Asian Journal of Information Technology 15 (10): 1552-1558, 2016
ISSN: 1682-3915

Trust Based Secure Data Aggregation for Privacy Protection and Integrity in WSN

J. Jean Justus and A. Chandrasekar
Department of Computer Science and Engineering, St. Joseph's College of Engineering,
600119 Chennai, Tamil Nadu, India

Abstract: Data aggregation method in Wireless Sensor Network (WSN) is usually vulnerable to various faults and security attacks which often results in alteration of data during the transmission. In order to overcome these issues, in this study, Trust based Secured Data Aggregation for Privacy Protection and Integrity is proposed for WSN. The aggregator analyses the data from each nodes and determines the different possibility of occurrence of the event to determine whether the action of node is correct or erroneous. Each node generates its local synopsis and sends it to the aggregator node. The aggregator node checks the trust worthiness of the received data. If valid, then the data is encrypted and data from all sensors are aggregated and forwarded to the base station. The generated report is then sent to the base station by the aggregator for verification of the malicious or liar node. The main advantage of our proposed technique is that it provides a robust method to increase security during data aggregation.

Key words: Data aggregation, synopsis, cluster, trust, malicious node

INTRODUCTION

Wireless Sensor Network (WSN): Wireless Sensor Network (WSN) is made up of spatially distributed sensor nodes. These nodes can cooperatively achieve one or more global functionalities. Sensor network can be mainly used to reply to the queries regarding the data gathered by the sensor nodes. Considerable amount of data has been generated by the large sensor networks. However, the sensor nodes can either be resources limited or energy constrained. Therefore, an effective data processing technique should be developed in order to efficiently utilize the data. The WSNs are also used in participatory sensing applications in which integrity and privacy of data are the major issues (He *et al*., 2008). Some other applications of WSN are monitoring of the physical parameters such as temperature, humidity and seismic activity, monitoring of ecological environment, law enforcement and military fields.

Without considering the applications, WSNs have two significant properties: WSN reaches a collective conclusion related to the outside environment that needs sensor level detection and coordination and WSNs operates under severe technical constraints such as limited computation, communication and power (battery) resources while operating in great spatial and temporal variability environment (Castelluccia *et al*., 2009).

Secure data aggregation: In WSN, data aggregation is a specific power-saving and efficient mechanism for query processing. Data are aggregated and processed within the network. These data are then returned to the base station. The nodes that are used for aggregating the information requested by the query are called aggregators. Aggregators gather the raw information from the sensors. After gathering, it processes the collected data locally. Finally, it sends reply to the aggregate queries of a remote user (He *et al*., 2008). The data aggregation techniques can be classified as centralized and in-network aggregation (Dhasian and Balasubramanian, 2013).

Data aggregation can be attacked by the malicious nodes that can inject wrong information or falsify aggregation values to perform a man-in-the middle attack. Adversaries can deploy sensors near existing sensors to conduct a known-plaintext attack or can tamper the sensors to force them to predetermined values thus conduct a chosen-plaintext attack (Huang *et al*., 2010).

To detect and avoid these attacks, several methods have been proposed that is based on the complex data authentication or the statistical features of specific aggregation. Persistent authentication is used for ensuring correctness. Providing persistent authentication is very safe in sensor networks.

Problem identification and solution: In our previous research, a fault tolerance data aggregation system has been proposed in WSN in which the nodes with maximum

Corresponding Author: J. Jean Justus, Department of Computer Science and Engineering, St. Joseph's College of Engineering, Old Mamallapuram Road, 600119 Chennai, Tamil Nadu, India

link quality, residual energy and coverage are selected as an aggregator node (Justus and Sekar, 2013). When the aggregator node transmits the data towards the sink, it constructs the routing tree with good quality links. A cross validation of the data recorded in the nodes is performed for preventing the data level faults.

As an extension work, we propose to design a fault-tolerant trust based secure data aggregation technique.

Lirterature review: John *et al.* (2013) have developed a Recoverable Concealed Data Aggregation (RCDA) technique. In RCDA, a base station can recover each sensing data generated by all sensors even if these data have been aggregated by aggregators. The base station can find out the integrity and authenticity of all sensing data and can execute any aggregation functions on them. But this technique imposes huge overhead on the base station.

Bhaskar (2014) proposed a genetically derived secure cluster-based data aggregation for WSN. The clustering process is done using genetic algorithm and the cluster heads or aggregators are selected based on the node connectivity. When a cluster member wants to transmit the data to the aggregator, data encryption technique is utilized that offers authenticity, confidentiality and integrity. However the trust worthiness of cluster members and aggregators are not ensured which leads to insider attacks.

Rezvani *et al.* (2015) have proposed an improvement for iterative filtering (IF) techniques. The main contributions of the work can be summarized as follows: identified a new sophisticated collusion attack against IF based reputation systems. Derived a new method for the sensor's faults which are not detected by known attacks. Designed an efficient and robust aggregation method inspired by the Maximum Likely hood Estimator (MLE). Provided an initial estimate of trustworthiness of sensors. Since sensors are compromised only relative to a particular batch, the framework is applied over consecutive batches of consecutive readings. However collecting the readings from the sensors and processing them involves huge overhead and time.

Mansouri *et al.* (2013) have proposed a new method which consists of three phases. In phase-1, best set of candidate sensors that participate in data aggregation are selected based on the transmission power between the cluster member and the cluster head. In phase-2, the malicious sensors are detected based on the information relevance of their measurements. In phase-3, the target position is estimated using Quantized Variational Filtering (QVF) algorithm. But the malicious sensors are detected only based on the relevance of data without considering the dishonest activities of sensor nodes.

Liu *et al.* (2013) have proposed an improved Reliable, Trust-based and Energy-efficient Data-Aggregation (iRTEDA) protocol for wireless sensor networks. It combines the reputation system, residual energy, link availability and a recovery mechanism to ensure that the network is secure, reliable and energy-efficient. Though it posses reputation mechanism for detecting insider attacks, it lack authentication and confidentiality techniques to avoid outsider attacks.

Stelte and Matheus (2011) have proposed a step towards a trustworthy sensor in-network data aggregation without the requirement of hardware tamper-protection or high cost modifications on sensor node equipment. Data aggregation can also be used to minimize the energy consumption. This makes the attackers to compromise the network often by changing the data or disrupting the transmissions. Secure data aggregation protocols detect the manipulation of the aggregation results on the basis of a criterion. In this approach, there is a possibility for multi-criteria decision-making on the basis of weighted Choquet Integral. The trust level calculation algorithm depends on a Gaussian probability function along with a Byzantine decision-making. However, this scheme does not completely reduce the attack of misbehaving nodes.

Trust based secure data aggregation

Overview: In this study, we propose to design a fault-tolerant trust based secure data aggregation technique for WSN. In this technique, each node sends reputation report to the aggregators. The aggregator node is selected based on factors like its validity, trustworthiness, higher residual power, etc.

Each sensor generates a local synopsis using the synopsis diffusion mechanism (Roy *et al.*, 2014). The synopsis generation function is based on the Sum () algorithm. Then using Stateful Public Key Encryption (StPKE) (Boudia *et al.*, 2015) each sensor encrypts the synopsis using homomorphic encryption and calculate the corresponding HMAC. Each trusted aggregator X generates its local synopsis QX and then using StPKE creates MAC(QX). It also receives the MAC values of synopsis from its m children X1, X2 ,..., Xm as MAC(BX1), MAC(BX2), ..., MAC(BXm)

The trusted aggregator X first checks the global trust values of each sensors and then applies the homomorphic aggregation to create the fused synopsis. If the trust value of any sensor is low, it will be omitted for aggregation. It then, forwards the aggregated synopsis to the base station. After receiving all sub-aggregates from each aggregator X, the BS invokes the decryption and verification processes. Figure 1 shows the block diagram of the proposed secure data aggregation technique.

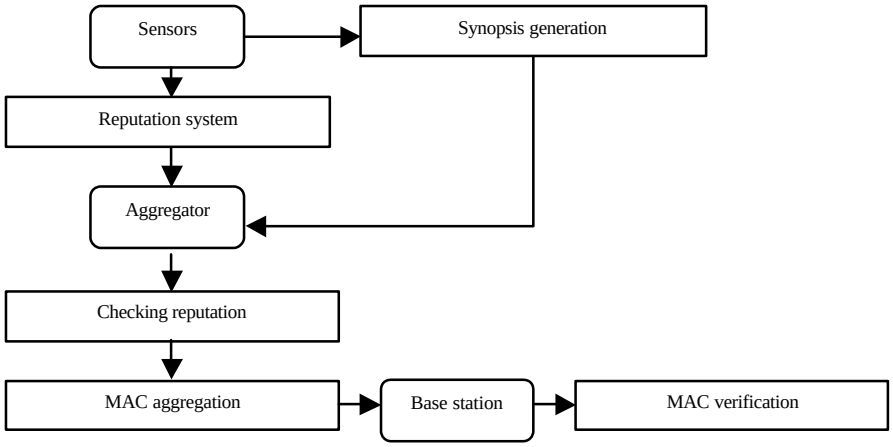

Fig. 1: Block diagram of secure data aggregation technique

MATERIALS AND METHODS

System model

Attacker model: In WSN, the nodes in the network are grouped to form a cluster. The nodes in the cluster select a node as Cluster Head (CH). Aggregator node is chosen among the nodes for performing certain security operations. Based on the ability of adversaries, the attack model can be defined. We consider the attacks and faults that can alter the data transferred. Aggregator nodes are used to identify the malicious nodes and informs about it to the base station.

Reputation system: The reputation model in our proposed work is based on beta distribution system. It provides a very stable and secure framework in order to increase the trust value of the nodes participating in clustering.

The beta probability distribution is mainly used to define the analytical probability of binary event based on past-observed outcomes of the event. The probability density function can be derived using gamma function (Γ) which is given as below (Ozdemir, 2008).

$$f(r \mid \alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1} \quad (1)$$

Where:

$(\alpha+\beta)$ = The past outcomes
r = The probability of occurrence of outcome α
(1-r) = The probability for occurrence of outcome β

Here, we consider node's action as binary events with possible outcomes as whether the action is correct or erroneous. After that, we will calculate the probability that the behavior of the node for next occurrence is correct or not.

Node i assumes that node j behaves correctly with probability θ. The outcomes are independently drawn on the basis of observations. The value of θ changes for every node j. Since, the parameter θ is undefined, node i model this uncertainty by assuming that θ is drawn from beta distribution which is updated when any new observations are made.

Here, Beta (α, β) distribution is used to compute θ where α and β represents the correct and erroneous action observed, respectively.

During the beginning of the system, initial estimate of θ corresponds to uniform distribution on [0, 1] or equivalent Beta (1, 1) without any former knowledge. Beta probability density function asymptotically approximates a Dirac at θ, when more observations are made. We define reputation rating $R_{i,j}$ as node i have about node j as the expected value of the beta distribution which is parameterized as below (Ozdemir, 2008):

$$R_{i,j}=E(\text{Beta}(\alpha+1,\beta+1))=\frac{\alpha+1}{\alpha+\beta+2} \quad (2)$$

The above-mentioned formula defines that reputation ring increases if more correct actions are observed and decreases if the actions are erroneous. Also, initially, it gives reputation of 0.5 which indicates that any action is equally likely from the node in the absence of any knowledge.

Representation of reputation and trust ratings: In proposed technique, each node maintains its own reputation ratings for all other remaining nodes with which it interacts. Nodes randomly operate for different classes of tasks. For example, a node act correctly while reporting data to its Cluster Head (CH) but shift the outcome of aggregation while performing CH duties.

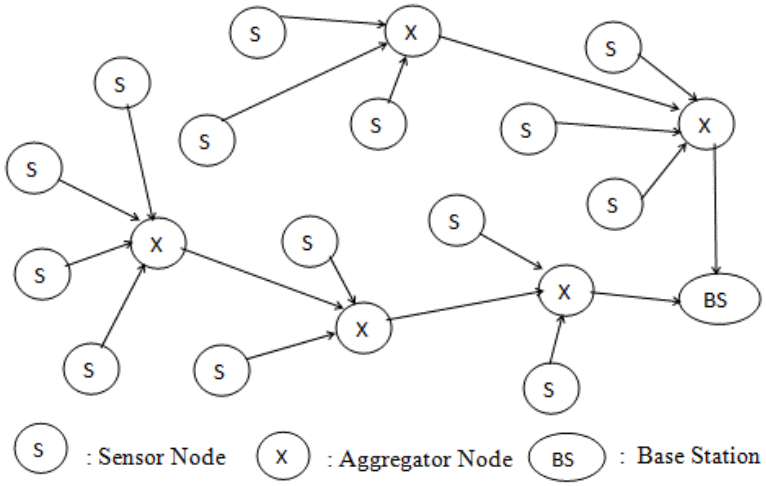

Fig. 2: Data aggregation architecture

Node ID	Sequence No.	Residual energy (E_{res})	Link Quality (LQ)	Node coverage (NC)

Fig. 3: Format of HELLO Mmessage

Hence, reputation is represented by a vector $R_{i,j}<q_1, ..., q_n>$ with a dimension and for each of the different classes of tasks node i keeps observing node j performance.

Reputation of a node is basically built from two observations: either from direct observations of the node or from second-hand reports of the node's behavior. Here, the second hand information is necessary in order to confirm first-hand information and to quickly update reputation as the nodes have only short live interaction with some node. But, this may make the system exposed to different type of attack such as bad mouthing and ballot-stuffing attacks.

In order to overcome this, a separate record of node's accuracy in reporting on the behavior of other nodes is represented in a metric called trust rating. The trust $TR_{i,j}$ which node i have node j is described using the beta distribution (Ozdemir, 2008):

$$TR_{i,j} = E(Beta(\gamma+1, \delta+1)) = \frac{\gamma+1}{\gamma+\delta+2} \quad (3)$$

Here parameter γ, δ represents correct and erroneous reporting action respectively. The Trust Rating (TR) helps to decrease the impact of liar nodes in the reputation system by discounting second hand reports according to node's trust value.

Architecture of data aggregation: The data aggregation method considers three parameters to provide fault tolerant technique: residual energy, Link quality and node coverage. These parameters are then sent by each node in the form of HELLO messages. These parameters are essential for selecting the trustworthy nodes in the cluster. We assume that the nodes can sent the link quality of the current link. As the residual energy is also considered, nodes with maximum energy are selected as aggregator node that can conserve the energy, thereby improving network lifetime. Figure 2 shows the data aggregation architecture. The format for HELLO message is given in Fig. 3.

Secure data aggregation and verification: In Fig. 4, nodes 0 and 1 generate their synopsis value and transmit the corresponding encrypted MAC values MAC_0 and MAC_1, respectively, to their aggregator node AGG1. The AGG1 in turn will generate its local synopsis and MAC value MAC_3. It then transmits the aggregated MAC value (MAC3 +MAC_0+MAC_1) to the base station BS, after checking the trust values of nodes 0 and 1.

Similarly, nodes 3 and 4 generate their synopsis value and transmit the corresponding encrypted MAC values MAC_3 and MAC_4, respectively, to their aggregator node AGG2. AGG3 then transmits the aggregated MAC value (MAC_5+MAC_3+MAC_4) to BS, after checking the trust values of nodes 3 and 4. BS thus receives the aggregated MACs from both the aggregators AGG1 and AGG2. Figure 5 shows the flow diagram of the aggregation process at AGG1.

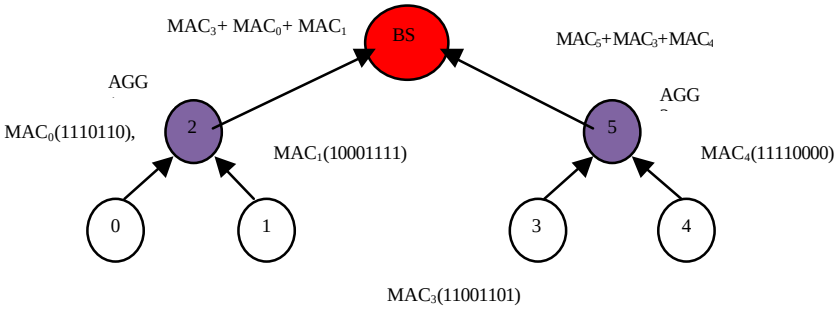

Fig. 4: Secure MAC aggregation

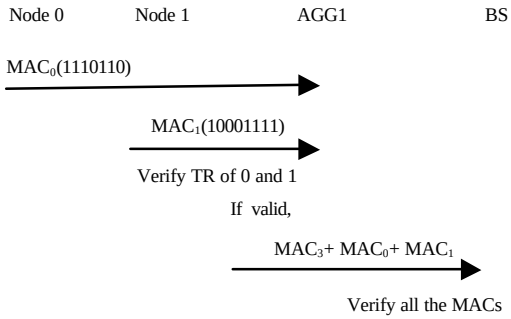

Fig. 5: Flow ddiagram of MAC aggregation at AGG1

Table 1: Aggrigation of notation

Notations	Description
S	Sensor node
X	Aggregator node
P, i, index, n	integer number
Q^S	Local synopsis value of the sensor node
Q^X	Local synopsis value of the aggregator node
key(i)	Key value generated for ith invocation
E^S	Encoded local synopsis data of sensor node
E^X	Ncoded local synopsis data of aggregator node
K^1, K^2	Key value
C	Cipher text
MAC	Message authentication code
MAC_{agg}	Aggregated MAC value
C_{agg}	Aggregated cipher text value
T^{thresh}	Threshold trust value
T^S	Trust Value of the sensor node
BS	Base station

Secure MAC aggregation technique: The trustworthy aggregator is used to perform the aggregation function according to the StPKE technique. The aggregated data is then forwarded to the base station by the aggregator. This process is described in the following Algorithm A (Table 1).

Algorithm A

Each sensor node, S generates a local synopsis using the local synopsis diffusion mechanism based on the Sum( ) algorithm.

In this mechanism, each sensor node S invokes a function called as the CoinToss( ) p times.

In the ith invocation where (1 = i = p), key(i) is estimated according to equation 1.

Key(i) =<S, i>

The index value is estimated according to Eq. 2

index = CoinToss (key(i), n)

The Q^S value is assigned according to Eq. 3

$Q^S_{[index]} = 1$

Then the i value is incremented according to Eq. 4

i = i + 1

Then Step 4, 5 and 6 is repeated till i = p

The final Q^S value is recorded

Next each S encrypts the Q^S value using the homomorphic encryption based on StPKE.

The Q^S value is encoded according to equation (5).

$E^S = Q^S \| 0^Z$

Based on the HKDF scheme, the X determines two key values: K^1 and K^2

Then, the C value is estimated according to Eq. 6

$C = K^1 + e \mod M$

Then, S estimates the MAC value of Q^S according to Eq. 7

$MAC(Q^S) = HMAC(C, K^2)$

The trusted aggregator node, X generates its local synopsis value, Q^X in the similar manner as followed by the sensor nodes by following step 4, 5 and 6 till i = p

Then the X encodes its Q^X according to Eq. 8

$E^X = Q^X \| 0^Z$

Based on the HKDF scheme, the X determines two key values: K^1 and K^2

Then X estimates the C value according to Eq. 6

Next, X estimates the MAC value of Q^X according to Eq. 9

$MAC(Q^X) = HMAC(C, K^2)$

Each X also receives the MAC value from the m surrounding S.

X checks the T^S of each S, by comparing it with T^{Thresh}

If $T^S < T^{Thresh}$, then the S is omitted

If $T^S > T^{Thresh}$, then the S is considered as trustworthy.

All the trustworthy S are aggregated by the X, by XORing the MACs according to equation (10) and C are aggregated by addition operation modulo M according to Eq. 11

$MAC_{agg} = MAC_{S1}$ XOR MAC_{S2} XOR ..... XOR MACSl

$C_{agg} = C_{S1} \mod M + C_{S2} \mod M + ... + C_{Sl} \mod M$

X forwards the aggregated data to the BS.

On receiving the aggregated data, the BS decrypts the data.

If the decrypted data matches the original data sent by corresponding X, then the data is considered as valid.

Thus, the data is securely aggregated and transmitted to the BS

RESULTS AND DISCUSSION

Simulation model and parameters: The Network Simulator (NS-2) is used to simulate the proposed architecture. In the simulation, 100 mobile nodes move in

Table 2: Simulation parameters

Parameters	Values
No. of nodes	100
Area size	500×500
Mac	IEEE802.11
Transmission range	250m
Simulation time	50 sec
Traffic source	CBR
Packet size	512
Initial energy	15.1J
Transmission power	0.660
Receiving power	0.035
Attackers	1-5

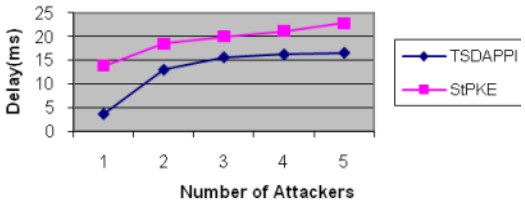

Fig. 6: Aggregation latency

a 500× 500 m region for 50 sec of simulation time. All nodes have the same transmission range of 250 m. The simulated traffic is Constant Bit Rate (CBR). The simulation settings and parameters are summarized in Table 2.

Performance metrics: The proposed Trust based Secured Data Aggregation for Privacy Protection and Integrity (TSDAPPI) is compared with the StPKE scheme (Boudia *et al*., 2015). The performance is evaluated mainly, according to the following metrics.

Packet delivery ratio: It is the ratio between the number of packets received and the number of packets sent.

Packet drop: It refers the average number of packets dropped during the transmission.

Delay: It is the amount of time taken by the nodes to transmit the data packets.

The number of misbehaving and false injector attackers is varied from 1-5 and the performance is evaluated for both the techniques. Figure 6 shows the aggregation latency of TSDAPPI and StPKE for varying attackers. When, the attackers are increased, it involves lot of verification procedures, thereby increasing the latency. However, TSDAPPI has 35% reduced latency when compared to StPKE since only the trusted reports are aggregated and the synopsis has smaller size.

Figure 7 shows the communication overhead of TSDAPPI and StPKE for varying attackers. When the attackers are increased, it involves lot of packet and key exchanges, thereby increasing the overhead. However,

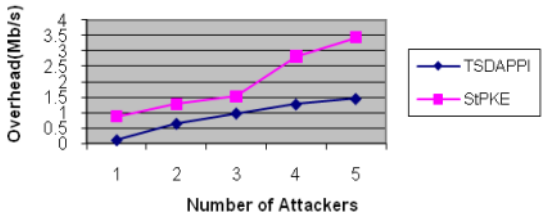

Fig. 7: Communication overhead

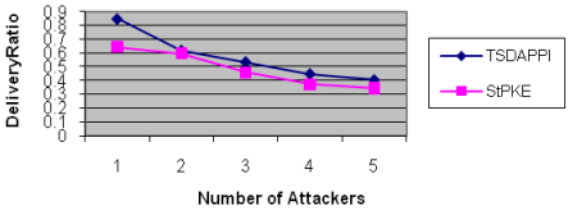

Fig. 8: Packet vs delivery ratio

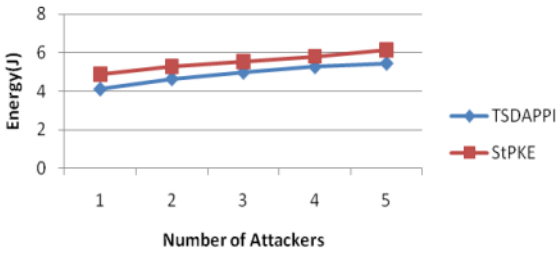

Fig. 9: Energy consumption

TSDAPPI has 56% reduced overhead when compared to StPKE since only the trusted reports are aggregated and the synopsis has smaller size.

Figure 8 shows the packet delivery ratio of TSDAPPI and StPKE for varying attackers. When the attackers are increased, there will be more packet drops thereby decreasing the delivery ratio However, Since TSDAPPI detects both insider misbehaving and outsider attacks, it attains 14% increased delivery ratio, than StPKE.

Figure 9 shows the energy consumption of TSDAPPI and StPKE for varying attackers. When the attackers are increased, it involves lot of verification and packet exchanges, thereby increasing the overhead. However, TSDAPPI has 11% reduced energy consumption when compared to StPKE since the aggregators are selected with high residual energy.

CONCLUSION

In this study, trust based secured data aggregation for privacy protection and Integrity has been proposed.

Here, reputation-based technique is used to analyze the behavior of the nodes participating in clustering process. Aggregator nodes are selected based on the efficiency of the nodes in terms of residual power, trust, etc. Then local synopsis is generated and aggregated using the StPKE mechanism. This data is encrypted, aggregated and then transmitted to the base station by the aggregator node. By simulation results, we can conclude that the proposed technique reduces the packet drops due to attacks and improves the packet delivery ratio.

REFERENCES

Bhasker, L., 2014. Genetically derived secure cluster-based data aggregation in wireless sensor networks. Inf. Secur. IET., 8: 1-7.

Boudia, O.R.M., S.M. Senouci and M. Feham, 2015. A novel secure aggregation scheme for wireless sensor networks using stateful public key cryptography. Ad Hoc Networks, 32: 98-113.

Castelluccia, C., A.C. Chan, E. Mykletun and G. Tsudik, 2009. Efficient and provably secure aggregation of encrypted data in wireless sensor networks. ACM. Trans. Sensor Netw., Vol. 5, 10.1145/1525856.1525858

Dhasian, H. and P. Balasubramanian, 2013. Survey of data aggregation techniques using soft computing in wireless sensor networks. Inf. Secur. IET., 7: 336-342.

He, W., H. Nguyen, X. Liu, K. Nahrstedt and T. Abdelzaher, 2008. IPDA: an integrity-protecting private data aggregation scheme for wireless sensor networks. Proceedings of the 2008 IEEE Conference on Military Communications (MILCOM), November 16-19, 2008, IEEE, San Diego, California, USA., ISBN: 978-1-4244-2676-8, pp: 1-7.

Huang, S.I., S. Shieh and J.D. Tygar, 2010. Secure encrypted-data aggregation for wireless sensor networks. Wireless Networks, 16: 915-927.

John M.J., S. Prince and A. Rakesh, 2013. Secure data aggregation and data recovery in wireless sensor networks. Intl. J. Eng. Adv. Technol., 2: 271-275.

Justus, J.J. and A.C. Sekar, 2013. A fault tolerance data aggregation scheme for wireless sensor networks. Intl. Rev. Comput. Software, 8: 1556-1563.

Liu, C.X., Y. Liu and Z.J. Zhang, 2013. Improved reliable trust-based and energy-efficient data aggregation for wireless sensor networks. Intl. J. Distrib. Sensor Netw., 2013: 1-11.

Mansouri, M., L. Khoukhi, H. Nounou and M. Nounou, 2013. Secure and robust clustering for quantized target tracking in wireless sensor networks. Commun. Netw. J., 15: 164-172.

Ozdemir, S., 2008. Functional reputation based reliable data aggregation and transmission for wireless sensor networks. Comput. Commun., 31: 3941-3953.

Rezvani, M., A. Ignjatovic, E. Bertino and S. Jha, 2015. Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. Dependable Secure Comput. IEEE. Trans., 12: 98-110.

Roy, S., M. Conti, S. Setia and S. Jajodia, 2014. Secure data aggregation in wireless sensor networks: Filtering out the attacker's impact. Inf. Forensics Secur. IEEE. Trans., 9: 681-694.

Stelte, B. and A. Matheus, 2011. Secure trust reputation with multi-criteria decision making for wireless sensor networks data aggregation. Proceedings of the Conference on Sensors, 2011 IEEE, October 28-31, 2011, IEEE, Limerick, Ireland, ISBN: 978-1-4244-9290-9, pp: 920-923.