Archives of Computational Methods in Engineering (2019) 26:1007–1027
https://doi.org/10.1007/s11831-018-9267-2

ORIGINAL PAPER

Secure Data Aggregation Techniques for Wireless Sensor Networks: A Review

D. Vinodha[1,2] · E. A. Mary Anita[2]

Received: 5 October 2017 / Accepted: 27 April 2018 / Published online: 3 May 2018
© CIMNE, Barcelona, Spain 2018

Abstract
Wireless sensor networks (WSN) are made up of energy constraint tiny sensing devices which are distributed geographically to monitor inhabited remote areas by collecting the physical phenomenon like temperature, pressure etc. They play a vital role in military surveillance, environment monitoring etc. Unstructured topology in WSN results in large amount of redundant data being transmitted over the resource constraint devices which leads to energy starvation problem. Since the nodes are prone to tamper, thanks to their environment, ensuring the privacy of sensitive data being aggregated and transmitted is important. Hence data aggregation schemes which minimize the data redundancy with the guarantee of security become the attraction of research. Many secured aggregation schemes have been proposed by researchers. In this survey the various existing solutions are surveyed and an attempt is made to classify them based on the node topology and mechanisms employed for assuring privacy.

1 Introduction

Wireless Sensor network (WSN) is the boon of latest technology which is capable of accumulating various phenomenon of the environment like temperature, humidity, pressure, speed, pollution, by using sensing devices distributed geographically. Hence WSN becomes vital for monitoring areas which are hazardous and inaccessible for human. Their applications in habitat monitoring, weather forecasting, military surveillance are enormous. The miniature nature of sensor devices with less memory capacity, poor processing power and low battery resource makes the network suffer from resource constraints. The research shows that most of the energy of miniature nodes are depleted due to transmission compare to computation. Hence it is important to minimize the communication in WSN. But the dense deployment of sensor nodes leads to redundancy in the data being sensed and transferred and adds to the communication overhead. Hence the techniques for minimizing the redundancy in the data transmission which in turn reduce the communication costs are inviting more attention in the recent research. Data aggregation is one such a technology which minimizes or removes the redundant data by aggregating multiple data packet with superfluous information into one packet.

Since most of the WSNs are deployed in unsecured area, they are prone to many attacks. Also the adhoc nature makes the routing of WSN as data centric. Hence there is a chance of data being tapped by the adversaries in the en route to the base station by monitoring the communication line and do threat full activities. So ensuring confidentiality of the data is must. Encryption is one of the mechanisms for confidentiality which converts the data into unreadable form before transmitting. In hop by hop data aggregation, the intermediate aggregator which receives the encrypted data from multiple sensors, decrypt and do aggregation before encrypting and transmitting further. Hence by compromising the intermediate sensor node, the confidential data may get accessed by adversaries. So to ensure end to end confidentiality, many aggregation schemes employs privacy homomorphism (PH) which allows arithmetic operation to be carried out on encrypted data such that the result is same as if the operation is done on the plain text.

✉ D. Vinodha
vinodha@saec.ac.in

E. A. Mary Anita
drmaryanita@saec.ac.in

[1] Anna University, Chennai, India

[2] CSE, S.A. Engineering College, Chennai, India

This enables the intermediate aggregator node to do the aggregation on the encrypted instead of plain data.

In PH based system, the adversaries can make modification in the end to end concealed data which may have malicious effect on the final result obtained by the Base station (BS). The existing proposed solutions ensure the integrity of the data, by doing verification either in centralized or in scattered manner. In centralised verification, only BS will take the responsibility of verification. It frees all the other nodes from the computation incurred due to verification process and allows the malicious data to travel all on the way to BS which leads to unnecessary consumption of the bandwidth. Detection of false data by BS may also lead to the rejection of other valid data. This rejection causes wastage of energy. Hence verifying the integrity of the data in the intermediate nodes and filtering out the malicious data at the earliest saves the bandwidth. But it in turn increases the number of computations at the intermediate node. So there is a need to maintain a balance between bandwidth consumption and computation. And the chain of privacy homomorphic encryption applied on the en route should not increase the noise level of data. This has notable impact on the accuracy of the data. The adversaries may also affect the privacy by generating packets resembling the valid data which may go undetected during integrity verification. These packets can be filtered out by ensuring the authenticity at the intermediate aggregators and at the Base station.

Also in real time, sensor nodes sensing different physical phenomenon are deployed in the same region. Application wise aggregation in these circumstances will increase the communication if the number of applications is more. And if the normal aggregation is applied over these sensors, the data of multiple applications get mixed up and retrieving application specific data at BS will not give the original data. By making the aggregation to support multiple applications, the bandwidth consumption can be minimised.

The another factor called data recovery which enables BS to recover individual sensor data from aggregate supports dynamic query which in turn reduces the communication. Hence securing the aggregated data against attacks with the support of confidentiality, integrity authenticity, multiple application and dynamic recovery becomes predominant research. In this survey, an attempt has been made to analyse the privacy factors addressed by the existing privacy preserving data aggregation schemes and their performance. Our observations have been organised as follows. In Sect. 2, the classification of existing schemes is presented. In Sect. 3, the various attack models are discussed. Section 4 explores the privacy homomorphic encryption schemes. Section 5 gives comparative analysis of Symmetric Privacy homomorphism based data aggregation schemes. Section 6 gives comparative analysis of Asymmetric Privacy homomorphism based data aggregation schemes. Section 7 explores the non PH based secured data aggregation schemes. Section 8 presents the conclusion of our survey.

2 Classifications

2.1 Classification Based on Network Model

We classify the existing secured data aggregation schemes based on their network model. Network model plays an important part in deciding the role of sensor nodes and the communication path between the nodes and BS. Some of the existing schemes [10, 15, 16] organized their nodes as a tree which is rooted at BS. The intermediate parent nodes play the role of aggregator and enable the possibility of multi hop aggregation en route to the BS. In tree topology (Fig. 1), the route between the sensor and BS is fixed and unique. Hence dynamic aggregator selection is not fully supported which makes the aggregator gets overloaded and its energy may get depleted overtime. Since the route is fixed, there is also the possibility of data loss. To avoid the data loss, some of the schemes [14, 17] adopt ring topology or layered model (Fig. 2) by grouping the nodes into one layer or ring based on their hop count from the BS. The nodes of layer x send their packets to any one of the nodes of layer $x - 1$ in their transmission range. Thus the nodes have multiple parents and multi paths exist between nodes and BS. This multi path enables the BS to handle data loss. In clustered model (Fig. 3), nodes are grouped into clusters and one of the nodes inside the cluster will act as a head and all the other cluster members communicate via cluster head. To prevent the cluster head (CH) from energy starvation, cluster head role can be shared dynamically among

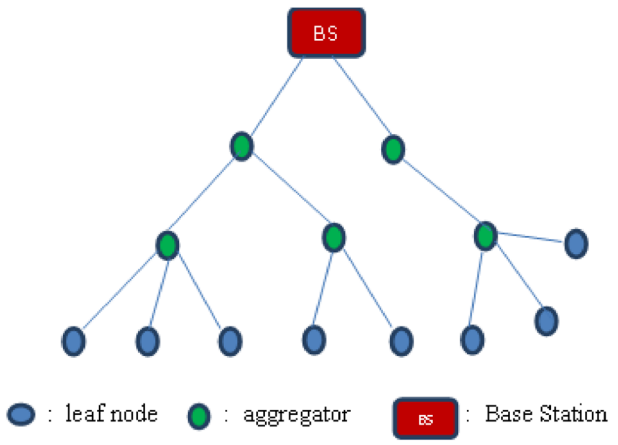

Fig. 1 Tree topology

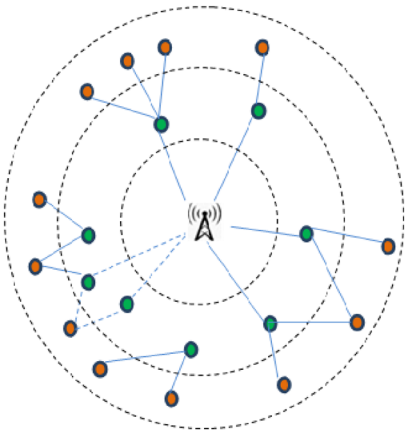

Fig. 2 Ring topology

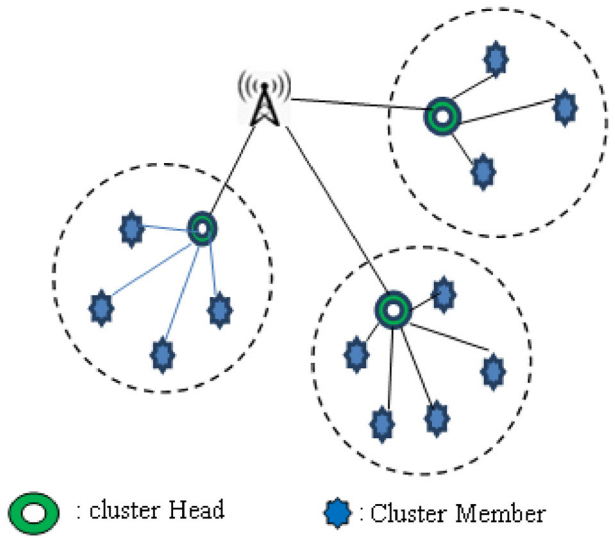

Fig. 3 Cluster topology

the cluster members. In hybrid model, the clustered and tree topology are combined together and the benefits of both the models are enjoyed. Comparison of this three network model is given in Table 1.

2.2 Classification Based on Keys

The aggregation schemes can again be classified based on how data is secured from adversaries either by employing encryption or not. With encryption, the original data is enciphered and the data is travelling in non understandable format.

In [27], the encrypted data is decrypted and aggregation is applied at every hop and again transformed into non understandable form using encryption. This process is repeated en route to BS. Since the sealed data is opened at every hop, the data can be easily revealed by compromising the intermediate aggregator. And it also consumes the processor cycle and time. Thus hop by hop encryption suffers from security threat and computation overhead. To overcome this, [1–3, 5–10] schemes conceals data from the source to destination by employing Privacy Homomorphism (PHM). PHM applies the aggregation operation over the encrypted data instead of plain data such that the decrypted aggregation has the same result as if it is applied on the plain text.

$$\mathrm{De}(\mathrm{f}(\mathrm{En}(\mathrm{x}),\mathrm{En}(\mathrm{y}))) = \mathrm{f}(\mathrm{x},\mathrm{y})$$

where De is decryption function, En is encryption function, x, y are plain text, f is aggregation function.

The nature of homomorphic function preserves the relationship of x and y between the encrypted value. But in PHM, the adversaries who are familiar with the structure of the plain text can modify the encrypted text into another valid encrypted text and cause malleability. And successive application of homomorphic function increases the noise level in the aggregated data which limits the supported arithmetic operation and makes the decryption fail. Table 2 gives the comparison of end to end concealed encryption and hop by hop schemes.

Further based on the kinds of operations supported, PHM schemes can also be grouped as additive PHM, multiplicative PHM and fully homomorphic. Additive PHM supports only addition operation where as multiplicative PHM allows only multiplication over encrypted data. But fully homomorphic becomes more flexible by

Table 1 Comparison of network model

Tree topology	Cluster topology	Ring topology
Each node has single parent	Each node has single parent	Each node has multiple parent
Multi hop communication	Two hop communication	Multi hop communication
Single path between nodes and BS	Single path between nodes and BS	Multipath between nodes and BS
High probability of communication loss	High probability of communication loss	Less probability of communication loss
Difficult to share the aggregator role dynamically	Supports dynamic sharing of aggregator role	Supports dynamic sharing of aggregator role
Comparatively less construction cost	Comparatively high construction cost	Comparatively less construction cost

Table 2 Comparison of hop by hop encryption schemes with end to end encryption

Hop by hop encryption	End to end encryption
Ciphering and deciphering takes place at every hop	Enciphering occurs only at source node, deciphering occurs only at BS
Aggregation function is applied on plain data	Aggregation function is applied on cipher text
Since aggregation is applied on plain data, the aggregated data is more accurate	Accuracy gets affected with number of hops
Less secure	Comparatively secure
Intermediate nodes'energy get depleted because of high computation	No depletion of energy due to computation

allowing any kind of operations (either + or *. Not both). But FHE increases the computation overhead due to its large public key size. Based on the type of keys, the PHM based schemes can be further classified as Symmetric Privacy Homomorphism (SPHM), Asymmetric Privacy Homomorphism (APHM) and Hybrid Privacy Homomorphism (HPHM). Figure 4 gives the classification of existing privacy preserved data aggregation schemes.

3 Attack Model

The wireless sensor network which plays vital role in monitoring inhabited area where the human intervention is difficult, is prone to different kinds of attack which can be classified as active and passive attacks. In active attacks, the intruders may try to affect the originality of the aggregated data by inserting, deleting or modifying the data. In passive attack, the intruders try to eavesdrop secret information. Some of the attacks which affect the privacy of the data are discussed here.

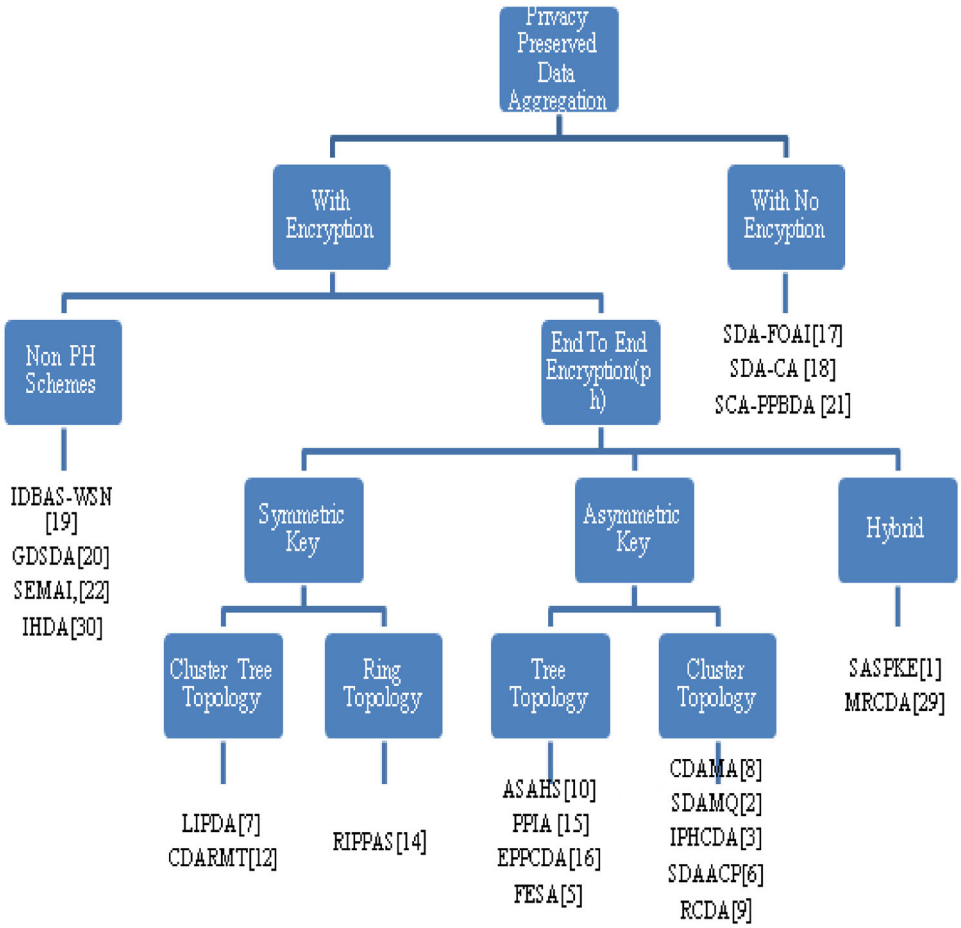

Fig. 4 Classification of privacy preserved data aggregation schemes

3.1 Confidentiality Attack (Eavesdropping Attack)

Attackers gain unauthorized access to data and keys. Known plain text, chosen plain text and chosen cipher text are example for this kind of attack. In known plain text attack, by knowing both plain text and cipher text, the attacker tries to access secret information. In chosen cipher text attack, partial information are collected by choosing encrypted text. In chosen plain text attack, cryptanalysis is done with random plain text.

3.2 Integrity Attack (Data Pollution Attack)

Unauthorized modification of actual data leads to integrity attack. By compromising the CH, the attacker can modify or inject false data into the aggregated value. It leads to propagation of false data to the base station.

3.3 Authenticity Attack

Attacker may act as a valid BS and inject false query or act as valid sensor node and inject false data.

3.4 Replay Attack

In this the already transmitted packets are retransmitted which leads to data corruption in the aggregated result received at the BS.

3.5 Sybil Attack

Attacker is able to present more than one identity within the network. An adversary can launch a Sybil attack and generate n or more witness identities to make the base station accept the aggregation results.

3.6 Collusion Attack

The secret information of the nodes can be extracted by making malicious nodes to collide with each other.

3.7 Malleability

The attacker can convert one cipher text into another valid cipher text from which a new plain text can be deciphered.

3.8 Byzantine Attack

The attacker can inject false data by compromising set of sensor nodes.

3.9 Coalition Attack

The adversary can inject invalid signature and get through the verification test by successfully generating valid aggregation signature.

3.10 Node Compromising Attack

By gaining control of aggregator node, the attacker replays the packets or selectively drops the packet which will affect the originality of the final data received by the BS.

4 Privacy Homomorphism Based Encryption Methods

4.1 Symmetric PHM Encryption Methods (SPHMs)

SPHMs use same key for both encryption and decryption. The Cipher text requires same number of bits as required by plain text. So it suffers from negligible message expansion. And computation cost is also significantly less as compared to asymmetric key based PH. Here we discussed two symmetry key based PH methods.

Domingo Ferror cryptosystem [23] proposed a scheme which is an additive and multiplicative fully homomorphic symmetric PH scheme, supports full arithmetic operations (e.g., addition, subtraction, multiplication, and division) over encrypted data and secure against chosen cipher text attack. This scheme is also proven to be secured against know plaintext attack provided the plain text is split into d values such that $d > 2$ during encryption and ciphertext space is greater than the plaintext space. Since same key is used, compromised intermediate node can decrypt the encrypted data and inject malicious data. Hence integrity is affected. Power consumption of Domingo Ferrer's system is high compared to RC5. It also suffers from threat of cryptanalysis (deciphering without using key).

Castelluccia (CMT) et al. [25] proposed an additive based PH scheme using symmetry key which is a modified version of Vernam cipher. The X-OR in Vernam is replaced by addition operation which makes it suitable for resource constrained WSN. In Vernam, keys are selected in random manner from the key space. But in CMT, keys are generated using the stream cipher and the privacy is enhanced by generating keys using a unique message identifier and the unique shared key between the node and BS. CMT cryptosystem also helps in balancing communication overhead evenly across all sensor nodes. This load balancing across WSNs eventually increases WSNs' lifetime.

4.2 Asymmetric PHM Encryption Methods (APHMs)

APHMs employ different keys for encryption and decryption. Absence of shared information between node and BS makes this more secure than SPHM. In Asymmetric key based PH, Public key based privacy homomorphism is costlier in terms of computation and communication for resource constraint WSN. It suffers from significant message expansion compare to symmetric PHM. Table 3 gives the comparison of SPHM and APHM.

4.2.1 APHM with Elliptic Curve Cryptosystems (APHM-ECC)

Elliptic curve based asymmetric PHM scheme enjoys the same level of security provided by non ECC based system with trapdoor function and smaller keys at high speed. Its strength relies on elliptic curve discrete logarithm. Finding discrete log for any random elliptic curve elements with respect to known points is intractable. That is, if p = m ° n, then finding m using the known p and n is difficult. Here we have discussed three elliptic curve systems.

4.2.1.1 Elliptic Curve Naccache–Stern's Cryptosystem ($n = pq$) In this elliptic curve cryptosystem, the strength of security resides on the intractable high degree residuosity problem. But homomorphic nature of this scheme leads to malleability. To generate keys, a group of even number of prime numbers are selected.

$$\mathrm{R} = \{r_1, r_2, \ldots . r\}_k$$

where k is even number.

Split this group into two sets.

$$\mathrm{x} = \prod_{\mathrm{i}=1}^{\mathrm{i}=\mathrm{k}/2} \mathrm{r_i y} = \prod_{\mathrm{i}=\frac{\mathrm{k}}{2}+1}^{\mathrm{i}=\mathrm{k}} \mathrm{r_i}$$

$$\mathrm{h} = \mathrm{xy}$$

Choose 2 large prime e, f such that

$$\mathrm{p} = 2\mathrm{ex} + 1 \quad \mathrm{q} = 2\mathrm{fy} + 1$$

where p, q are primes

$$\mathrm{n} = \mathrm{pq}$$

Select a random g mod n such that

$$\mathrm{g} = \varphi(\mathrm{n})/4$$

Parameters (h, n, g) are made public key and p and q are kept secret.

Encryption The plain text m is encrypted

$$E(m) = d^h g^m \bmod n$$

where d is random number

Decryption For each r_i, m mod r_i is calculated as follows

$$c_i \equiv c^{\frac{(\phi(n)}{r_i}} \; mod \; n$$

where c is encrypted text

Hence

$$c_i \equiv d^{\frac{h(\phi(\mathrm{n})}{r_i}} g^{\frac{m(\phi(\mathrm{n})}{r_i}} \; mod \; n$$

$$c_i \equiv g^{(m_i + y_i p_i)(\phi(\mathrm{n})/r_i} \; mod \; n$$

$$c_i \equiv g^{m_i(\phi(\mathrm{n})/r_i} \; mod \; n$$

where

$$m_i = \; m \; mod \; r_i$$

m_i requires exhaustive search by choosing the minimum r_i. From m_i, the m can be recovered by Chinese reminder theorem.

4.2.1.2 Elliptic Curve Okamoto–Uchiyama's Cryptosystem (OK–UC)(EC-OU) ($n = p^2q$) It is secure against intractability of high degree residuosity problem. The most powerful algorithm available for integer factorization is the number field sieve method, and the complexity of the number field sieve method is dependent on the size of a composite number to be factored. Hence, elliptic curve Okamoto–Uchiyama's cryptosystem can use prime numbers p and q of size 341bit ($\approx n = p^2q$ of size 1024-bit) to achieve the equivalent security level as provided by the 1024-bit RSA cryptosystem.

Given two cipher texts E (m0), E (m1) and their corresponding plaintexts m0 and m1, a polynomial-time attacker with the privilege to access the public parameters of the Okamoto–Uchiyama cryptosystem is able to randomly guess the correct cipher text linked to m0 with probability 1/2.

The Okamoto–Uchiyama cipher texts have public randomization. This property states that (1) ciphertexts can be randomized by merely using the public key (2) given two randomized ciphertexts E (m0, r0), E (m1, r1) and their corresponding plaintexts m0 and m1, an attacker is able to guess the randomized cipher text linked to m0 with probability 1/2.

Key Generation In this scheme $n = p^2q$, Where p, q are large prime numbers.

g is selected such that $g^{p-1} \neq 1 \; mod p^2$

$$h = g^n \; mod \; n$$

(n, g, h) are made public and p, q are kept secret.

Encryption

$$E(m) = h^r g^m \; mod \; n$$

Table 3 Comparison oF SPHM and ASPHM

SPHM	APHM
Shared secret information make this system less secure	Highly secure. Because no secret is required to be shared between receiver and sender for secured communication
No significant message expansion problem	Suffers from message expansion problem
Less computation and communication overhead	Dependency on complicated mathematical operations leads to high computation and communication overhead
Requires storage at intermediate nodes	Not required
Requires significant key distribution mechanism	Not required

where m is data and r is the random factor

Decryption

Let D = E(m)

$D_p = D^{p-1} \; mod \; p^2$

$L(D_p/g_p) modp$

This technique is based on logarithmic function defined over the sub group of $(\mathbb{Z}/n^2\mathbb{Z}^*)$. It is semantically secure and its strength lies on the intractability of p subgroup problem.

4.2.1.3 Elliptic Curve Paillier's Cryptosystem [13] ($n = p \cdot^2 q \cdot^2$)

Paillier proposed three schemes based on Naccache–Stern's cryptosystem (n = pq), Okamoto Uchiyamas cryptosystem ($n = p^2q$) and third scheme which is the extension of the previous two to Paillier cryptosystem. All these schemes which are additive homomorphic ensure semantic security and secure against chosen cipher text attacks (CCA2). Their security strength relies on the intractability of high degree residuosity problem. The major issues of the above three schemes are message expansion problem. The message expansion increases the communication overhead and eventually reduces the lifespan of WSNs. Despite the significant communication overhead, Paillier's cryptosystems are vulnerable and allow the secret key to be recovered from the publicly available information.

4.2.2 EC-ElGammal Cryptosystems [34]

It is based on an intractability of solving the Elliptic Curve Discrete Logarithm Problem (ECDLP). It is defined on an elliptic curve over a finite field F. Therefore, EC-ElGamal cryptosystem (EC-EG) requires only 160-bit key-size to achieve the same level of security as provided by the 1024-bit RSA cryptosystem. The reduction in the size of elliptic curve parameters improves bandwidth efficiency, energy utilization, and storage capabilities in WSNs. The EC-ElGamal cryptosystem has at least a 4 to 1 message expansion ratio, where a plaintext is converted into two cipher texts, and each cipher text has at least two coordinate values. The point compression technique [35] can efficiently compute a value of their coordinate from an x-coordinate value and with the help of an additional sign bit. Therefore, the EC-ElGamal cryptosystem requires only two extra bits as compared to the ElGamal cryptosystem. Mapping function (which maps the plain text to elliptic curve) needs to be homomorphic and deterministic such that it can uniquely associate each plaintext value to an elliptic curve point and vice versa. It should be independent from the encryption and decryption operations of the EC-ElGamal cryptosystem.

4.3 Hybrid PHM Encryption Methods

It is the combined version of SPHM and APHM. It enjoys the benefit of both the schemes. It addresses the message expansion problem of APHM by encrypting large messages by symmetric key while achieving the highest security by asymmetric key using public key cryptosystem.

5 SPHM Based Secured Aggregation Schemes

5.1 Lightweight and Integrity-Protecting Oriented Data Aggregation Scheme: LIPDA [7]

In this the authors proposed a method to ensure the end to end confidentiality and integrity by structuring the data as a complex number using private key and by using additive homomorphic encryption. So the computational and communication overhead is minimized with high accuracy and fullness compare to elliptic curve based homomorphic encryption. Cluster Tree based network model with base station as root node is used for aggregation. By propagating a HELLO packet from base station and by calculating the

probability of a node to become a cluster head based on distance, the hierarchical cluster tree is formed. Each cluster is assigned a cluster number (ID) in the order of cluster formation. Once the cluster is formed, the sink node generates clusters of key pairs and they are broadcast along with the cluster number using RC4 algorithm which ensures the secure distribution. The cluster head which receives this key pair checks the belonging of the key using the cluster number. If the cluster number is not matching, it is rebroadcast. Hence, there is a great threat of leakage of the keys received by the compromised cluster head. One of the key in the pairs, is used to form the complex number and another key is used to encrypt this complex number. The imaginary part of the complex number is used for verifying the authenticity of the data. Each cluster head aggregates the encrypted complex number using additive PHM and forward it to its parent with its cluster ID.

Integrity is verified only at base station. The BS separates the aggregated encrypted data from the aggregated privacy factor. Since the privacy factors of all the clusters are generated and available with BS, the BS can now form the new aggregation of all the participated clusters and compared with the received privacy factor. If the difference is less than the threshold, the aggregated data is accepted or else it is rejected. To prevent CH from energy depletion due to overload, CH role is shared dynamically among the cluster members. When the energy level of the CH becomes below the threshold value, HELLO packets are transmitted and each node which receives this packet calculates the distance. Each node transmits this distance and the remaining energy level to the cluster head. The CH will select the appropriate node as new CH based on the distance and available energy.

Though the author's claim that the distribution of the privacy factor from the sink node is safe because of the RC4 algorithm, the key pairs are visible to all cluster head on the way to the destination clusters. So even if any one of the cluster head is compromised, the attackers easily gain the keys of all clusters whose key pairs are moving via the compromised node. The authors also claim that because of RC4, retrieving the privacy data from the sensor node by querying is impossible. So by probability the LIPDA is more secure than the IPDA [36]. Also even if the raw data is obtained by colluding the neighbor nodes, deducing the correct data is difficult because of the RC4. Since the integrity is verified using the complex structure of the data, tampered data can be easily detected.

5.2 Ring-Based Privacy-Preserving Aggregation Scheme: RiPPAS [14]

The authors proposed a method using penname (pseudonyms) mechanism with anonymous communication by hiding source node identification and applied symmetric homomorphic encryption technique to secure data.

This method supports all kinds of queries like sum, min and max. Nodes are organized with layered topology with sink as the root node and each node may have more than one parent. Nodes with same number of hop are grouped into one layer. The root node decides the unique key, penname codes for each node and loads the nodes with this detail. Before the aggregation process, each node sends its location in encrypted form to the sink and it is stored in the table along with the key and penname. The source node selects a random pen name and sends that with the encrypted data. These pennames enable the sink to identify the source node of the received data and identify the key corresponding to that node for decrypting the data. The reliability of this method depends on the uniqueness of the pennames for each node. No two nodes should have common pennames. Each pair of nodes also share an independent unique key with each other for secure communication between them. This method proposed two different data sending scheme. One is Cipher text unicasting and another one is plaintext broadcasting without source ID which facilitates anonymous communication.

In cipher text unicasting, each node x selects the parent node y randomly and encrypt the data using the shared key between x and y. The receiving node decrypts and performs aggregation. Thus hop by hop encryption is carried in the en route to sink. In the plaintext broadcasting, the data is broadcast in a packet which carries only the receivers' address not the senders ID. Using symmetric homormorphic encryption the data is aggregated in the en route to sink. Each node maintains the count of receiving packet to ensure the reception of packets from all its successor nodes. Since authentication is not verified, there is a chance of duplicate packet or falsified packet which will mislead the data aggregator. Because of the randomness in the parent and anonymous source node, it is very difficult for the intruders to do long term analysis or breaking the down/up link of an intermediate data aggregator and extract information of pennames without compromising. Thus it provides robust privacy with less communication overhead.

5.3 CDA-RMT [12]

The authors proposed a method which exploits the additive homomorphic nature of privacy homomorphism and provides end to end privacy. This scheme uses additively PHM based on Domingo Ferror which provides security against chosen ciphertext attack. This frees the intermediate aggregate nodes from doing decryption, encryption and storing sensitive information. Hence any node can be elected to play the role of aggregator. But Wagner has shown that this is vulnerable against chosen plain text

attack. Impact of node compromising attack is reduced by proposing topology aware key pre distribution algorithm. Keys are distributed to subgroup of nodes which form the reverse multicast routable region. This group keying restricts the impact of compromised node to sub group of nodes and prevents the failure of entire network. But if any one node is compromised during bootstrapping phase, the concealment of whole network will be broken.

5.4 Comparison of SPHMs

The existing SPHMs provide secured data aggregation with comparatively less communication overhead and no significant message expansion problem. Sharing of Symmetric key between source and destination requires reliable key distribution method. The PH mechanism makes these methods secure against eavesdropping attack. In LIPDA, thanks to RC4 the information leakage through collusion attack is limited and integrity verification at BS has made this method secure against data pollution attack. The security against eavesdropping attack is strengthened in RiPPAS by adding some randomness where as in CDA-RMT the impact of compromised node attack is reduced. The originality of the data is not verified in all SPHMs except LIPDA. Comparison of the SPHM based data aggregation schemes is given in Table 4.

6 APHM Based Data Aggregation Schemes

6.1 Asymmetric Homomorphism for Secure Aggregation in Heterogeneous Scenarios: ASAHS [10]

The authors proposed a scheme which is secure, scalable and many to one kind of communication using asymmetric encryption. Nodes are organized as tree rooted at BS. The BS sends a symbol array where each array index represents the symbol for different range of values to be sent by the sensor. The sensor nodes with data to be sent increment the corresponding array element which is the range of the current value to be sent. And this array is forwarded to the aggregator. Value of each array element refers the number of sensors contributed the value. And also there is an array index for null value, which refers the number of non contributing nodes. This concept keeps the size of the message moving from nodes to BS constant and makes the network scalable. It enables the BS to apply different aggregation function. Confidentiality is ensured by encrypting the array element using OK–UC which is probabilistic based additive, asymmetric homomorphism. The BS collects information about participating nodes by receiving acknowledgment for the broadcasted hello

Table 4 Comparison of SPHM based data aggregation schemes

Scheme	Encryption/ method	Network model	Dynamic aggregator selection	Attacks addressed	Authentication	Integrity	Scalability	Multiple application	Data recovery
LIPDA [7]	Additive PH using complex number	Clustered tree model	Supported	Eavesdropping attack, collusion attack, data pollution attack	No	Integrity is verified at BS using the imaginary part of the ciphered complex number	No	No	No
RiPPAS [14]	Symmetric PH with pseudonym	Layered architecture/ ring topology	NA	Eavesdropping attack	No	No	No	No	No
CDA-RMT [12]	Symmetric additive PH-with Domingo Ferror	Clustered tree model	Supported	Leakage of limited information by compromising node. Chosen ciphertext attack	No	No	No	No	No

NA not applicable

message. The BS generates OK–UC parameters x, g, h, k, p, e. Among these (g, h, k, p, e) is kept secret. x is known to all nodes in the tree. Each node i has

1. public key x
2. secret key K_i shared between node and BS
3. cipher text $w_i = gh^{IDi}$ where ID_i is the random number used to identify the nodes by BS
4. a separate secret key shared between nodes and its parent

Each array element is initialized

$$I(s) = g^{\alpha_s} h^{\beta_s} \bmod x$$

where

α_s *is random integer such that* $\alpha_s + n < p$

where n is total number of leaves

β_s *is random integer with no restriction*

Each node generates a_i using the k_i

Each cipher text element in the received array is modified using a_i

$$I(s)^{a_i} = g^{\alpha_s . a_i} h^{\beta_s . a_i} \bmod x$$

Then the index for the indented data alone is multiplied by w_i.

Intermediate aggregator node aggregates the received arrays from all children using additive homomorphic operation. This semantic security and public randomization enables the method secure against eavesdropping attack. Before the beginning of aggregation process, the BS collects the ID of all participating node by broadcasting hello message and receiving acknowledgement. Using this, the BS recalculates the aggregation of all participating nodes ID and compares it with extracted aggregation ID from the received message. If the verification succeeds it respond with acknowledgement. Each intermediate node stores the message received from its child nodes until an acknowledgement is received from BS. If consistency check fails, the BS traces the malicious node using the message and signature which are stored in the intermediate aggregator node. Thus the BS verifies the integrity of the received message.

Authenticity of all communication is verified by attaching a signature generated using a unique key shared between all the nodes and their parents. Thus the authenticity of all participating nodes and integrity of leave nodes are ensured. And the deviation of aggregated value caused by the collusion of compromised intermediate and sensor node is minimized. But this method requires the entire active node whether it is having data or not, to participate in the aggregation process by sending array of null values, this consumes the bandwidth of energy constraint sensor network.

6.2 Privacy Preserving In-network Aggregation in Wireless Sensor Networks: PPIA [15]

The authors proposed a method using Paillier cryptosystem based asymmetric homomorphic encryption scheme to resilient false data injection in data aggregation. The kind of aggregation supported are sum, count and mean. Existing solutions for false data injection by compromised node by allowing multipath routing is very complex. Hence the authors proposed this solution. The leaf node encrypts the data using the

$$\mathrm{Ency(M)} = \mathrm{rM.pq \bmod qq}$$

where M is the message to be encrypted.

$$M = mT$$

where m is sensed data and T is the threshold, p is random value, $p \varepsilon Z_q$.

The encrypted M is forwarded to the parent node in tree where it get aggregated and forwarded to its parent. This process gets repeated till the sink. Using the threshold and the private key, the sum, count and mean can be calculated by the sink. Thus the security is ensured by using homomorphic encryption and energy consumption is reduced by using the threshold.

6.3 An Efficient Privacy-Preserving Compressive Data Gathering Scheme: EPPCDA [16]

The authors proposed a secure data compression technique based aggregation using asymmetric homomorphic encryption scheme which ensures end to end confidentiality by preventing data flow analysis and flow tracing. It employs the compressive sensing technique (CST) which can recover a sparse signal using optimization problem. In CST, the raw sensed data is encoded and expanded to X-dimension vector using the measurement matrix X*Y where Y is number of nodes. And this vector is transferred upward towards the sink and gets aggregated in intermediate node in the tree based network topology. Each node gets the Measurement Matrix (MM) using seed mechanism. This seed mechanism avoids the explicit transmission of MM to all nodes and avoids communication overhead. Each node generates their corresponding column vector using the local seed and preinstalled pseudo random number. Local seed is generated using the nodes own ID and global seed is broadcast by the sink. Hence the ID of each sensor node must be kept secret between the node and sink. The sink generates the MM using the ID and extracts the raw data using greedy algorithm called orthogonal

matching pursuit. By employing end to end enciphered compressive sensing based data aggregation the energy consumption by encryption and decryption and the latency problem of hop based aggregation scheme is reduced. The traffic analysis attack is made difficult by hiding monitoring round number and resisting size correlation and content correlation analysis. Since the aggregation is based on multiplication, the size correlation analysis is made difficult. And by preventing the attackers from intercepting the packets of same monitoring round, content correlation is avoided. It also prevents the adversaries from knowing the source of message through tracing the forwarding path. This is done by keep changing the encrypted message at every intermediate node. Thus by preventing flow analysis and tracing, the brute force attack is addressed in this proposed scheme.

6.4 Secure Data Aggregation with Fully Homomorphic Encryption in Large-Scale Wireless Sensor Networks: FESA [5]

Authors proposed a scheme named FESA to provide end to end data confidentiality, MAC based integrity verification and to support multiple aggregation operation on encrypted data. Integrity can be verified at BS, during aggregation as well as forwarding phase. So propagation of false data is avoided at the earliest and hence the bandwidth/communication cost due to false data is eliminated. For early detection of false data, MFN (Mobile node, Forwarding node, Neighboring node) group of multi hop network structure with static topology is proposed in this paper. And to support fully homomorphic encryption (FHE), DGHV scheme [24] is used. Squash the decryption circuit is used for including additional information about the private key in the public key, which can be used for post processing the cipher text. Thus the bootstrapping minimizes the noise level caused by the chain of homomorphic operations and make the scheme FHE.

The nodes are grouped and each group is assigned a group key, which is used as seed for Pseudo Random Number Generator function. This function is run to find the bit position of the subMACs generated by the group member of the monitoring node and the order of the subMACs are determined by the aggregator and is informed to individual group members. Since this order is not known to the adversary, injecting false message becomes difficult. Secure data aggregator selection protocol and monitoring node selection algorithms are used for selecting aggregators and monitoring nodes. The involvement of aggregators and all neighboring nodes in the selection process of monitoring nodes adverse the impact of any compromising nodes. Using pairwise key establishment algorithm, a symmetric key is established between two consecutive aggregators which helps to detect the false data. The formation of this group network enables the forwarding as well as the neighboring nodes to verify the integrity of the aggregated data and detect false data injection.

The aggregator verifies the integrity of the received data (FHED) using MAC. If valid, broadcast to all its neighboring nodes. The neighboring nodes in turn verify the integrity by generating MAC using the group key. If verification process is successful separate aggregation is done by aggregator node as well as monitoring node. Separate MAC is calculated by each aggregator and monitoring node. All these subMACs along with FHE aggregated data are forwarded whose integrity is verified by the forwarding nodes. Because of the symmetric key, group key and probabilistic location of subMACs, it is very difficult for the adversaries to break all the encryption keys. Thus this scheme becomes more secure against cipher text based attack and plain text only attack.

Computation overhead is increased due to large public key size. Same aggregation operation is done by the aggregator as well as monitoring node. The received data from leaf aggregators are forwarded to all the neighbors for verification which consumes the bandwidth. It is the contradiction to the purpose of aggregation. The computation overhead is directly propositional to the number of Monitoring nodes. Integrity and detection of false data injection depend on the number of subMACs. But the computation and the data expansion increase with the subMACs i.e. number of monitoring nodes. So determining the balanced number of monitoring nodes is critical issue of this scheme. Since node wise data can't be extracted, dynamic queries are not supported.

6.5 Concealed Data Aggregation Scheme for Multiple Applications: CDAMA [8]

Authors proposed a solution called CDAMA which will aggregate the data of different application without mixing them up and provide a way to extract the correct aggregate value of a particular application. This employs a modified asymmetric privacy homomorphism scheme which enables aggregation of data from multiple applications. Basically CDAMA is a modification from [37] public-key PH encryption system and integrates the Paillier with the Okamoto-Uchiyama encryption schemes.

The concept of CDAMA is, it generates multiple points with different order for each application. Using this set of points, keys are generated for each application using elliptic cryptography. Only one private key is generated and kept in base station. By using appropriate combination of point orders, the base station can extract the aggregated result of different application. Though this scheme support multiple application, it prerequisites the generation of

public keys for each application and secure distribution of keys to the corresponding sensor nodes. This consumes the bandwidth of the energy constraint WSN. And only the additive value of a particular application can be extracted by the base station. The base station cannot recover individual sensor node. Hence sensor node level integrity verification is not available.

Problem of compromised aggregator nodes like repeated or selective aggregation is addressed by making the BS to receive the secure count which represents the number of messages which are aggregated. This doubles the number of groups which in turn increases the ciphered text size. Because the cipher text size $\propto (G+1) * b + 1$, where G is the number of groups i.e. 2 k (where k is number of application) and b is number of bits for prime number. It is also secure against unauthorized aggregation and malleability since the point information is kept with sensor node.

6.6 Secure Data Aggregation for Multiple Queries: SDAMQ [2]

The authors proposed a method based on additive homomorphic and elliptic curve cryptography to publish queries in authenticated manner and aggregate data belongs to multiple sum based queries into single packet so that communication cost can be reduced. The validity of the query is done by signature verification process. For this a common key is shared between the BS and all CHs.

In this system, the heterogeneous nodes are grouped into clusters and node with high capacity is elected as cluster head. This is implemented in three phases. In Phase I, the BS spreads the query. It generates unique ID and pair of public key and private key for each query. A query message consists of query specific public key along with signature and validity duration is broad cast to all CH. In Phase II, each sensor node responds to the query by generating data. After receiving the query, the sensor node checks the possibility of data contribution and generates data. This data is encrypted using the public key and signature is attached with encrypted data which is forwarded to CH. In Phase III data is aggregated by CH.

The CH verifies the authenticity and aggregate using additive homomorphic encryption based on elliptic curve. Since aggregation is done on encrypted data, confidentiality is ensured. This can be opened only with private key which is available with BS. Thus even if the CH is compromised, it cannot access the data.

Since authentication is done at each level of communication, attackers are prevented from collecting information by behaving like a BS. The timer information attached with each query prevents the replay attack. But since a common key is shared among the BS and CHs, if the CH is compromised, the attacker can easily gain access to this common key and do all mal activities. There is a need for secure methodology to distribute keys to all cluster members and cluster head.

6.7 Integrity Protecting Hierarchical Concealed Data Aggregation: IPHCDA [3]

The authors proposed a method to aggregate data encrypted using different keys by employing modified elliptic curve based additive homomorphic encryption [4]. This scheme ensures confidentiality and integrity of the data using MAC. The whole network terrain is split into different region and a separate public key is assigned to each region which is used for encrypting the data sensed in that region. These regions are hierarchically organized. An aggregated data can be classified region wise by BS using encryption key. It is a kind of multiple aggregator model. Aggregators are chosen dynamically to distribute the load evenly. Replay attack is addressed by time stamping all the data packets and attaching a nonce. Since a separate private/public key pair is assigned to each region, each sensor node is deployed with region specific public key and each CH shares unique symmetric MAC key with the BS.

To verify the integrity of the aggregated data sent by the CH, MAC is calculated by using aggregated data and symmetric key shared between BS and CH. In the hierarchical tree, the CH which receives aggregated data and MAC from each sub CH, will XOR the MAC. And this aggregated data and a single XOR is forwarded further in the tree towards BS. Since Hash-MAC [31] is used, MAC becomes un forgeable.

The BS extract region wise encrypted data using discrete log and private key and calculate new MAC for each region and they are XORed which will be compared with received MAC to ensure the integrity. Since the data is encrypted using public key as well as the random number, the probabilistic nature of ciphertext make the system resilient to chosen plain text attack. Replay attack is not directly addressed by this IPHCDA. But it may be handled by time stamping the packets. When the CH is compromised, it may inject false data which cannot be detected by BS. But the compromised node does not affect the confidentiality of the aggregated data.

The computational overhead of IPHCA is high compare to EC-EG and close to EC-OU. But both EC-EG and EC-OU will not support concealed aggregation with different encryption keys. While comparing the communication cost, the IPHCDA outperforms EC-OU and EC-EG when the number of deployment region is less than 3. But the communication cost increases with number of regions. Integrity of final aggregated data is verified only at base station. Integrity of individual sensor node is not

guaranteed. So the false data injected by malicious nodes go undetected by BS. And authentication is not verified. Its computation and communication complexity is not negligible. Since the ciphertext size is large, it cannot be sent in one time. This leads to communication overhead. This multi transmission of ciphertext packet may also leads to more noise and inaccuracy.

6.8 A Secure Data Aggregation Scheme Based on Appropriate Cryptographic Primitives in Heterogeneous Wireless Sensor Networks: SDAACP [6]

The authors proposed a scheme based on additive EC-EG homomorphic encryption to ensure end to end privacy. This scheme is developed for heterogeneous clustered sensor network organized in three layers. Top layer has BS, middle layer is made up of CH and last layer is with low power sensor nodes. And CH is assumed to tamper resistant. The BS generates parameters and they are published and get loaded in sensor nodes. The BS also keeps some master secret key for extracting the private key of ID based signature scheme and a secret key for deciphering the cipher text of EC-Elgammal. Each sensor node will get ID and private key from BS before joining the network. The BS and CH maintains this list of member IDs, which is used for verification while receiving message from its members. Each sensor node encrypts the message using the public key of BS.

The Identity Based Signature (IBS) scheme proposed by Bellare et al. [38] allows pairing free IBS scheme. Bellare et al. scheme along with Schnorr signature scheme [11] is secure under the intractability of discrete logarithm. But batch verification of this scheme is insecure against forgery attack. To address this weakness, modified version of this scheme in which the private key is generated using ID is proposed by this author. And this ID based signature is generated using the Private key and is sent along with the encrypted message. All CHs will verify the signatures received from sensor nodes, aggregate and generate sign for the aggregated cipher text. BS verifies the authentication of CH and extracts the plaintext by applying elliptic curve based discrete logarithm and pollards lambda algorithm. This is suitable only for plaintext with small size. Batch verification technique is used to filter out injected false data and the authenticity of transmitted encrypted data is verified at the intermediate node as well as at the BS using pairing free identity based signature scheme. But the data integrity is not verified at BS and transmission path.

6.9 Recoverable Concealed Data Aggregation: RCDA [9]

This is a secured data aggregation scheme for cluster based wireless sensor network. It ensures end to end privacy of the data by using EC-EG privacy homomorphism encryption. The raw data of individual sensor nodes can be dig out by BS from the aggregate value and the BS can generate dynamic queries based on the need of the application. The different queries like finding maximum, minimum and average of the sensed data can be applied on the same aggregated data. Also there is no restriction on the type of aggregation applied by the base station and the base station can verify the integrity and authenticity of the sensed data. Hence consistency, accuracy and reliability of the data can be maintained. But RCDA fails to ensure the authenticity and integrity at aggregator node level. Message replay attack and message dropping attack by the compromised cluster header cannot be detected by the BS. Any adversary, who follows the whole communication, can separate the cipher text from the aggregated cipher text and signature. But it cannot aggregate the cipher text of different applications. The authors proposed different schemes for homogeneous (RCDA-HOMO) and heterogeneous network (RCDA-HETERO).

6.9.1 RCDA-HOMO

The BS generates a separate private/public key for all sensor nodes using the key generation algorithm of BGLS [32] signature scheme. And generates its own public/private key pair using the algorithm MGW [33]. And each sensor node is loaded with its own key pairs and public key of BS. Sensor data is encoded for generating signature by using the private key of sensor node which is known only to BS. So the signature cannot be opened and modified by adversaries. And the data is encrypted using its public key of BS, map function and random number which increases the probabilistic characteristics of encoded data. The BS can extract the individual senor node's data using decoding and rmap function and the integrity can be verified collectively. Because of the discrete log used in decryption, size of final aggregated value should be small enough for efficient recovery operation. This constraint puts a limit on the number of sensors in a cluster and CHs. The BS always first decrypts only then the originality of the data is verified. But it cannot verify the validity of encrypted aggregated value. If more number of bogus aggregated values is coming, most of the time is wasted in fruitless decryption process on invalid data.

6.9.2 RCDA-HETERO

In this the nodes with high capacity are designated as cluster heads. All the cluster members share a symmetric key with cluster head. Each sensor nodes encrypt their sensed data with this symmetric key and forward to CH. The CH decrypts data received from all sensor nodes and aggregate them. Now the CH encrypts and generate signature as in RCDA–HOMO. This process repeats till the BS. Since the authenticity of the sensor nodes are not verified by cluster head, the adversary can modify the encrypted data sent by sensor node, which will go unnoticed and leads to false data in the data received by the BS.

6.10 Secure Aggregation Scheme for Wireless Sensor Networks Using Stateful Public Key Cryptography: SASPKE [1]

Authors proposed a method to provide end-to-end confidentiality and integrity with reduced computation and communication cost during aggregation. By supporting a handful of aggregation queries the limitations of existing conceal based schemes are addressed which makes this encryption scheme as Fully Homomorphic encryption.

This method employs Stateful Public Key Encryption (StPKE) proposed by Bellare et al. [26] which avoids the repeated computation of the same value between aggregations. This in turn minimizes the computation overhead. Confidentiality is achieved by encrypting data with symmetric key which is unique for each message. This key is generated using state, nonce and asymmetric key which is shared between sensor node and base station. Since this unique key is valid only for short duration, the replay attack can be identified by Base station by verifying the MAC. Thus it combines the benefits of both symmetric and asymmetric homomorphism. This method requires initial setup of sensor nodes. During network deployment, each sensor should be loaded with a large number M and function for generating keys for encryption using H-MAC. The Base Station (BS) generates private key k and public key p. The elliptic curve domain parameters along with sensor specific secret key shared with BS are loaded in the sensor nodes.

The aggregation occurs in two phases. Before aggregation, the authentication of individual node is checked by BS by collecting state information and MAC from individual sensor node. MAC is generated using the state and dynamic key which is generated using Hash based MAC (H-MAC) function and the secret key shared between sensor node and BS. Thus the state of individual sensor nodes is stored in BS. Since the keys are kept secret between the Base station and sensor node, false data introduced by compromised cluster head in the plain data of cluster member node can easily be identified by BS. But the false modification by the compromised cluster head will go undetected by BS. SASPKE provides efficiency in terms of computation and communication only for stream data. Though the authors claim that computation cost is reduced by maintaining state information in the sensor node, state and dynamic key need to be computed for every aggregation round. Security of symmetric key used for state generation is ensured with the assumption of Diffie–Hellman that is computing f^{xy} is infeasible by knowing the f^{x} and f^{y} where f is generator function.

The integrity of the message is verified at the base station with the help of aggregated MAC. Each sensor node sends a MAC along with the encrypted message. MAC is generated with the help of H-MAC function using state and nonce. Since this method enables BS to retrieve individual sensor node data, dynamic queries are supported and malicious nodes can easily be identified by BS without knowing the list of all participating nodes. The shared state information and nonce based dynamic keys enable this method to address malleability, packet forgery and replay attack but selective forwarding attack problem is not addressed.

6.11 Malleability Resilient Concealed Data Aggregation in Wireless Sensor Networks: MRCDA [29]

The authors proposed a scheme to make privacy homomorphism based data aggregation secure against the malleability attack. Though concealed aggregation scheme provides security against eavesdropping attack by aggregating encrypted data without decrypting them at intermediate node, PH schemes are malleable. The attacker can inject false data without knowing the original data. In this proposed scheme, the EC-EG based PH is employed to protect the confidentiality of the data. And two different MACs are employed to ensure the originality of the data at intermediate aggregator node as well as at the base station. Symmetric key based MAC (SMAC) is used to ensure the freshness of the data at aggregator node and helps to detect the malicious activities by compromised intermediate nodes. For end to end integrity, Homomorphic MAC (HMAC) is generated using the key pair shared between the BS and all leaf nodes. The fresh data is encrypted using the public key of EC-EG system and HMAC is generated from this encrypted data. SMAC is generated using HAMC and counter which randomize the SMAC and prevents chosen cipher text attack. Though the Usage of EC-EG cryptosystem make MRCDA secure against known plain text, cipher test analysis, it considerably increases the message size which leads to communication overhead. But compare to the CMT cryptosystem [25, 28] and the

Westhoff et al.'s cryptosystem [12], MRCDA witnessed notable reduction in bandwidth consumption in the presence of non responding nodes. Though this method ensures the integrity of the leaf node at BS level, the genuineness is not ensured.

6.12 Comparison of APHMS

Compared to SPHM, APHM schemes offer improved security by addressing compromising node attack. The malleability attack which converts the cipher text into another valid cipher text is addressed by SASPKE [1], MRCDA [29], CDAMA [8]. The replay attack which has a noticeable impact on the duplicate sensitive aggregation function is nullified by using time stamp and nonce. The forgery attack on ID based signature is addressed by existing schemes by making appropriate modification to the encryption schemes. The leakage of valuable information by colluding nodes is addressed by ASAHS. In EPPCDA, brute force attack is prevented by thwarting the intruder from analyzing the traffic and flow. Table 5 gives the comparison of APHM based data aggregation schemes. ASAHS [10] verifies the genuineness of individual nodes at both CH and BS and in SASPKE [1] it is only at BS. Verification at cluster head helps to avoid the participation of the unauthorized nodes in the aggregation at the beginning which avoids bandwidth consumption. In MRCDA [29], IPHCDA [3] validity of data is verified at both CH and BS. Verification at CH level helps to prevent the propagation of false data at the earliest. Secured aggregation scheme for supporting multiple application has been proposed in CDAMA [8].

7 Non PH Schemes

7.1 Secure Data Aggregation in Wireless Sensor Networks Filtering Out the Attacker's Impact: SDA-FoAI [17]

The authors proposed a secure algorithm based on synopsis diffusion which enables the BS to sieve the dishonest aggregate data introduced by compromised nodes and receive the correct aggregation even in the presence of falsified data. It can compute duplicate sensitive aggregate function like SUM, COUNT. The synopsis diffusion algorithm addresses the communication loss problem present in the tree network model by using ring topology. The algorithm is implemented in two phases. In phase 1, Initial estimation of aggregate is collected using the authentication information received from nodes. In phase 2, group of nodes are decided using the estimation of phase 1 and more authentication information is collected from these nodes only. All other nodes are not considered.

In this scheme, Ring topology is formed around BS based on the number of hops from BS to nodes. Nodes with same number of hops form rings. The algorithm works using 3 functions for generating, blending and assessing Synopsis. The node generates a synopsis value, a bit vector which represents the data to be aggregated using random toss count and it is broadcast. The receiving node in the inner ring, fuse the received synopsis with its own and broadcast further. Let X be the inner node. The fused synopsis of all received synopsis from its n children F_1, F_2,… F_n.

$$F_x = \text{synopsis of X } ||F_1||F_2||\ldots||F_n$$

where || is bitwise OR operator. Synopsis of X is bit vector of the data of node. MAC is generated for originality verification using the ID

$$MAC_x = \text{fun}(Key_x, L)$$

where Key_x = secret key shared between node and BS, L = {ID_x, M_x, b_1, b_2, … b_n, seed}, ID_x = ID node x, M_x = message of node X, b_1, b_2, … b_n = bit positions which are 1 in the synopsis, seed = random value broadcast by BS while deploying node.MAC is forwarded by the node along with synopsis L. The BS recalculates the MAC using L and detects the false MAC generated by the compromised node. The BS ensures the truth of the final synopsis by receiving at least one valid MAC for each 1 bit. To address the bandwidth consumption of the MAC, during 1st phase, only randomly selected MACs are forwarded by the nodes. During 2nd phase nodes which contributed the 1 bits in the final synopsis forward the MACs. So to ensure the correctness of the final aggregate, the BS does not need to receive authentication message from all nodes. Thus communication overhead is reduced. This scheme assumes even distribution of Compromised nodes. Otherwise the communication overhead is high. Only false data by compromised node (CN) is addressed, not the data leakage problem by CN.

7.2 Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks: SDA-CA [18]

In this the authors proposed an improved iterative filtering technique (IFT). This makes the aggregation robust against collusion attack and more accurate by calculating the initial trust of sensor node. Iterative filtering algorithms address the problem of cryptography based aggregation methods in which the adversary can gain complete access of information by compromising nodes. The compromised nodes can also skew the aggregation by injecting the false data. In

Table 5 Comparison of asymmetric PH based data aggregation schemes

Scheme	Encryption/ method	Network model	Attacks addressed	Authentication	Integrity	Scalability	Multiple application	Data recover
ASAHS [10]	Asymmetric additive PH using OK–UC	Tree topology	Minimised the impact of collision attack. But one intermediate node to be the part of colluding nodes to reduce the impact	Authentication of All Participating Nodes	Yes	Yes	No	No
PPIA [15]	Asymmetric PH with threshold	Tree topology	False data injection attack	No	No	No	No	No
EPPCDA [16]	Asymmetric PH with ID based compressive sensing technique	Tree topology	Brute force attack, node compromising attack, traffic analysis attack	No	No	No	No	No
SASPKE [1]	Hybrid PH (using both symmetric and asymmetric key) fully homomorphic encryption	Clustered tree model	Replay attack, Malleability, packet forgery	Authentication of individual node is verified at BS	Yes	No	No	No
CDAMA [8]	Asymmetric additive PH with EC	Clustered based model	Aggregator node compromising attack (repeated aggregation or selective aggregation), cipher text only attack, known plain text attack, chosen plain text attack, unauthorised aggregation, malleability	No	No	No	Yes	No
SDAMQ [2]	Additive PH with EC	Clustered model	Replay attack	Authentication of query's origination is verified at each node	No	No	No	No
IPHCDA [3]	El additive PH with EC	Clustered tree model	Replay attack (using time stamp and nonce), chosen plain text attack,	Supported at CH and BS	Yes	No	No	No
SDAACP [6]	Additive PH with EC	Clustered model	Adaptive chosen message attack, eavesdropping, forgery attack	Supported	No	No	No	No
RCDA [9]	EC-EG based PH, fully homomorphic	Clustered model	Eavesdropping attack, forgery attack	Supported at BS	Yes	No	No	Yes
MRCDA [29]	Hybrid PH	Support both tree and cluster topology	Eavesdropping attack, known plain text attack, chosen cipher text attack, malleability, replay attack	Supported at BS and intermediate node	Yes	No	No	No
FESA [5]	Asymmetric PH with fully homomorphic	Multi hop network model	Reduced impact of node compromising attack, plain text only attack	No	Yes	No	No	No

IFT, by assigning a trust to each sensor node, the impact of false data on the final aggregation is reduced. The trust of each node is calculated by comparing the current reading with the weighted average of the previous round. If the difference is high, the sensor is assigned less weight which reduces their contribution or impact in the aggregation process of current round. But in the traditional method each sensor are initially given an equal weight which gives room for byzantine attack scenario which is highlighted in this paper and is overcome by calculating the initial trust of sensor node using a novel approach proposed by this authors. In this novel approach, initial reputation is

calculated with reduced number of iteration and initial weight of each sensor node is calculated based on the distance between current reading and the initial reputation. Thus this improved IFT addresses the byzantine attack model. It also allows the aggregation to converge quickly with comparatively high accuracy. But Problem of compromised aggregator nodes is not addressed. Only the impact of compromised sensor nodes is reduced by measuring the trust.

7.3 A Secure and Efficient ID Based Aggregate Signature for Wireless Sensor Network: IDBAS-WSN [19]

The authors proposed a secured data aggregation scheme which ensures the integrity of individual sensor node, reduces the bandwidth and storage cost. It employs ID based ciphering and signature aggregation scheme. Use of ID based cryptography eliminates the need of trusted third party for certificate generation which in turn reduces the communication, computation overhead and usage of memory. Using this, the authors addressed the coalition attack in which the adversary can inject invalid signature and get through the verification test by successfully generating valid aggregation signature. To achieve this the network is modelled as clusters and one node with high computing capability is designated as aggregator which aggregate the data and signature received from each sensor node and forwards to cluster centre (CC) which verifies the integrity. The CC is loaded with pair of security key namely PB_{cc}, PR_{cc}. PR_{cc} is generated using ID and PB_{cc} is known to all members. This scheme also requires each cluster member to be loaded with the security parameters and ID based Secret key before deployment. This scheme uses 6 different probability based algorithm which function in polynomial time, for setting up the network, generating ID based keys, signature generation for integrity verification, aggregation and final verification. Each sensor node generates a randomized signature using its ID, secret key and Hash function which is used by CC for integrity verification. Usage of polynomial time based algorithm remains robust against the coalition attack by ensuring that the aggregate signature verification will succeed only if all the individual sign are valid. And the security of this scheme is justified by using computational Diffie hellman assumption which says that the problem of finding xyM ϵ G given M, xM, yM ϵ G where G is a cyclic group and x, y are randomly chosen from Z_p* and M is a generator of G, is hard. This scheme addresses only the integrity and coalition attack.

7.4 Genetically Derived Secure Cluster-Based Data Aggregation in Wireless Sensor Networks: GDSDA [20]

The author proposed a secured data aggregation scheme using genetic algorithm. The efficiency of clustered wireless sensor network is improved by clustering the network using genetic algorithm. Initially the cluster head is selected by analysing the connectivity and clustering is optimized by executing genetic algorithm. After electing the initial CH, the sensor nodes are assigned to cluster head by considering the node distribution, cost estimated based on the distance, communication cost which include transmission cost, receiving cost, sensing cost and idle state cost. The fitness value of individual cluster member is computed by estimating the total distance between the node and sink and distance measured via new CH. The chromosomes with higher fitness factors are elected as offspring generator using fitness proportionate selection genetic operator. Then cross over techniques is applied to verify the eligibility of new CH and to enable the nodes to identify their new CH. This process highly reduces the energy consumption which in turn increases the lifetime of energy starving sensor network. The false data packets are identified and filtered by encrypting using the keys which are generated dynamically using the residual energy of the node. During the initial transmission, each sensor node generates the key using initial residual energy and initialization vector which is pre distributed to all nodes. In the next rounds of data transmission, the new key is generated using the residual energy and previous round key. Security is enhanced by generating permutation code using ciphered data and ID by applying RC5 algorithm. This permutation code along with ID is sent to aggregator, which extracts the key by using the previous round key and calculated residual energy. This encryption scheme requires some space for storing the generated keys. Hence by compromising the nodes the intruder can easily gain knowledge of keys and induce mal activities.

7.5 Scalable Privacy-Preserving Big Data Aggregation Mechanism: Sca-PPBDA [21]

In this, the authors proposed a secured data aggregation scheme with scalability support by organizing the nodes into hierarchy of clusters such that size of each cluster remains equal i.e. number of cluster members are equal to the number of cluster heads in each level. This enables the usage of same secured aggregation scheme in all clusters independent of the network size and hence supports the big data aggregation in wireless sensor network. To ensure confidentiality, sensed true data is embedded among the

camouflage values whose locations in the indexed data packet is decided by the configuration message sent by the sink node to each cluster. True value is hidden from the adversary by dividing the camouflage values into restricted and unrestricted. This also prevents the true sensed value from getting mixed with camouflage value. Privacy of this method highly depends on the secure transformation of configuration message. If an adversary is able to gain the configuration message, he can easily extract the true sensed value.

7.6 Secure and Energy-Efficient Data Aggregation with Malicious Aggregator Identification in Wireless Sensor Networks: SEMAI [22]

The authors proposed a scheme which enables to identify false aggregator nodes by making the child nodes verify the consistency of their parent node's aggregate value with constant communication overhead. It also verifies the genuineness of aggregator node by generating signature using the secret key of individual aggregator. Communication between nodes is secured by pair wise shared secret keys. The compromised intermediate aggregate node is identified by making its children to recalculate the aggregation by using the data of all its siblings and is compared with aggregated value of its parent node which is received from the grand parent. If any discrepancy is found, alarm is broadcast in the network and prevents BS from accepting false aggregate. By receiving the aggregate from grandparent instead of parents, the compromised aggregator node is prevented from playing the trick by sending original unmodified aggregate. If both parent and grandparent nodes are compromised, the attacker can escape from verification process and create colluding attack. This is overcome by making the child to report to BS. The involvement of all intermediate nodes in the verification and aggregation operation increases the computation overhead. This is against the purpose of aggregation which is to avoid redundancy. The intermediate node after receiving each packet decrypts them using the public key which can be calculated using ID of the source node. Authentication of each data packet is verified by attaching signature which is generated by applying the hash function over the ID, count and original data. Hence by knowing the ID and by compromising the node, the public key can be easily generated by adversaries. So it requires ID should be kept secret.

7.7 Hilbert-Curve Based Data Aggregation Scheme to Enforce Data Privacy And Data Integrity For Wireless Sensor Networks: iHDA [30]

In this, a novel scheme is proposed to ensure the privacy and integrity using Hilbert curve based aggregation. Privacy is ensured by aggregating the data seeds received from multiple neighbours and transformed as coordinates along Hilbert curve. Usage of seed reduces the communication cost comparatively by not using extra message for privacy. The authors proposed integrity verification algorithm using private information retrieval (PIR) method. First the tree topology is established by broadcasting HELLO packet and all the nodes identify their parents and siblings. A balance is maintained by limiting the maximum number of child of a node. Using EC key exchange algorithm, the seeds are exchanged among the nodes. The sensed data is protected from adversaries by hiding in its own seed and the seeds received from its neighbour nodes. This hidden data is forwarded to parent node. Each node determines the direction and level of the Hilbert Curve (HC) and the one dimensional data is encrypted by mapping to the 2 dimensional points in the HC. The encrypted data is of the form

$$<\text{key}(\text{Dr}, \text{L}), \text{m}, \text{n})$$

where Dr is the direction of the HC, L is the level of the HC, m, n are coordinates in HC.

The parent node decrypts the received encrypted data by matching the direction and level of the HC. And it aggregates and encrypts the new aggregate using its own HC and forwards. The aggregator ensures the integrity by sending PIR message to the child node. The child responds by generating HC which is compared for integrity verification. In the PIR message, which is split into cells, the desired cell which carries the data is selected using the prime number and modified n value, which makes it difficult for adversaries to find the exact desired cell. Thus this scheme ensures privacy and integrity with comparatively reduced communication cost by eliminating extra messages

7.8 Comparison of Non PH Based Schemes

The computationally efficient non encryption based data aggregation schemes have addressed coalition, byzantine, collusion attack and node compromising attack. In existing methods, collective authentication or individual node level authentication is verified either at BS or at both Cluster head and BS. In SDA-FoAI [17] and IDBAS-WSN [19] collective verification is done at BS and in GDSDA [20] endorsement is verified at both cluster head and BS. In SDA-FoAI [17], IDBAS-WSN, [19], iHDA [30], the

Table 6 Comparison of non PHM based data aggregation schemes

Scheme	Encryption/ method	Network model	Attacks addressed	Authentication	Integrity	Scalability	Multiple application	Data Recovery
SDA-FoAI [17]	Synopsis diffusion	Ring topology	Compromised node-falsified sub aggregate attack by compromising node	Collective authentication	False aggregation is sieved at BS	No	No	No
SDA-CA [18]	IFT (Iterative filtering technique)	Clustered model	Compromised node, byzantine attack, collusion attack using compromised node	No	No	No	No	No
IDBAS-WSN [19]	IBAS with encryption	Clustered model	Coalition attack	Collective authentication	Supported., verified at BS	No	No	No
GDSDA [20]	Dynamic key based encryption	Clustered model	Eavesdropping attack	Supported. verified at CH, BS	Supported, verified at CH, BS	Scalability affects the efficiency	No	No
Sca-PPBDA [21]	Camouflaging based non encryption scheme	Clustered model	Eavesdropping attack	No	No	Scalable	No	No
SEMAI [22]	ID based key for encryption	Tree topology	Compromised aggregator node attack, colluding attack	Authentication of aggregator node	No	Scalability increases the communication and computation overhead	No	No
iHDA [30]	Hilburt curve based encryption scheme	Tree topology	Eavesdropping attack	No	Supported, verified at BS and aggregator node	Communication cost increases with number of nodes	No	No

validity of the aggregated data is ensured by carrying out the verification process at BS where as in GDSDA [20] validity of data is verified at both CH and BS. Verification at CH level helps to prevent the propagation of false data at the earliest. Sca-PPBDA [21] supports scalability by grouping the nodes into clusters of equal size without affecting the performance of the network where as in SDA-FoAI, IDBAS-WSN, GDSDA, SDA-CA, SEMAI and iHDA increase in the number of nodes affects the performance on the network. Table 6 gives the comparison of non PH based data aggregation schemes.

8 Conclusion

The comprehensive survey on existing privacy preserved data aggregation techniques for wireless sensor network explores the various mechanisms for data aggregation for preserving energy of sensor nodes by eliminating the redundant data transmission. In this survey the existing schemes are classified based on the networking model and the mechanisms employed for ensuring the privacy. This survey reveals that the topology of nodes has significant impact on the performance of the data aggregation schemes. Since the sensor nodes are employed mostly on the hostile, human inaccessible environment they are exposed to major security threats. This survey highlights the various security attacks such as eavesdropping, node compromising, coalition, collusion, byzantine attack and the solutions proposed to overcome the same. The existing schemes are compared using major security factors like confidentiality, integrity which ensures the freshness of the data and authentication which guarantees the genuineness of the nodes. Since sensor nodes of different applications which sense different environment factor are deployed in the same area, there is a need to extract application specific data from the aggregation. And the ability to recover data from the individual sensor nodes supports multiple

dynamic queries. Comparison of the schemes based on the privacy factors such as confidentiality, integrity, and authentication unfolds how far the schemes support factors like scalability, multiplication, and data recovery. To meet the growing need of wireless sensor network with large number of nodes, finding proactive mechanism which ensures the privacy, scalability and support of multiple applications with dynamic query becomes an open research issue.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Boudia ORM, Senouci SM, Fehama M (2015) A novel secure aggregation scheme for wireless sensor networks using stateful public key cryptography. Ad Hoc Netw 32:98–113. https://doi.org/10.1016/j.adhoc.2015.01.002
2. Prathima EG, Shiv Prakash T, Venugopal KR, Iyengarc SS, Patnaik LM (2016) SDAMQ: secure data aggregation for multiple queries in wireless sensor networks. Procedia Comput Sci 89:283–292
3. Ozdemir S, Xiao Y (2011) Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. Comput Netw 55:1735–1746. https://doi.org/10.1016/j.comnet.2011.01.006
4. Boneh D, God E-J, Nissim K (2005) Evaluating 2-DNF formulas on cipertexts. In: Proceedings of theory of cryptography conference, LNCS, vol 3374, pp 325–321
5. Li X, Chen D, Li C, Wang L (2015) Secure data aggregation with fully homomorphic encryption in large-scale wireless sensor networks. Sensors 15:15952–15973. https://doi.org/10.3390/s150715952
6. Shim K-A, Park C-M (2015) A secure data aggregation scheme based on appropriate cryptographic primitives in heterogeneous wireless sensor networks. IEEE Trans Parall Distrib Syst 26(8):2128–2139. https://doi.org/10.1109/tpds.2014.2346764
7. Zhao X, Zhu J, Liang X, Jiang S, Chen Q (2017) Lightweight and integrity-protecting oriented data aggregation scheme for wireless sensor networks. IET Inf Secur 11(2):82–88. https://doi.org/10.1049/iet-ifs.2015
8. Lin Y-H, Chang S-Y, Sun H-M (2013) CDAMA: concealed data aggregation scheme for multiple applications in wireless sensor networks. IEEE Trans Knowl Data Eng 25(7):1471–1483. https://doi.org/10.1109/TKDE.2012.94
9. Chen C-M, Lin YH, Sun H-M (2012) RCDA-recoverable concealed data aggregation for data integrity in wireless sensor networks. IEEE Trans Parall Distrib Syst 23(4):727–734. https://doi.org/10.1109/TPDS.2011.219
10. Viejo A, Qianhong W, Domingo-Ferrer J (2012) Asymmetric homomorphisms for secure aggregation in heterogeneous scenarios. Inf Fusion 13:285–295. https://doi.org/10.1016/j.inffus.2011.03.002
11. Schnorr CP (1990) Efficient identification and signatures for smart cards. In: Proceedings of the 9th annual international cryptol conference advances in cryptology, pp 239–252
12. Westhoff D, Girao J, Acharya M (2006) Concealed data aggregation for reverse multicast traffic in sensor networks: encryption, key distribution, and routing adaptation. IEEE Trans Mobile Comput 5(10):1417–1431. https://doi.org/10.1109/tmc.2006.144
13. Paillier P (2000) Trapdooring discrete logarithms on elliptic curves over rings. Okamoto T (ed) ASIACRYPT 2000, LNCS 1976. Springer, Berlin, pp 573–584
14. Zhang K, Han Q, Cai Z, Yin G (2017) A ring-based privacy-preserving aggregation scheme in wireless sensor networks. Sensors 17:300. https://doi.org/10.3390/s17020300
15. Singh VK, Verma S, Kumar M (2016) Privacy preserving in-network aggregation in wireless sensor networks. Procedia Comput Sci 94:216–223. https://doi.org/10.1016/j.procs.2016.08.034
16. Xie K, Ning X, Wang X, He S, Ning Z, Liu X, Wen J, Qin Z (2017) An efficient privacy-preserving compressive data gathering scheme in WSNs. Inf Sci 390:82–94
17. Roy S, Conti M, Setia S, Jajodia S (2014) Secure data aggregation in wireless sensornetworks: filtering out the attacker's impact. IEEE Trans Inf Forensics Secur 9(4):681–694. https://doi.org/10.1109/tifs.2014.2307197
18. Rezvani M, Ignjatovic A, Bertino E, Jha S (2015) Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. IEEE Trans Dependable Secure Comput 12(1):98–110. https://doi.org/10.1109/tdsc.2014.2316816
19. Shen L, Ma J, Liu X, Wei F, Miao M (2017) A secure and efficient id based aggregate signature for wireless sensor network. IEEE Internet Things J 4(2):546–554. https://doi.org/10.1109/jiot.2016.2557487
20. Bhasker L (2014) Genetically derived secure cluster-based data aggregation in wireless sensor networks. IET Inf Secur 8(1):1–7. https://doi.org/10.1049/iet-ifs.2013.0133
21. Wu D, Yang B, Wang R (2016) Scalable privacy-preserving big data aggregation mechanism. Digit Commun Netw 2:122–129. https://doi.org/10.1016/j.dcan.2016.07.001
22. Li H, Li K, Wenyu Q, Stojmenovic I (2014) Secure and energy-efficient data aggregation with malicious aggregator identification in wireless sensor networks. Future Gener Comput Syst 37:108–116. https://doi.org/10.1016/j.future.2013.12.021
23. Domingo-Ferrer J (2002) A provably secure additive and multiplicative privacy homomorphism. In: ISC 2002, LNCS 2433, pp 471–483. Springer, Berlin
24. Dijk MV, Gentry C, Halevi S, Vaikuntanathan V (2010) Fully homomorphic encryption over the integers. In: Proceedings of the 29th annual international conference on the theory and applications of cryptographic techniques (EUROCRYPT'10), Riviera, France, 30 May–3 June 2010, pp 24–43. https://doi.org/10.1007/978-3-642-13190-5_2
25. Castelluccia C, Mykletun E, Tsudik G (2005) Efficient aggregation of encrypted data in wireless sensor networks. In: The second annual international conference on mobile and ubiquitous systems: networking and services (MobiQuitous 2005), July 2005. https://doi.org/10.1109/MOBIQUITOUS.2005.25
26. Bellare M, Kohno T, Shoup V (2006) Stateful public-key cryptosystems: how to encrypt with one 160-bit exponentiation. In: Proceeding of the 13th ACM conference on computer and communications security, October 2006, ACM, Alexandria, VA, 2006, pp 380–389. https://doi.org/10.1145/1180405.1180452
27. Cam SOH (2010) Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks. IEEE/ACM Trans Netw 18:736–749. https://doi.org/10.1109/TNET.2009.2032910

28. Castelluccia C, Chan ACF, Mykletun E, Tsudik G (2009) Efficient and provably secure aggregation of encrypted data in wireless sensor networks. ACM Trans Sens Netw (TOSN) 5(3):20:1–20:36. https://doi.org/10.1145/1525856.1525858
29. Parmar K, Jinwala DC (2016) Malleability resilient concealed data aggregation in wireless sensor networks. Wirel Pers Commun 87:971. https://doi.org/10.1007/s11277-015-2633-6
30. Kim Y-K, Lee H, Yoon M, Chang J-W (2013) Hilbert-curve based data aggregation scheme to enforce data privacy and data integrity for wireless sensor networks. Int J Distrib Sens Netw 9(6):217876. https://doi.org/10.1155/2013/217876
31. Bellare M, Canetti R, Krawczyk H (1996) Keying hash functions for message authentication. Crypto. https://doi.org/10.1007/3-540-68697-5_1
32. D Boneh, C Gentry, B Lynn, H Shacham (2003) Aggregate and verifiably encrypted signatures from bilinear maps. In: Proceedings of 22nd international conference on the theory and applications of cryptographic techniques (Eurocrypt), pp 416-432. https://doi.org/10.1007/3-540-39200-9_26
33. Mykletun E, Girao J, Westhoff D (2006) Public key based cryptoschemes for data concealment in wireless sensor networks. IEEE Int Conf Commun 5:2288–2295. https://doi.org/10.1109/ICC.2006.255111
34. Koblitz N (1987) Elliptic curve cryptosystems. Math Comput 48(177):203–209. https://doi.org/10.1090/S0025-5718-1987-0866109-5
35. Hoffstein J, Pipher J, Silverman JH (2008) An introduction to mathematical cryptography. Springer, Berlin. https://doi.org/10.1007/978-1-4939-1711-2
36. Bista R, Kim Y-K, Song M-S, Chang J-W (2012) Improving data confidentiality and integrity for data aggregation in wireless sensor networks. IEICE Trans Inf Syst E95-D(1):67–77. https://doi.org/10.1587/transinf.E95.D.67
37. Boneh D, Goh E, Nissim K (2005) Evaluating 2-DNF formulas on ciphertexts. Proc Second Int'l Conf Theory Cryptogr (TCC) 3378:325–341. https://doi.org/10.1007/978-3-540-30576-7_18
38. Bellare M, Namprempre C, Neven G (2004) Security proofs for identity-based identification and signature schemes. Proc Adv Cryptol Int Conf Theory Appl Cryptogr Techn. https://doi.org/10.1007/s00145-008-9028-8