes such as nodes identity, message information, and time stamp.

Our protocol consists of three main algorithms, namely, the data fragmentation, secure node authentication, and fully homomorphic encryption algorithms based on the access control model as shown in Figure 1. The data fragmentation algorithm breaks the data into smaller pieces before the data are transmitted to the next-hop nodes to hide them from being attacked. We use a fragmentation algorithm to keep the original data from the attacker. For example, if the attackers can access the network, they will able to read and transmit data to other malicious nodes or attackers will just drop the original data. Therefore, to avoid these issues and to prevent the attacker from accessing the original data, data is fragmented into blocks. Meanwhile, the secure node authentication algorithm checks if any node is leaving or joining the network to prevent the data between nodes from being tampered or interrupted. The secure node authentication algorithm utilizes an access control model that has the ability to distribute the operation between nodes. The secure node authentication algorithm is helpful for authentication between nodes, for example, the sensor nodes can send data between them directly. If the attacks are carried out and act as valid nodes in the network, attackers will be able to steal the data and can cause transmission delay or causing network interruption. The fully homomorphic encryption algorithm which can protect end-to-end data confidentiality will be applied in this protocol. This ability allows more operations to be implemented without increasing the communication overhead. Thus, the proposed protocol can maintain or reduce the energy consumption in the network while implementing the secure node authentication algorithm. Details, descriptions of the data fragmentation, secure node authentication, and the fully homomorphic encryption algorithms are provided in the next section.

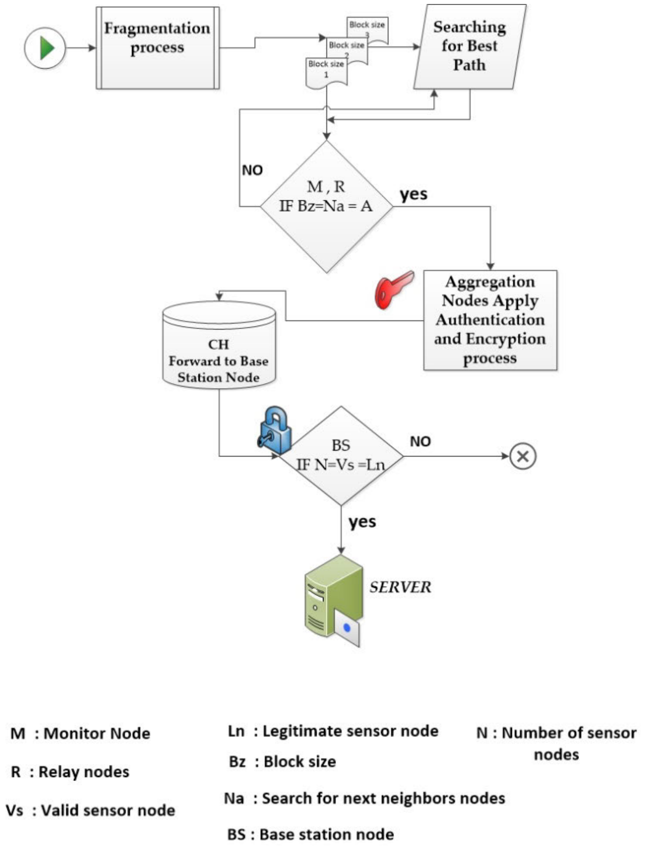


FIGURE 1. The access control model.

We propose a secure and energy-efficient data aggregation protocol to aggregate the data and to make the network highly secure from attacks by checking the aggregated data before transmitting it to the server. To support the energy consumption and to prolong the network lifetime, we employ cluster network topology involving static and mobile sensor nodes. The hierarchical cluster (as in Figure 2) is built using six types of sensor nodes namely child node, monitor node, relay node, aggregation node, cluster head, and base station node. The format of each type of node is described in Algorithm 1. Each node has its own operation and data transmission procedure. They are built to preserve and reduce the energy of nodes, lessen communication overhead, and preventing transmission redundancy in the network.

In this simulation, we use a different type of nodes. To determine and understand the node's type, each node in the network is assumed to have different energy and bandwidth, depending on the deployment location of the node. The energy and bandwidth of different type of nodes are as follows; the relay nodes are set to 11 joules energy and 55 kbps bandwidth, the cluster head nodes are set to 15 joules energy and 80 kbps bandwidth, the aggregation nodes are set to

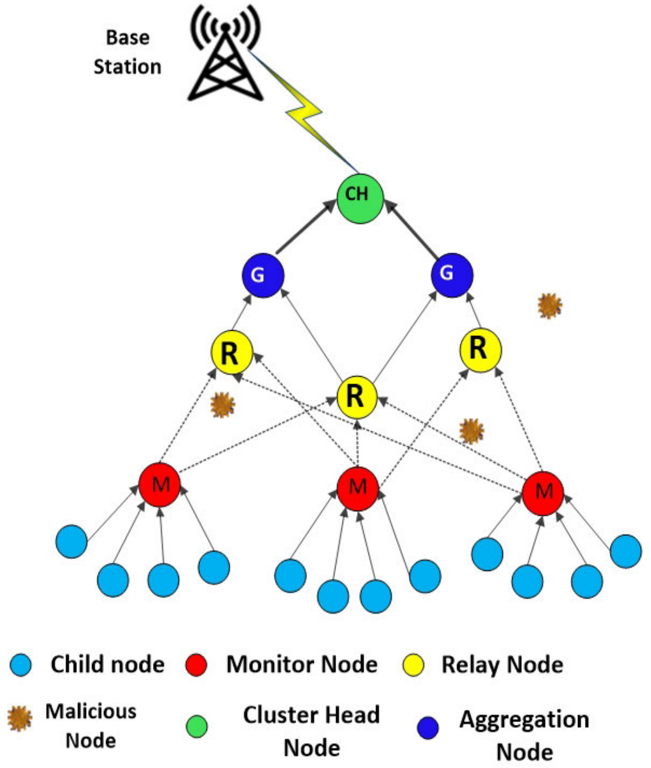


FIGURE 2. The network system model for secure data aggregation.

Algorithm 1 Formatting Type of Nodes

```
1. hierarchical clustering (child, M, R, A, CH, BS)
2. for each cluster member in access control   do
3. child nodes start to send data to next neighbor nodes
4.   if M&R nodes receive data        then
        fragment data into block size
5. while there are more than one cluster members
6.   find the nodes closest distance with M nodes
7.     if the C1 closest with C2     then select and called
   relay nodes (R)
8.   C1== C2      and M == R nodes    Else
9.   repeat to    step 6
End for
  End if
   End if
10.   if the E ≥ 11.2 ≤ 14 J    then    select aggregation
   nodes (A)
8.   R== A nodes and A receive all data information
9.   the A nodes calculate MAC
10. if the E ≥ 10 ≤ 16 J    then select cluster head nodes
   (CH)
11.                    CH== BS
End if
   End if
```

Algorithm 2 Process of Monitor Nodes & Relay Nodes

```
1.  If M collect the data from child nodes then
2.  The M fragment the data block '$D_b$', into block size
             $D_b = B_{z1}, B_{z2}, B_{z3} \ldots B_{zn})$
3.   The monitor and relay nodes search for the best path
                next neighbor nodes
4.         If     the '$B_z$', = M,'$R$'= '$N_a$',     then
5.  the monitor and relay nodes sending fragment data to
                the next neighbor nodes
    End if
```

14 joules energy and 95 kbps bandwidth and the monitor nodes are set to 7 joules and 45 kbps bandwidth. The following describes the function of each node.

A. CHILD NODES

Sense and send the data to the monitor nodes.

B. MONITOR NODES (M), RELAY NODES (R)

The monitor nodes and relay nodes fragment the data into smaller pieces using a data fragmentation algorithm as described in Algorithm 2. The purpose of data fragmentation is to protect and prevent attackers from stealing the data. Consequently, the monitor node will send the fragmented data to the relay nodes and the relay nodes will group the fragmented data and send the information data to the next neighbor nodes.

C. AGGREGATION NODES (G), CLUSTER HEAD NODES (CH)

Aggregation nodes, as the name suggests, will aggregate the data using aggregation functions. Firstly the authentication is performed on the new nodes attempting to join the network to verify their legitimacy as described in Algorithm 3. After that, the aggregation nodes will perform encryption processes and send the encrypted data to the cluster head nodes. The cluster head nodes receive and send aggregated data to the base station node without decrypting the aggregated data. Other than that, the cluster head node sends a broadcast query message to all cluster member nodes to verify their identity and nodes information. The broadcast query message includes address identification of cluster head node, cluster member nodes, and the time as well as data information of the member nodes. The query message with all the information is then sent to the base station node to store the information.

D. BASE STATION NODES (BS)

The base station nodes will receive the aggregated data from the cluster head nodes. The base station node analyzes the aggregated data and checks for the fake aggregated data before sending them to the server by checking the authentication process such as a random value, random timestamp, and secret key, as described in Algorithm 4. Furthermore, the base station node utilizes the distance and timestamp between nodes and checks them with cluster head node information when the new nodes join the network. The base station is

Algorithm 3 Process of Aggregation (G) & Cluster Head Nodes (CH)

```
1. The CH sends a broadcast query message to all cluster members
2. The cluster members receive the query message and send the nodes information to cluster head nodes
3. If the CH = $M_q$ then
4.   CH send and store all information to BS
5.   The aggregation nodes calculate MAC
6.   the aggregation nodes encrypt the aggregated data before sending to CH nodes
7.   The aggregation nodes send the encryption data to CH
8. End if
```

Algorithm 4 Process of Base Station Nodes

```
1. The CH forward aggregated data to BS
2.   BS decrypt the aggregated data
3.   BS investigate the data information from cluster members and check the broadcast message with cluster head nodes
         BS= Cer ($N_{id}$, $D_P$, T, $D_C$)
         A = ($B_s$, $M_i$, $D_P$, T, $D_C$, $N_{id}$)
4. If N== $V_s = L_n$(N) then
5.     $BS$ approval N == $V_s \in S_n$ Else
6.         $B_s \neq$ N and not approval N
7. End if
```

assumed to have substantial energy and memory compared to other nodes.

1) DATA FRAGMENTATION ALGORITHM

The fragmentation algorithm is used to hide and preserve the original data from being tampered by the malicious nodes. In Figure 2, it can be seen that the monitor node, M, and the relay node, R, will fragment the data into smaller blocks using the data fragmentation algorithm. The fragmented blocks information is shown in Figure 3, where it contains SEEDA protocol version, type of service, block size, fragmentation data, address of the source node (i.e. monitor node), address of the destination node (i.e. relay node), and the data packet size, which enable them to be reconstructed back. The malicious node needs to group all the blocks generated from the data fragmentation algorithm in order to intercept the original message. Since each block produced by the fragmentation algorithm have privacy protection component acquired from cluster head node and aggregation nodes, they are not easily tampered. The monitor node and relay node will then search for the best path based on the distance calculation and distribute the data to the relay node. Following the data distribution, the relay node will reconstruct the fragmented data and send them to the aggregation node.

The data fragmentation process is also described in Algorithm 5. The input and output parameters (M, $\boldsymbol{D_b}$, $\boldsymbol{S}$, $\boldsymbol{U}$) of the algorithm is specified in 1–2. In Step 3, the monitor node 'M' collects the data block $'\boldsymbol{D_b}'$, from the child nodes and checks the number of block data. The monitor node fragments the data block into small size blocks $'\boldsymbol{B'_z}$, ($\mathbf{D_b} = \mathbf{B_{z1}}, \mathbf{B_{z2}}, \mathbf{B_{z3}} \ldots \boldsymbol{B_{zn}}$) in Step 4-5. In Steps 6, due to the random deployment of sensor nodes in the network, the monitor node searches for the next neighbor nodes $'\boldsymbol{N'_a}$, with the nearest distance to send the fragmented data to the relay nodes '$\boldsymbol{R}$'. This process decreases the energy consumption. When the monitor node finds the neighbor nodes, the monitor node keeps sending the data until all the data are forwarded to the relay node $\boldsymbol{M} = \boldsymbol{N_a} \rightarrow \boldsymbol{B_z to\ R}$, as described in Steps 7–8. In Step 9, all the fragmented data are received in relay nodes $\boldsymbol{B_z} = \boldsymbol{R}$, then the relay nodes send the fragmented data to the aggregation node '$\boldsymbol{D'_a}$, for the aggregation process S ∨ U. After the aggregation and encryption process is completed, the aggregation node forwards the encrypted data to the cluster head node, 'CH', as described in Step 10. Finally, in Steps 11–12, the base station receives the encrypted aggregation data from the cluster head node CH ‖‖ B_s using the algorithm defined in [9].

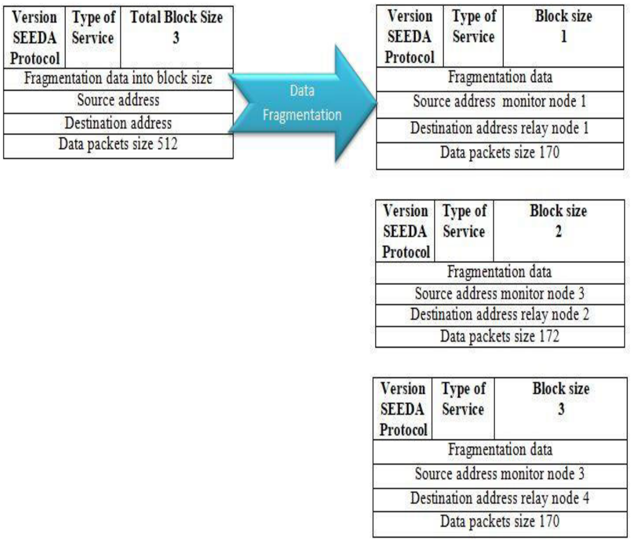


FIGURE 3. The data fragmentation process.

2) ACCESS CONTROL MODEL

There are three important modules involved in the access control model, namely the authentication and authorization process, medium access control an data integrity, and authentication and redundancy. The proposed secure node authentication algorithm detects and prevents malicious attacks from accessing the network by checking the secure authentication of the new nodes. The medium access control generates and calculates the MAC and data aggregation functions. The authentication process describes the authentication procedure and checks the distance between nodes. Our method aims to enhance the authentication between nodes and to decrease redundancy. This is because many researchers focused on

Algorithm 5 The Data Fragmentation

```
M: monitor nod; $D_b$: data block; $B_z$: block size; $N_f$: number
of fragment block size; data; $N_a$: search for next neighbors
hope; $B_s$: base station; CH: cluster head node; R: relay
nodes; Da: data aggregation; S: set functions for aggregation; U: authentications.
Input (M, $R$, $D_b$, S, U)
Output ($B_s$, $N_f$)
 If the M collect the data from child nodes;     then
 The M, R, fragment data $D_b$ into small size blocks $B_z$
              ($D_b = B_{z1}, B_{z2}, B_{z3} \ldots B_{zn}$)
    The M search the $N_a$ to send the $B_z$
      If    the M find the $N_a$; and M $\epsilon N_a$     then
              The $M = N_a \rightarrow B_z to$ R
          $N_a \rightarrow B_z to$ R forward process, and
          $N_a$ *repeats to steps* 6 − 8 until
                        $D_a$ *received* $D_b$
          The $D_a$ collect all data $D_b$, $N_f$
          The $D_a$ gather the data fragment by
        applying a process
                $S \vee U$
    CH collect the aggregated data from $D_a$
          The CH forwards the encrypted data to base
        station node
                    CH ||| $B_s$
    End if
End if
```

the security and aggregation process without addressing the issues of the authentication and authorization between nodes. The details of thethree important modules are described in the following subsections which discuss authentication and authorization process, MAC and data integrity detection, and authentication and redundancy.

3) SECRET KEY

In this section, we discuss the secret key with the SEEDA protocol. We are using a fully homomorphic encryption algorithm that can protect end-to-end data confidentiality as defined in [9] and perform encryption on the aggregated data before sending it to the base station nodes. We are employing the encryption process when the data information gets to the aggregation node because the monitor nodes and relay nodes are focusing on the fragmentation process to hide the original data from the adversary. By doing so, the encryption process from monitor nodes is not needed. This action will reduce the transmission delay. The aggregation nodes will aggregate the received data from relay nodes and check the authentication process before sending it to cluster head nodes. When the aggregation nodes completed the authentication process, the aggregation nodes will encrypt the aggregated data using an encryption algorithm to make the data highly secure and preventing attackers from stealing the data in the network. At the same time, the aggregation nodes send the encryption data to cluster head nodes, and the cluster head nodes will then forward it to the base station nodes without decrypting the aggregated data.

a: AUTHENTICATION AND AUTHORIZATION PROCESS

Prior works on authentication lack of focus on authentication with authorization method. Authentication is the process of verifying the legitimacy of the new nodes that join the network. This process is performed at the base station. The aim of this process is to prevent the adversary nodes from joining the network and act as original nodes to collect data from the network. The authorization is the process that allows only authorized users to read and transmit the data. The implementation of both authentication and authorization processes in the network is important because if the malicious nodes have successfully joined the network, the authorization process can prevent these nodes from accessing the data in the network. Therefore, we include both secure node authentication and authorization algorithms in the proposed SEEDA protocol as described in Algorithm 6.

Steps 1-2 presents the input and output parameters of the algorithm. In Steps 3-7, the cluster head node sends the broadcast query message to the cluster sensor nodes. The sensor nodes that received the query message from the cluster head node computes its node identity and information messag $M_q = (CH_{id} \| S_{nid} \| M_i \| T)$. In steps 8-12, the aggregation node informs the base station node to verify the aggregated data from the new sensor node by checking the node ID, message, timestamp and distance (N_{id}, D_P, T, D_C). The base station checks the query message in the cluster head node as described in steps 13-14. The authorization process is performed if the distance and certificates for the new sensor node are similar to the original nodes in which, the new node is considered as valid and authorized. After completing the authorization process, the base station authorizes the new node to join the network and allows the data to be sent between nodes, as described in Steps 15-17. Conversely, if the new sensor node is malicious and unauthorized, the base station will reject the new node from joining the network. This process is described in Steps 18-19.

b: MEDIUM ACCESS CONTROL AND DATA INTEGRITY

This section explains the access control model and the procedure of the base station to check authentication process and to secure the data aggregation. The base station utilizes the distance and timestamp to examine the authenticity of all nodes in the network. We assume the node N acts as a malicious node that creates fake sub-aggregate data for authentication. First, the malicious node N creates false data with a random value. Then, node N sends the medium access control (MAC), which contains the random value and personal identity of node N to the base station for authentication. When the base station receives the MAC, the base station will verify the legitimacy of node N by checking the node identity, data packets, distance between nodes, and their timestamp.

Algorithm 6 Secure Node Authentication

```
$B_s$: base station; N: number of the sensor node; CH: cluster head nodes; Cer: certificate of the sensor node; $M_i$: information message; $M_q$: query message; A: authorization; $A_p$: approval; $S_n$: sensor the network; $D_P$: data packets; $D_C$: distance between nodes; $T$: broadcast the time of nodes; $V_s$: valid sensor node; $L_n$: legitimate sensor node;
$L_i$: illegitimate sensor node; En: entry the network
Input $M_i, N_{id}$, N, T, $D_C$)
Output (A, Cer, )
for each member sensor nodes in
        access control model    do
    CH sends a query message to it
        all member sensor nodes ($M_q$)
after receiving $M_q$)(the sensor nodes compute the
        $M_q = (CH_{id} \| S_{n\,id} \| M_i \| T)$
            CH → $M_q$
    If  the new sensor node join the network
        N → En ∈ $S_n$    then
            $CH$ inform $B_s$
                $B_s$ recall N
        $B_s$ investigate Cer ($N_{id}, D_P$, T, $D_C$) for the $V_s$
            set A for N; A = ($B_s, M_i, D_P$, T, $D_C, N_{id}$)
            $B_s$ check broadcast A($M_q$) with CH
                If  N== $V_s = L_n$(N) and
                        N== $M_q$    then
                    $B_s$ approval N == $V_s \in S_n$
                Else

                    N ≠ $V_s$ and N ≠ $M_q$ $L_i = V_s$
                    $B_s$ ≠ N and not approval N
        End if
    End if
End for
```

The MAC is calculated using equation 1 given as below

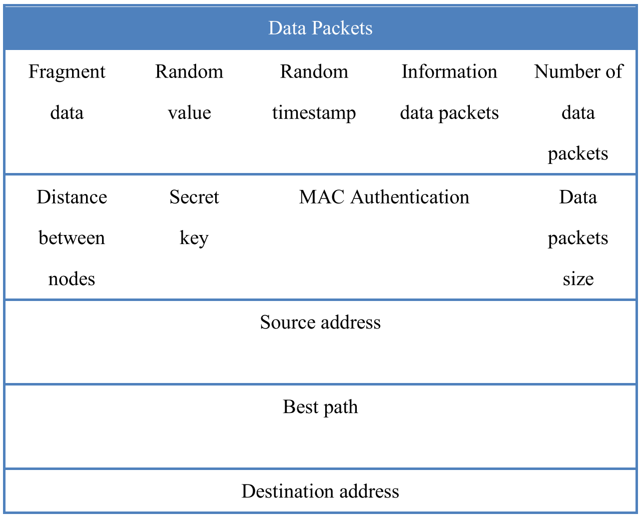
$$MAC = \int_{k+1}^{n} \{k\,(R_v) + n_i \tag{1}$$

where, k is a set of sensor nodes, R_v is the random value, n_i is the node ID. In Equation 2, the base station creates the random value and random timestamp to authorize node.A random value, R_v, is an arbitrary number that is used to avoid malicious attacks due to duplication. A random timestamp, R_t, is a timestamp encoded with a random number. The malicious node needs to know the time it takes for a specific node to transfer data to the base station and their random number in order to masquerade an attack.

The base station with a random value and random timestamp can be generated as follows:

$$B_{s(v,t)} = \sum\nolimits_{j+1}^{j} R_t + \sum\nolimits_{i+1}^{i} (R_v 1)^{gv} * n\,(s) \tag{2}$$

where, $B_{s(v,t)}$ is the base station with a random value, and random timestamp, R_t is the random timestamp, gv is the random value generated by a malicious node, n(s) is the number of the sensor node, i, j are the set of a random value and random timestamp. The data aggregate can be computed as follows:



Data Packets				
Fragment data	Random value	Random timestamp	Information data packets	Number of data packets
Distance between nodes	Secret key	MAC Authentication		Data packets size
Source address				
Best path				
Destination address				

FIGURE 4. The data packets format.

$$DA = \sum\nolimits_{n=1}^{n} (B_s) * MAC \tag{3}$$

where, DA is the data aggregation, n is the set of sensor nodes.

Lemma: The malicious sensor nodes unable to create MAC with fake data that is similar to the original data recorded at that base station.

Proof: Let's assume node N can create the random value with false data and send to the base station for authentication

$$N = (n_{id}, R_v, b1 \ldots bn) \tag{4}$$

To improve the authentication and allow the base station to determine the fake data, we not only create the random value with aggregated data, but we also create random timestamp and secret key. The following equation shows the medium access control with the aforementioned security measures.

$$\begin{aligned} N_{MAC} &= (k_e + R_v + R_t + D_p + b1..bn) \\ N_{MAC} &= N \end{aligned} \tag{5}$$

where, the N_{MAC} is the medium access control, k_e *is the secret key*, D_p is the data packets, R_t is the number of random timestamp, R_v is the number of a random value, b1..bn is the number of bit data. We design the secret key by using the fully homomorphic encryption to make the network highly secured. We use an encryption process between aggregation and base station nodes to preserve energy. Our protocol distributes data to all nodes to enable the valid nodes to share the data packets between them in the network. The data packets format is as shown in Figure 4. This message security is very helpful to prevent attacks from accessing the network. The malicious nodes cannot create similar messages such as the time and the secret key. Apart from that, the base station also holds the distance and ID between nodes from the cluster head nodes. For this reason, the malicious node will not be able to join the network and share its data.

c: AUTHENTICATION AND REDUNDANCY

The authentication process makes it challenging for an attacker to join the network. This authentication method is expected to enhance the security of the network since the design of the network and key encryption only allows authorized users to transmit the data.

We assume the aggregation node (G) sends the secure aggregated data to the cluster head nodes (CH). This operation can be described as:

$$\boldsymbol{p_e} = \{\boldsymbol{n_{id}}, \boldsymbol{CH_{id}}, \boldsymbol{N_{MAC(n,CH)}}, \boldsymbol{D_P}, \boldsymbol{D_c}\} \tag{6}$$

where , p_e is the packet encryption, $MAC_{k(n,CH)}$ is the key of message authentication code for cluster member nodes and cluster head nodes, D_P is the data packets, D_c is the distance between nodes. The cluster head nodes receive the packet encryption from aggregation nodes and then the cluster head node forwards the packet encryption to the neighbors of cluster head nodes or to the base station. The process of cluster head node (CH) forwarding the data to the base station nodes can be written as:

$$\boldsymbol{p_{e1}} = \{\boldsymbol{CH_{id}}, \boldsymbol{NCH_{id}}, \boldsymbol{N_{MAC(CH,NCH)}}, \boldsymbol{D_P}, \boldsymbol{D_c}\} \tag{7}$$

where, the NCH_{id} is the identity of the next hop cluster head node, $MAC_{k(CH,NCH)}$ is the group of encryption key transmission between cluster member nodes or cluster member nodes with base station.

We propose these equations to enhance the authentication and integrity of the message encryption when the data are sent to the nodes to reach the base station node. Finally, substituting equation (6-7) into equation (8): where, T_s is the total data encryption sent through the network.

$$\begin{aligned} \boldsymbol{T_S} = [&\{\boldsymbol{N_{id}}, \boldsymbol{CH_{id}}, \boldsymbol{N_{MAC(n,CH)}}, \boldsymbol{D_P D_c}\} \\ &+ \left\{\boldsymbol{CH_{id}}, \boldsymbol{NCH_{id}}, \boldsymbol{N_{MAC(CH,NCH)}}, \boldsymbol{D_p}, \boldsymbol{D_c}\right\}] \end{aligned} \tag{8}$$

The base station node calculates the distance between nodes to determine the best path to the next nodes for data transmission and to check for a node that join in the network. This operation helps to avoid the redundancy of data transmission in the network because the nodes will send the data in a short time and will not generate traffic control through the sending process. The distance between the nodes can be calculated as:

$$\mathbf{D_c} = \mathbf{N} \times \frac{\mathbf{T_r}}{\boldsymbol{S}} + \mathbf{C_p} \tag{9}$$

where, D_C is the distance between nodes, N is the number of nodes in the network, T_r is the transmission range, C_p is the ID number of clustering node, **S** is propagation speed of the signal. Due to the random deployment of the nodes, equation (9) checks the distance between nodes that helps to choose the best path for data transmission to next-hop nodes. Consequently, it also reduces the delay and data redundancy in the network.

IV. EVALUATION METRICS

The lemma and proof of the proposed method have been presented in the previous section. To further test the reliability of the proposed method, four evaluation metrics are considered. These metrics measure the security and performances of the network namely, the detection rate of the malicious nodes, energy consumption and accuracy, end to end delay and resilient time in the network. These metrics are explained in the following sub-section.

A. DETECTION RATE

The performance of the proposed SEEDA protocol is evaluated by simulating Sybil and sinkhole attacks. The malicious nodes are detected by checking the false data inserted into the aggregated data. The detection rate of malicious attacks can be written as equation 10 [35] :

$$\boldsymbol{D_m} = \frac{\boldsymbol{A_d}}{\boldsymbol{A_d} + \boldsymbol{F_d}} \tag{10}$$

where,

D_m is the detection rate of fake aggregated data, A_d is the number of aggregated data and, F_d is the number of false aggregated data. The number of false aggregated data depends on how many malicious nodes in the network.

B. ENERGY CONSUMPTION AND ACCURACY

The efficient management of energy consumption in the network is very important for secure data aggregation. One of the goals of our proposed protocol is to reduce or maintain the energy even when the malicious attack occurs in the network. It is also designed to prolong the network lifetime by reducing the communication overhead. Let's assume the clusters sensor nodes P_{d1} and aggregation nodes P_{d2} send the data packets and messages between them. The equation can be written as:

$$\boldsymbol{C_1} = (1 - \boldsymbol{P_{d1}}) * \boldsymbol{S_n} * \boldsymbol{P_{d2}} \tag{11}$$

$$\boldsymbol{C_2} = (1 - \boldsymbol{P_{d1}}) * \boldsymbol{S_n} * \sum_{\boldsymbol{S=0}}^{\boldsymbol{Pd2}} \boldsymbol{S} * (\boldsymbol{D_a} - 1) \tag{12}$$

where,

The C_1 is the communication overhead for node 1, C_2 is the communication overhead for node 2, P_{d1}, P_{d2} are the data packets sent between the nodes, S_n is the number of sensor nodes, D_a is the aggregated data. On the other hand, the total communication overhead C_t, of the exchanged message can be written as:

$$\boldsymbol{C_t} = \boldsymbol{C_1} + \boldsymbol{C_2} \tag{13}$$

where , C_t is the total communication overhead.

The energy consumption can be evaluated and computed as:

$$\boldsymbol{E_C} = \boldsymbol{C} * \boldsymbol{V_S} \tag{14}$$

where, E_c is the energy consumption, C is the initial energy to send data, V_S is the average of send bit data per second.

Thewasted energy when node N transmits the packets to the next node N_1 can be calculated as:

$$\boldsymbol{W_E\left(N, N_1, D, P_Z\right) = \left(C1 + C2 * D\left(N, N_1\right)\right.} \\ \boldsymbol{* V_S * P_Z} \qquad (15)$$

where, W_E is the wasted energy, D(N, N_1) is the distance between (N, N_1) nodes, P_Z is the packet size., C1 is the cluster node 1, C2 is the cluster node 2. The energy when packets are received between nodes can be written as:

$$\mathbf{E} = \boldsymbol{(N, N_1, P_Z) = (V_S * P_Z)} \qquad (16)$$

C. END TO END DELAY

The end to end delay is defined as the time difference between the time when the packet aggregation occurs and the time when the packet arrives at the aggregate queue. The end to end delay can be computed as:

$$\mathbf{D} = \frac{\sum_{i+1}^{P} \boldsymbol{(T_{rec\ i} - T_{send\ i})}}{\boldsymbol{P}} \qquad (17)$$

where,

D is the end to end delay,

T_{rec} is the time when packets are received, T_{send} is the time when sending packets, and

P is the total number of packets.

D. RESILIENT TIME IN THE NETWORK

After collecting a set of time offsets from multiple nodes, the malicious time offsets from Sybil and sinkhole attacks are identified. These malicious time offsets will be excluded and the rest of the time offsets are used to estimate the actual time offset. The resilience time shows the performance of the network after the network is compromised. The resilience time can be written as:

$$\mathbf{R_S = T_i - T_j} \qquad (18)$$

where, R_S is the resilience time, T_i is the set of time offsets from nodes, T_j is the set of time offsets under malicious nodes.

V. COMPLEXITY ANALYSIS OF THE PROTOCOL

We suppose that they are N total number of nodes in the network, and C the constant of complexity. In our protocol, we have one loop in algorithm and iteration, so the computational complexity is *O(N).* The complexity is acceptable for a large network with a large number of sensor nodes in the network. In addition, the SEEDA protocol has found the optimal distance and best path to reduce the energy consumption, and significant improvement over the many O(N) algorithms and protocols in the literature. The following describes the communication control message:

1. Calculate the cluster member's access control in the network, these are called C0.

2. The CH send a broadcast message to all cluster members N.

3. If the new node joins the network, the base station will check the authentication in the network, these can be described as C1+C2+C3.

The overall control message can write as:

$$\begin{aligned} \mathrm{C0} &= \mathrm{N}^*(\mathrm{C1} + \mathrm{C2} + \mathrm{C3}) \\ &= \mathrm{C0} + \mathrm{C1\ C2\ C3N} \\ &= O(N) \end{aligned}$$

TABLE 1. **Simulation parameters.**

Parameters in SEEDA	Value
Transmission range	50 meters
Initial energy of relay node, monitor node, aggregation node, cluster head node	7,11,14,15 joules
Simulation time	22 minutes
Network size	400*400 m²,1000*1000 m²,
Number of sensor nodes	400
Bandwidth for relay , monitor, cluster head, aggregation nodes	45,55,80,95 kb/sec
Buffering capacity	50 packets at each node
Data packet size	512 bytes
Initial pause time	16 seconds
Power intensity	-14dbm to 13dbm

VI. SIMULATION SETUP

The proposed protocol used an access control model to secure the data aggregation and to efficiently conserve the energy. We design the model and simulation as in SDAACA protocol [2]. The proposed protocol was simulated using the network simulator 3.25 running on Ubuntu operating system 16.4 LTS version. Several scenarios of attacks were performed to evaluate the ability of the proposed protocol to detect malicious nodes. 8% to 30% of Sybil and sinkhole attacks in the network were considered during the evaluation.

A total of 400 sensor nodes such as child nodes, monitor nodes, relay nodes, aggregation nodes, cluster head nodes, and base station nodes were deployed in the area of 400 × 400 m^2 and 1000 × 1000 m^2. In addition, several different scenarios are generated to show the performance and to check the ability of the proposed protocol to detect malicious attacks. Table 3, presents the different scenarios generated such as a different number of nodes and different area sizes. All the examined scenarios show similar results

TABLE 2. Malicious node detection rate % comparison with different protocols.

Malicious Node	SDA	SDAT	EESDA	SDALFA	EESSDA	SDAACA	SEEDA
0	0.88	0.93	0.93	0.96	0.93	0.96	0.984
1	0.89	0.89	0.92	0.95	0.904	0.94	0.97
2	0.89	0.895	0.902	0.952	0.906	0.942	0.99
3	0.9	0.902	0.906	0.954	0.912	0.93	0.972
4	0.92	0.89	0.912	0.957	0.93	0.97	0.976
5	0.91	0.896	0.915	0.95	0.934	0.968	0.96
6	0.93	0.898	0.928	0.94	0.936	0.968	0.967
7	0.92	0.897	0.94	0.942	0.938	0.957	0.97
8	0.9	0.906	0.928	0.901	0.94	0.96	0.969
9	0.89	0.8997	0.89	0.89	0.942	0.94	0.98
12	0.896	0.8995	0.889	0.895	0.946	0.95	0.981
15	0.8955	0.89355	0.893855	0.88	0.95	0.97	0.93856
18	0.88	0.8923	0.89283	0.886	0.956	0.96	0.928395
21	0.892	0.8912	0.89182	0.891821	0.91821	0.95	0.918292
24	0.891	0.8901	0.89	0.89001	0.9001	0.9	0.900198

when the nodes and area sizes of the network are smaller or larger. The parameters used in the simulation are presented in Table 1. Various values of energies and bandwidths for the sensor nodes were tested depending on the deployment location of the node. For example, the relay nodes are set to 11 joules energy and 55kbps bandwidth, the cluster head nodes are set to 15 joules energy and 80kbps bandwidth, the aggregation nodes are set to 14 joules energy and 95kbps bandwidth, the monitor nodes are set to 7 joules and 45kbps bandwidth. The standard energy for the sensor network is 3.5 joule.

A. SIMULATION RESULTS

The simulation results show that the performance of the proposed protocol successfully keeps the network highly secure. This is because the proposed protocol enhances the authentication by generating a random value and random timestamp with a secret key which makes it difficult for the adversary to replicate. Thus, prevent unauthorized node to join the network. The sensors nodes are also protected by fragmenting the data into small pieces before transmitting it to the next-hop nodes. The base station node verifies the fake aggregated data and checks the certificate nodes. Furthermore, the base station node utilizes the distance between nodes and timestamp to secure the network. Apart from security, the distance information is used to speed up the time by choosing the optimal next-hop nodes.

The performance comparison between the proposed protocol and the other six protocols in securing data aggregation are presented in Figure 5 to Figure 8.

In Table 2, it can be seen that the detection rate of the proposed SEEDA protocol is approximately 98.84% with the presence of 24% malicious nodes in the network as shown in Figure 5. This indicates that the proposed protocol is only slightly affected by the increment of the malicious nodes in the network. The detection rate of other prior protocols ranges between 88.02–96.6% with 24% of malicious nodes in the network. The proposed protocol uses the authentication process in all nodes to validate the new nodes and only allows the authorized nodes to join the network.

Energy consumption is an important factor for a secure data aggregation system. The data packets and messages that send between clusters sensor nodes and aggregation nodes consume energy and will continue to consume more energy with the participation of the adversary. This evaluation will investigate the effect of Sybil and sinkhole attacks on energy consumption. Figure 6 plots the energy consumption in joules as a function of process time in days. From the figure, it can be observed that the proposed SEEDA protocol consume the least energy in all three different scenarios with 10% (Figure 6(a)), 20% (Figure 6(b)), and 30% (Figure 6(c)) malicious node with 2.51 joules, 2.90 joules, and 3.4 joules respectively in 16 days. The other protocols consume about 3.5-4.0 joules, 3.71-3.89 joules, 3.60-4.0 joules with

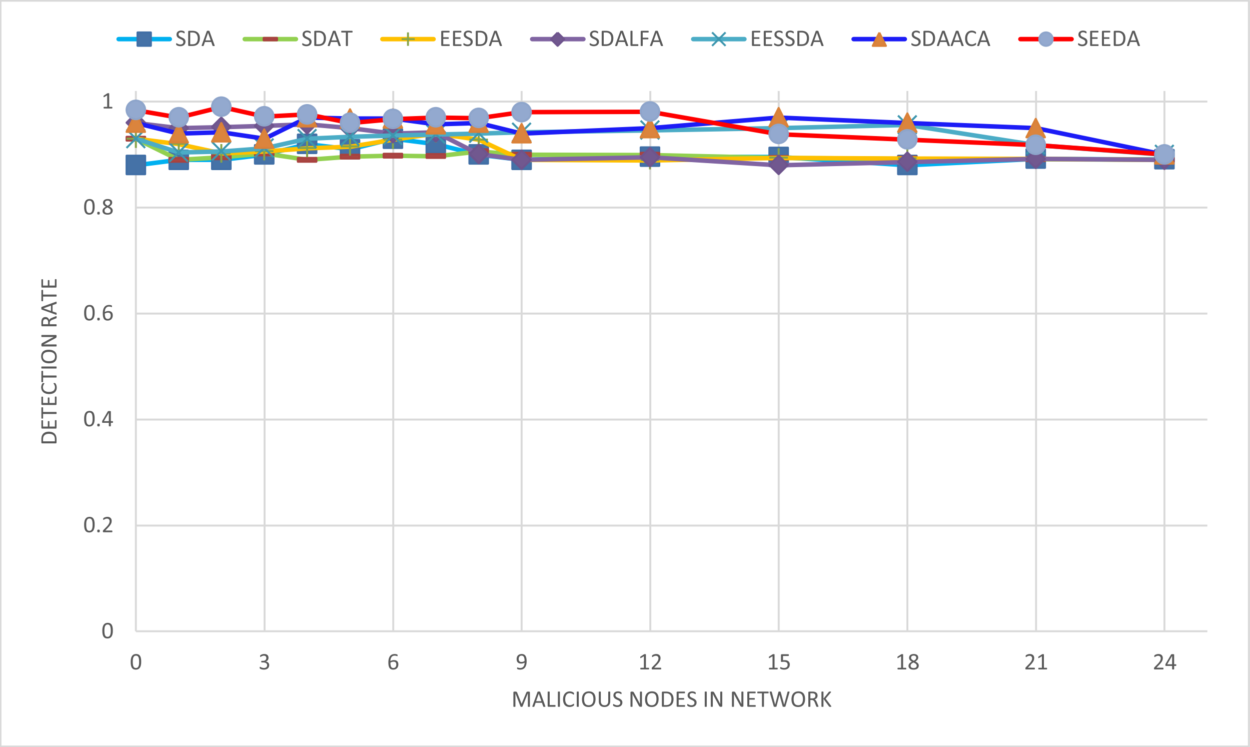


FIGURE 5. The detection rate with malicious nodes from 0% to 24%.

10%, 20%, 30% malicious nodes in the network respectively in 16 days. The reason for the efficient performance of our protocol is caused by reducing of the communication overhead and reducing the delay among nodes in the network.

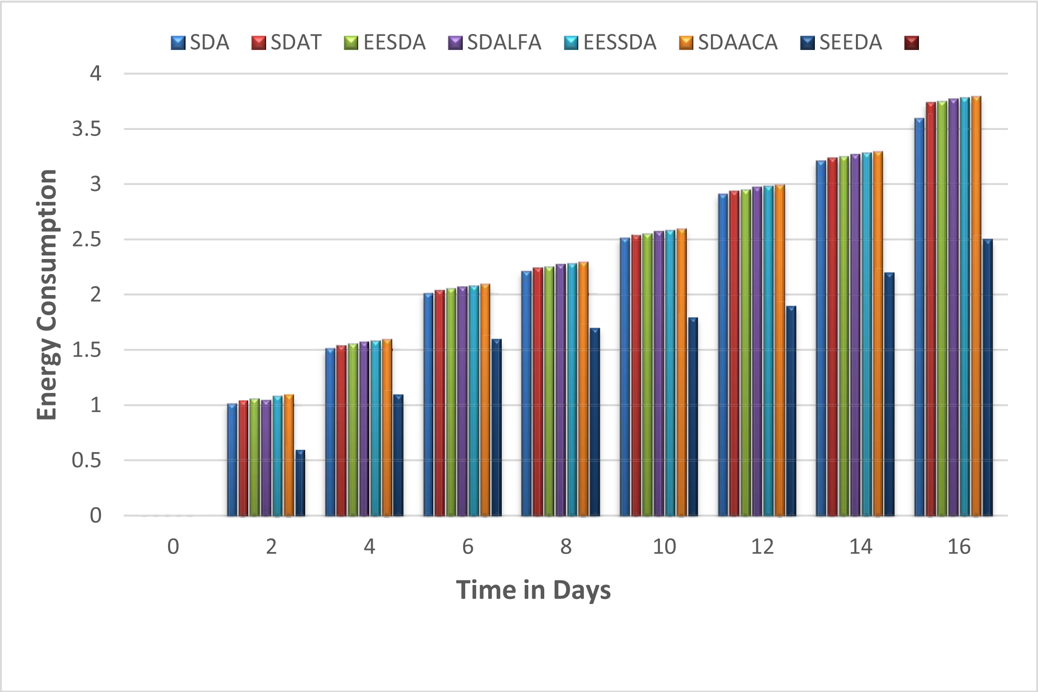
The end to end delay (in seconds) for all protocols is shown in Figure 7. The delay of the proposed protocol is minimal compared to other data aggregation protocols because the base station nodes utilize the distance and timestamp between the nodes to prevent the attacks from accessing the network, consequently, it helps to reduce the delay and avoiding the network traffic. The proposed protocol recorded a maximum delay of 0.038 seconds whereas the prior protocols have higher latency between 0.059–0.08 that considered as very high for sensitive applications.

Figure 8 (a) shows the resilient time when 0% to 8% malicious node affected the network, while, Figure 8(b) shows the resilient time when 0% to 16% malicious node affected the network. Both simulations are conducted separately. Compared to other protocols, the proposed SEEDA protocol shows the best result with 0.054 seconds resilient time when affected by 8% malicious node (Figure 8(a)) and 0.075 seconds when affected by 16% malicious node (Figure 8(b)). The least resilient time shown by other protocols are at 0.068 seconds when affected by 8% malicious node (Figure 8(a)) and 0.094 seconds when affected by 16% malicious node (Figure 8(b)). The proposed SEEDA protocol outperforms the others because of the secure node authentication and the base station node checks for the false data aggregation to avoid attacks. Also, the SEEDA protocol can identify the integrity of the nodes. These operations have led to a low resilient time compared to other protocols.

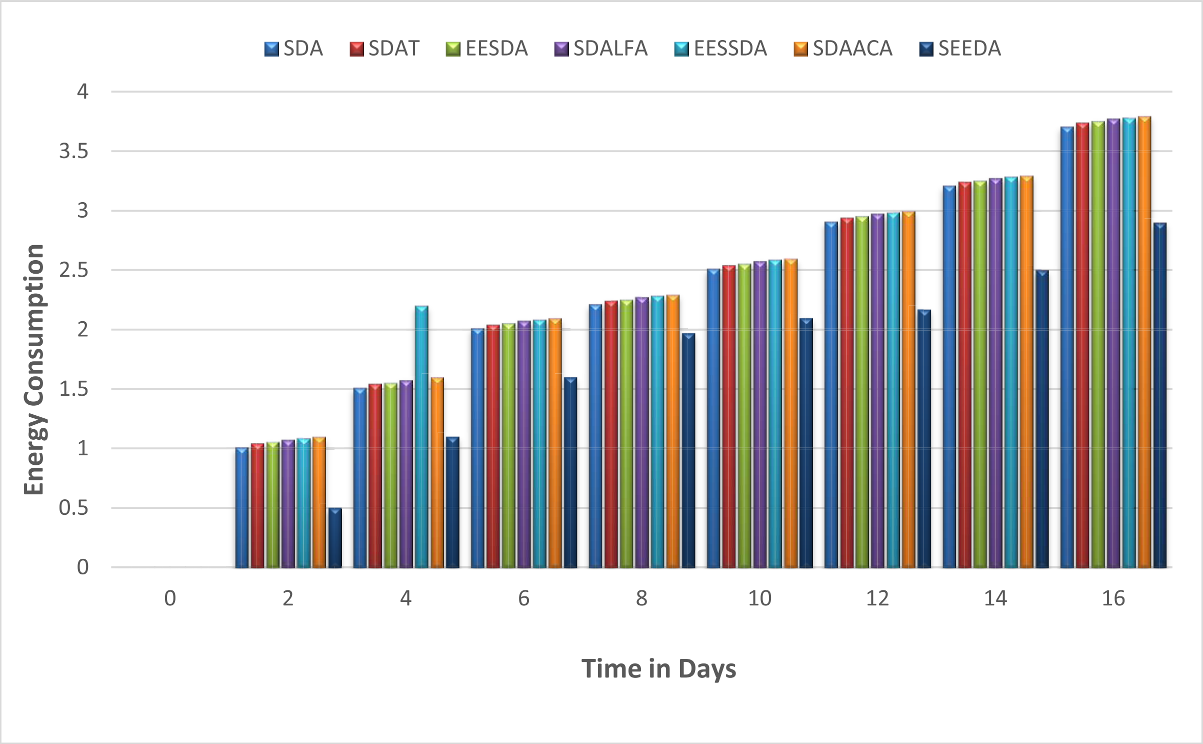
VII. DISCUSSION

The proposed SEEDA protocol is an improved version of the SDAACA protocol. Both methods utilize data fragmentation algorithm and random value but the proposed SEEDA protocol also includes query mechanism, distance information, and random timestamp. The data fragmentation algorithm is used to hide and preserve the original data from being tampered by the malicious nodes. The malicious node needs to group all the blocks generated from the data fragmentation algorithm to reconstruct the original message. In the proposed SEEDA protocol, the cluster head node sends a broadcast query message. This message consists of data such as node ID, and distance, cluster head node ID, data packets that are used to protect the fragments and data. The fragmented data can be distributed to different or multiple node locations. If one block is compromised, the other block can be used as a backup. In a situation where data retransmission is needed, the fragmentation method still has an advantage since only a compromised block will need to be retransmitted. The data distribution is also able to increase the network lifetime due to the selection of high energy nodes.

As the location of the legitimate node is known, the utilization of distance may differentiate them from the malicious node. If a node has a different estimated distance measurement as the legitimate node, it can be regarded as a malicious node. The distance can be calculated via



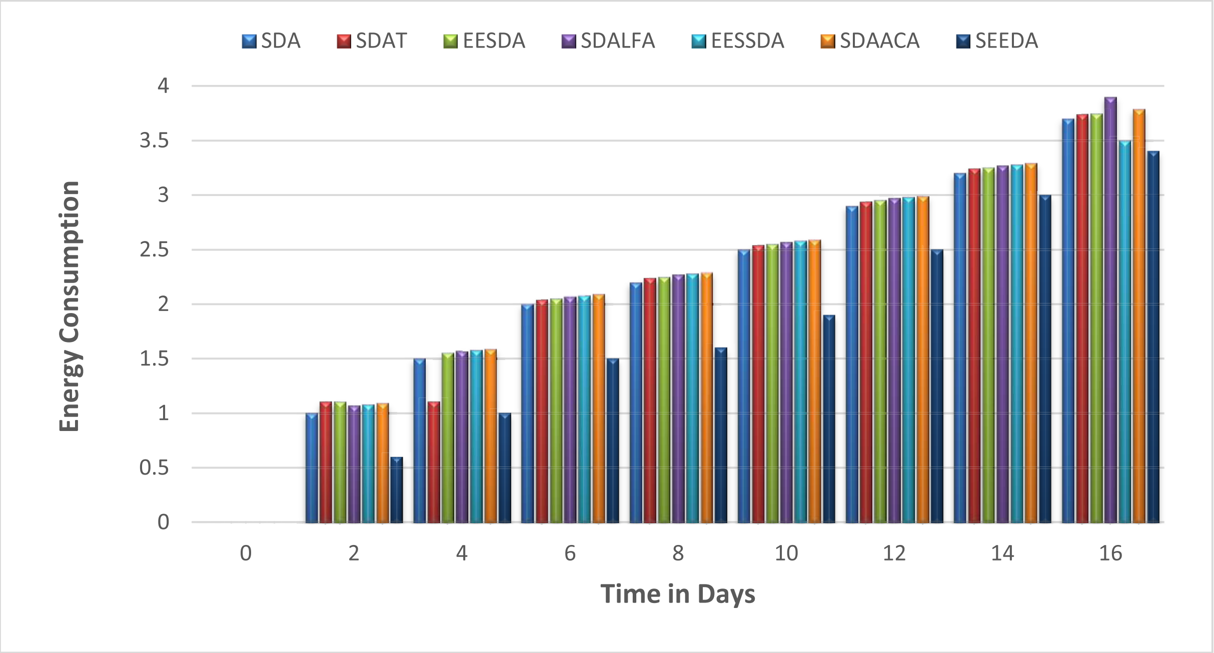
(a)



(b)

FIGURE 6. The energy consumption with a) 10% malicious nodes, b) 20% malicious nodes, c) 30% malicious nodes.

transmission range and signal strength, therefore, it is not easily tampered. Other than that, the utilization of the distance information between nodes is used to search for the best path and to reduce transmission cost. Rather than transmitting data



(c)

FIGURE 6. ***(Continued.)*** **The energy consumption with a) 10% malicious nodes, b) 20% malicious nodes, c) 30% malicious nodes.**

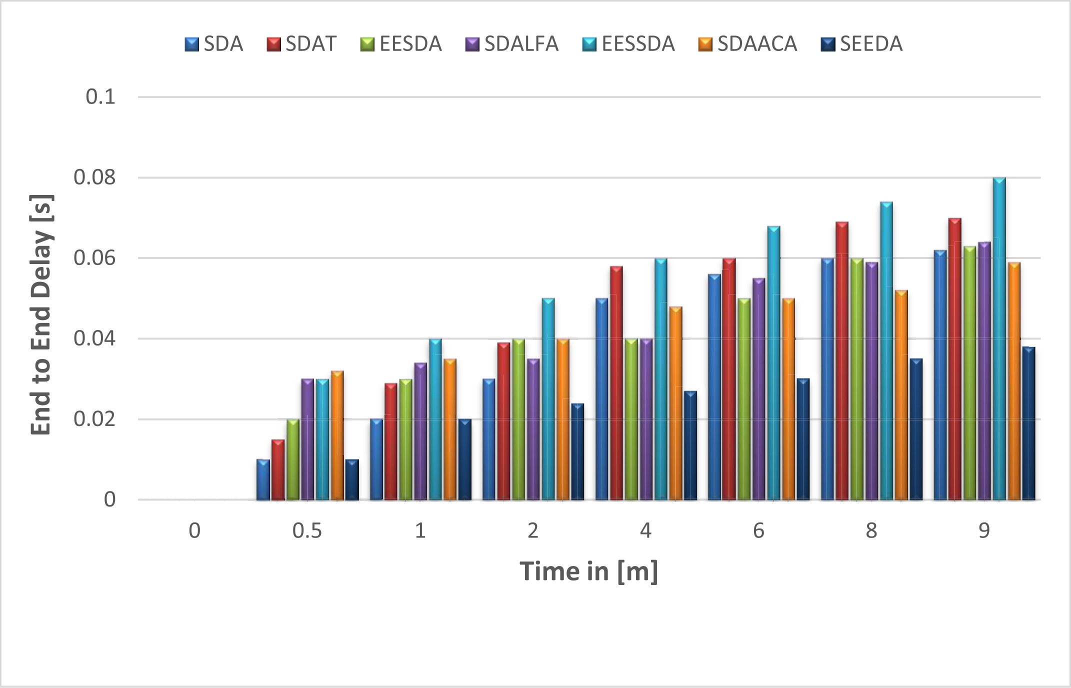
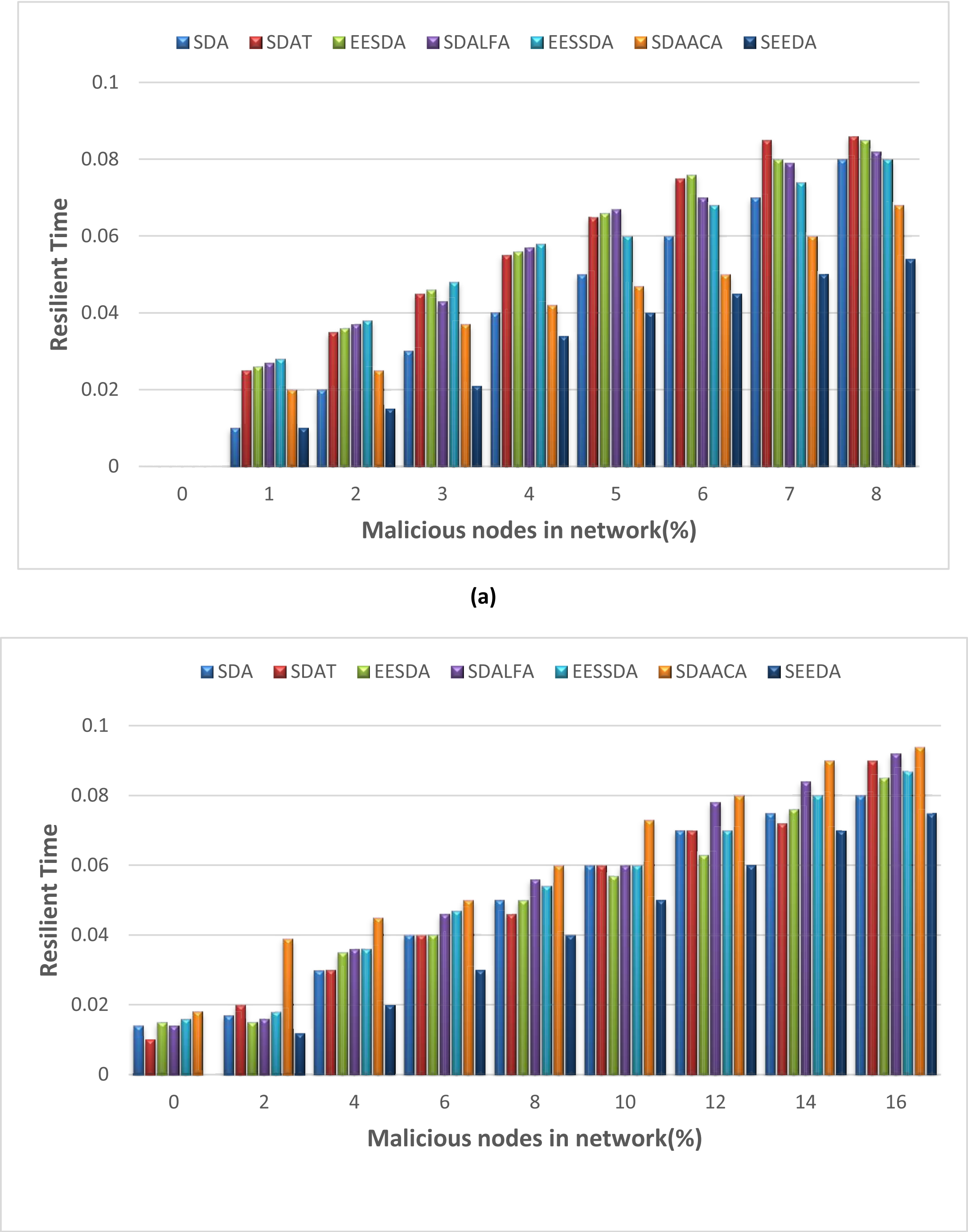


FIGURE 7. **End to end delay of our proposed SEEDA.**

to further distance node which requires more energy, distance information allows efficient task distribution among available nodes. This can reduce the delay in the network and also able to conserve energy.

A timestamp is a sequence of characters or encoded information identifying when the event occurred. The timestamp is used to calculate the time it takes to transmit data from nodes to the base station. This transmission time is



(a)

(b)

FIGURE 8. **(a) Resilient time when 0% to 8% malicious nodes affected a network, (b) Resilient time when 0% to 16% malicious nodes affected a network.**

recorded and is used as a signature to determine the legitimacy of a node. To strengthen the security, the timestamp is encoded with a random number to produce a random timestamp. A random timestamp is proposed so that the malicious node will not be able to duplicate the data packet even though they know the legitimate node typical events occurrence.

The base station nodes check the distance and the timestamp of all the nodes broadcasted by the cluster head node. If they are different from the recorded value, they can be regarded as an adversary. The adversary can also be detected by comparing the transmission time and distance. Usually, faraway nodes would have higher transmission time compared to nearer nodes.

TABLE 3. The average outcome of energy consumption comparison with different number of nodes and area.

Total nodes	Deployed area	Malicious nodes %	Other protocols	SEEDA protocol
100	400×400 m²	10%	3.34-3.9 joules	2.45 joules
		20%	3.66-3.96 joules	2.87 joules
		30%	3.57-4.0 joules	3.32 joules
250	400×400 m²	10%	3.5-4.0 joules	2.51 joules
		20%	3.71-3.89 joules	2.90 joules
		30%	3.60-4.0 joules	3.4 joules
350	1000×1000 m²	10%	3.5-4.0 joules	2.48 joules
		20%	3.68-3.98 joules	2.89 joules
		30%	3.61-4.0 joules	3.36 joules
400	1000×1000 m²	10%	3.6-4.0 joules	2.48 joules
		20%	3.73-3.96 joules	2.86 joules
		30%	3.56-4.0 joules	3.4 joules

VIII. CONCLUSION

In this paper, we propose SEEDA protocol which is secure and energy-efficient data aggregation for WSN using an access control model. The proposed protocol enhances the authentication of MAC by generating a random value and random timestamp with a secret key. The base station node verifies the fake aggregated data before sending it to the server. Other than that, the proposed protocol detects and prevents attacks such as Sybil and sinkhole. The base station nodes also utilize the distance and timestamp between nodes to avoid delay in the network. These operations reduce the redundant data transmission, energy consumption, and help prolong the network lifetime. Our protocol consists of three algorithms, namely, data fragmentation, secure node authentication, and fully homomorphic encryption algorithms. The data fragmentation algorithm partition the data into small pieces before transmitting them to the next hop nodes to hide them from being attacked. The secure node authentication algorithm checks the authentication of the node that is leaving or joining the network to prevent tampering or interrupting the data transmission between nodes. The fully homomorphic encryption algorithm encrypts the aggregated data before sending it to the base station nodes. Additionally, the proposed protocol reduces energy consumption using an access control model by reducing redundant data transmission and communication overhead. The simulations result shows that our proposed protocol outperforms other prior protocols in terms of security and energy usage. In the future, the improvement of the proposed protocol will focus on preventing more attacks and solving more challenges especially when involving mobile nodes.

REFERENCES

[1] H. Li, K. Li, W. Qu, and I. Stojmenovic, "Secure and energy-efficient data aggregation with malicious aggregator identification in wireless sensor networks," *Future Gener. Comput. Syst.*, vol. 37, pp. 108–116, Jul. 2014.

[2] A. Razaque and S. S. Rizvi, "Secure data aggregation using access control and authentication for wireless sensor networks," *Comput. Secur.*, vol. 70, pp. 532–545, Sep. 2017.

[3] N. Alsaedi, F. Hashim, A. Sali, and F. Z. Rokhani, "Detecting sybil attacks in clustered wireless sensor networks based on energy trust system (ETS)," *Comput. Commun.*, vol. 110, pp. 75–82, Sep. 2017.

[4] T. A. Zia and M. Z. Islam, "Communal reputation and individual trust (CRIT) in wireless sensor networks," in *Proc. Int. Conf. Availability, Rel. Secur.*, Feb. 2010, pp. 347–352.

[5] P. Kumar, A. Gurtov, J. Iinatti, M. Sain, and P. H. Ha, "Access control protocol with node privacy in wireless sensor networks," *IEEE Sensors J.*, vol. 16, no. 22, pp. 8142–8150, Nov. 2016.

[6] H. Hu, Y. Chen, W.-S. Ku, Z. Su, and C.-H. J. Chen, "Weighted trust evaluation-based malicious node detection for wireless sensor networks," *Int. J. Inf. Comput. Secur.*, vol. 3, no. 2, pp. 132–149, 2009.

[7] K. Hsu, M.-K. Leung, and B. Su, "Security analysis on defenses against Sybil attacks in wireless sensor networks," *IEEE J.*, to be published.

[8] V. Jariwala, H. Patel, P. Patel, and D. C. Jinwala, "Integrity and privacy preserving secure data aggregation in wireless sensor networks," *Int. J. Distrib. Syst. Technol.*, vol. 5, no. 3, pp. 77–99, 2014.

[9] X. Li, D. Chen, C. Li, and L. Wang, "Secure data aggregation with fully homomorphic encryption in large-scale wireless sensor networks," *Sensors*, vol. 15, no. 7, pp. 15952–15973, 2015.

[10] C.-M. Chen, Y.-H. Lin, Y.-C. Lin, and H.-M. Sun, "RCDA: Recoverable concealed data aggregation for data integrity in wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 4, pp. 727–734, Apr. 2012.

[11] S. Ozdemir and H. Cam, "Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 736–749, Jun. 2010.

[12] K. N. Raja and M. M. Beno, "Secure data aggregation in wireless sensor network-Fujisaki Okamoto(FO) authentication scheme against Sybil attack," *J. Med. Syst.*, vol. 41, no. 7, p. 107, 2017.

[13] G. Santhi and R. Sowmiya, "A survey on various attacks and countermeasures in wireless sensor networks," *Int. J. Comput. Appl.*, vol. 159, no. 7, pp. 7–11, 2017.

[14] H. Shafiei, A. Khonsari, H. Derakhshi, and P. Mousavi, "Detection and mitigation of sinkhole attacks in wireless sensor networks," *J. Comput. Syst. Sci.*, vol. 80, no. 3, pp. 644–653, 2014.

[15] H. Li, K. Lin, and K. Li, "Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks," *Comput. Commun.*, vol. 34, no. 4, pp. 591–597, Apr. 2011.

[16] K. A. Mehr and J. M. Niya, "Security bootstrapping of mobile ad hoc networks using identity-based cryptography," *Secur. Commun. Netw.*, vol. 9, no. 11, pp. 1374–1383, 2016.

[17] J.-H. Cho, K. S. Chan, and I.-R. Chen, "Composite trust-based public key management in mobile ad hoc networks," in *Proc. 28th Annu. ACM Symp. Appl. Comput.*, 2013, pp. 1949–1956.

[18] L. Zhu, Z. Yang, J. Xue, and C. Guo, "An efficient confidentiality and integrity preserving aggregation protocol in wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 10, no. 2, 2014, Art. no. 565480.

[19] K. T.-M. Tran, S.-H. Oh, and J.-Y. Byun, "Well-suited similarity functions for data aggregation in cluster-based underwater wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 9, no. 8, 2013, Art. no. 645243.

[20] R. B. Manjula and S. S. Manvi, "Cluster based data aggregation in underwater acoustic sensor networks," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2012, pp. 104–109.

[21] H. Çam, S. Özdemir, P. Nair, D. Muthuavinashiappan, and H. O. Sanli, "Energy-efficient secure pattern based data aggregation for wireless sensor networks," *Comput. Commun.*, vol. 29, no. 4, pp. 446–455, 2006.

[22] K. Wu, D. Dreef, B. Sun, and Y. Xiao, "Secure data aggregation without persistent cryptographic operations in wireless sensor networks," *Ad Hoc Netw.*, vol. 5, no. 1, pp. 100–111, 2007.

[23] S.-I. Huang, S. Shieh, and J. D. Tygar, "Secure encrypted-data aggregation for wireless sensor networks," *Wireless Netw.*, vol. 16, no. 4, pp. 915–927, 2010.

[24] Y. Zhang, L.-N. Zhu, and L. Feng, "Key management and authentication in ad hoc network based on mobile agent," *J. Netw.*, vol. 4, no. 6, pp. 487–494, 2009.

[25] H. Yeh, T. Chen, P. Liu, T. Kim, and H. Wei, "A secured authentication protocol for wireless sensor networks using elliptic curves cryptography," *Sensors*, vol. 11, no. 5, pp. 4767–4779, May 2011.

[26] H. Zhong, L. Shao, J. Cui, and Y. Xu, "An efficient and secure recoverable data aggregation scheme for heterogeneous wireless sensor networks," *J. Parallel Distrib. Comput.*, vol. 111, pp. 1–12, Jan. 2018.

[27] A. Albakri, L. Harn, and S. Song, "Hierarchical key management scheme with probabilistic security in a wireless sensor network (WSN)," *Secur. Commun. Netw.*, vol. 2019, Jul. 2019, Art. no. 3950129.

[28] S. B. Othman, A. Trad, H. Youssef, and H. Alzaid, "Secure data aggregation with MAC authentication in wireless sensor networks," in *Proc. 12th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Jul. 2013, pp. 188–195.

[29] S. Roy, M. Conti, S. Setia, and S. Jajodia, "Secure data aggregation in wireless sensor networks: Filtering out the attacker's impact," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 681–694, Apr. 2014.

[30] T. Wang, X. Qin, and L. Liu, "An energy-efficient and scalable secure data aggregation for wireless sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 9, no. 12, 2013, Art. no. 843485.

[31] S. B. Othman, A. Trad, H. Alzaid, and H. Youssef, "Secure and energy-efficient data aggregation for wireless sensor networks," *Int. J. Mobile Netw. Des. Innov.*, vol. 5, no. 1, pp. 28–42, 2013.

[32] M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks," *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 1, pp. 98–110, Feb. 2015.

[33] K. Haseeb, N. Islam, A. Almogren, I. U. Din, H. N. Almajed, and N. Guizani, "Secret sharing-based energy-aware and multi-hop routing protocol for IoT based WSNs," *IEEE Access*, vol. 7, pp. 79980–79988, 2019.

[34] D. Qin, Y. Zhang, J. Ma, P. Ji, and P. Feng, "A distributed collision-free data aggregation scheme for wireless sensor network," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 8, 2018, doi: 10.1177/1550147718795847.

[35] M. Wazid, A. K. Das, S. Kumari, and M. K. Khan, "Design of sinkhole node detection mechanism for hierarchical wireless sensor networks," *Secur. Commun. Netw.*, vol. 9, no. 17, pp. 4596–4614, 2016.

AHMED ABDULHADI JASIM received the bachelor's degree in computer science from the University of Baghdad, and the master's degree in computer science from the University of Malaya, where he is currently pursuing the Ph.D. degree with the Faculty of Computer Science and Information Technology. His current research interests include the area wireless sensor networks and security systems.

MOHD YAMANI IDNA BIN IDRIS received the Ph.D. degree in electrical engineering and has vast experience in research. He is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya. His expertise is in the area of the sensor networks, security systems, and signal/image processing.

SAADIAL RAZALLI BIN AZZUHRI received the Ph.D. degree from the University of Queensland, Australia, in wireless network system, specializing in wireless ad-hoc routing protocol. His current research focusing in microfiber laser, wireless network protocols, blockchain, and autonomous unmanned aerial vehicle (UAV).

NOOR RIYADH ISSA received the bachelor's degree in computer science from the University of Baghdad, and the master's degree in computer science from the Faculty of Computer Science and Information Technology, University of Malaya. Her current research interests include image processing and sensor networks.

NOORZAILY BIN MOHAMED NOOR received the bachelor's and master's degrees from the University of Malaya. He is currently a Senior Lecturer with the Faculty of Computer Science and Information Technology, University of Malaya. His current research interests include energy, mathematic and materials science.

JAGADEESH KAKARLA received the Ph.D. degree in computer science and engineering from the National Institute of Technology (NIT), Rourkela, India. He is currently an Assistant Professor with the Indian Institute of Information Technology, Design & Manufacturing (IIITDM), Chennai, India. His current research is focusing in wireless sensor networks, ad-hoc networks, and the IoT.

IRAJ SADEGH AMIRI received the M.S. and Ph.D. degrees in physics from the University of Technology Malaysia (UTM). He is currently a Senior Researcher in photonics with Ton Duc Thang University (TDTU), Vietnam. He has authored or coauthored more than 100 international peer-reviewed journals. His main research interests include modeling plasmonic photonics devices plasmonics photonics devices, optical wireless networks, and ad-hoc networks.