Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 94 (2016) 216 – 223

Procedia
Computer Science

The 11th International Conference on Future Networks and Communications (FNC-2016)

Privacy preserving in-network aggregation in wireless sensor networks

Vishal Krishna Singh[a]*, Saurabh Verma[b], Manish Kumar[c]

[a, b, c]*Indian Institute Of Information Technology, Jhalwa, Allahabad, India*

Abstract

The use of attack-prone hardware makes designing secure schemes for data collection a complex task. Addressing the continuous threat of attacks at the aggregator nodes, this work proposes a privacy preserving secure in-network data aggregation (PPSDA) technique for wireless sensor networks (WSNs). Using the pallier crypto system, a secure scheme which is resilient to false data injection attacks, is devised to compute the SUM, COUNT and MEAN at the sink node. Extensive theoretical analysis and simulation results show that the proposed scheme outperforms other existing aggregation approaches

Keywords: In-network data aggregation; network lifetime; pallier cryptosystem; privacy; wireless sensor networks

1. Introduction

Tactically placed sensor nodes collaborate with each other to form an ad-hoc network which is capable of reporting occurrence of events to a data collection sink. An aggregate function e.g. SUM, COUNT, MEAN is typically used to compute the aggregated values at sink. The intermediate nodes can be used to combine partial results, during the routing of the message, to reduce the network traffic load and thus improve the network lifetime[1]. Transmitting partial results, in a tree-based structure involves each node collecting response from all its children before sending the final result. However, event based systems, are often compromised by false data injection attack, which aims at compromising the intermediate nodes and send false data to the sink, causing energy depletion in the network. In such a scenario, the tree based approaches fail to maintain the data confidentiality and integrity as they are not robust to cope up with communication losses. At each failure, an entire sub tree is lost resulting in unaccounted data at the sink and thus wrong aggregated values. Computing aggregates with multi path routing could

* Corresponding author. Tel.: +919559626633; *E-mail address:* vashukrishna@gmail.com

doi:10.1016/j.procs.2016.08.034

be a probable solution but preventing an insider attack imposed by compromised intermediate nodes is a complex task[2]. Considering the possibility of compromised nodes, recent works have focused on various encryption techniques for maintaining the data confidentiality but every time the data is decrypted for aggregation, at the intermediate nodes, it becomes vulnerable to attacks.

Secure data transmission in WSN involves encrypting the sensed data and decrypting the data at the intermediate nodes for aggregation. The decrypted data is aggregated and is encrypted again before forwarding it to the next designated node in the communication channel. The process is followed until the data reaches the sink where the final aggregated values are obtained. In such a scenario, the confidentiality of the sensed data as well as the aggregated value is hard to maintain[3]. Also, the amount of energy consumed in encryption and decryption might not be an issue in a small network, but as the size of the network grows, it becomes increasingly difficult to manage the energy consumption without compromising the security of the data. Compromising the sensor nodes and using them for imposing various insider attacks on the network, can easily disrupt the operation of sensor networks[3,4]. This work however, is aimed at a special insider attack known as false data injection attack, where the intermediate nodes are attacked to insert malicious data as soon as the data is decrypted for aggregation. Specifically, a data aggregation scheme, which is resilient to such attacks, is presented for WSNs. The proposed scheme uses homomorphic encryption for maintaining the confidentiality of the data and in-turn preserves huge amount of energy, which would have been lost in the decrypting the data at the intermediate nodes.

The rest of the paper is organized as follows: the section 2 presents the related work, in section 3 the proposed confidentiality preserving data aggregation scheme is explained, section 4 describes the experimental and simulation setup, in section 5 the results are presented with their detailed explanation and finally the section 6 concludes the work.

2. Related Work

Sensor nodes are typically constrained in terms of communication bandwidth, memory, energy resources and computation power. Implementing security protocols in such an environment becomes increasingly difficult with the attackers physically tampering the nodes. Moreover, because of the wireless medium, the confidentiality of the data is compromised by the attackers by accessing the communication among the nodes. Existing encryption techniques prevent unauthorized nodes from injecting malicious information into the data. However, such encryption techniques fail to prevent attack launched by compromised nodes. A recent work[5] studied the routing protocols and their vulnerabilities with compromised nodes in the network. The authors presented several attacks, such as wormhole attack, hello flood attack etc. to obstruct normal data routing in a WSN. Energy efficient methods of in-network aggregation[6,7], identify tree based routing for computing the aggregates such as SUM, COUNT and MEAN. The aggregation technique[6] (TAG) to compute the Count and the Sum. Partial counts, from the child nodes, are added to each node and the value is incremented by one. The sub-aggregate is then forwarded to the parent node until the sink is reached. An aggregation protocol[8], for the decentralized aggregate computation is a gossip-based protocol which assumes the nodes to form an overlay network. Such an assumption is impractical for WSNs as it allows any pair of nodes to be considered as neighbors.

Privacy preservation schemes[9,10,3] aim at keeping the individual readings secret from the intermediate nodes acting as aggregators. Such schemes are advantageous because of their ability to provide end-to-end confidentiality at minimal energy consumption.

Thus, this work is motivated by the need of providing end-to-end security and preserving the privacy of the sensed data. Maintaining the end-to end privacy is achieved by using the homomorphic properties of the paillier cryptosystem, and the use of a predefined threshold allows easy computation of the aggregates at the sink.

3. Proposed Privacy Preserving In-Network Aggregation

Paillier cryptosystem is a homomorphic encryption technique that uses an asymmetric algorithm based on public key cryptography. The proposed scheme uses the paillier cryptographic technique to ensure the security and energy constraints of the network. The proposed PPSDA aggregates data in such a way that the aggregated value of SUM, COUNT and MEAN is easily calculated at the sink without decrypting the sensed data at any point without imposing any extra burden over the network.

3.1 The steps of the proposed PPSDA are as follows:

i. Let the S_i is the sensed value by node i.

ii. The sensed value i.e. S_i , is appended by multiplying the sensed value with a predefined threshold, k.

iii. Leaf node encrypts the appended value, i.e. $e(kS_i)$

iv. Suppose m is a message to be encrypted such that $m \in Z_n$.

v. Select random s where $s \in Z_n$

vi. Compute cipher text as: $C = rm.sn \bmod n^2$

vii. The parent node performs the aggregation operation on encrypted values $(\text{Agg}(\text{e}(\text{Si})))$ and forwards the encrypted data to the next node in the communication chain until the data reaches the sink.

viii. At the sink, the data is decrypted to obtain the final aggregated value.

ix. For calculating the SUM at sink, the aggregated value i.e. X is divided by the predefined threshold of k. The floor of the decimal will give the SUM of all the sensed data.

x. For finding the COUNT, modulo of X is calculated.

xi. Finally, the MEAN is obtained by dividing the SUM by the COUNT.

xii. The intermediate nodes are used only for the aggregation. The homomorphic property of paillier cryptosystem allows addition in the data without decrypting the original data.

xiii. The proposed PPSDA approach makes use of the additive homomorphic property of paillier algorithm.

xiv. All nodes perform SUM on the aggregated values without decrypting them, thus maintaining the confidentiality and integrity of the original data.

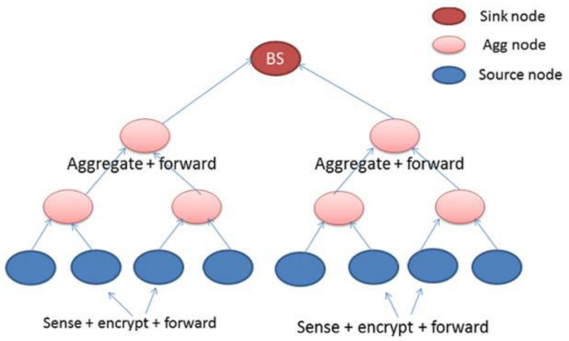

Fig. 1. Privacy Preserving Secure In-network Data Aggregation (PPSDA)

3.2 Determining the Threshold

As discussed, the predefined threshold, ‘ k ’ is an important factor for computing the aggregate functions i.e. SUM, COUNT and MEAN. The threshold ‘ k ’ is defined as:

$$\left\{k = 10^{\alpha} \quad , \alpha = number\ of\ digits \text{ in } N\right\}$$

where, N is the number of deployed nodes.

3.3 Energy Consumption Analysis

The energy consumption analysis in a WSN involves parameters such as the energy dissipated by the transmitter to run the radio electronics and the power amplifier, energy dissipated by the receiver to run the radio electronics etc. To analyse the energy consumption, a simple model for the hardware energy dissipation is considered. The free space model (d^2 power loss model) and the multipath fading (d^4 power loss) channel models were considered. If the distance is less than the threshold " C_0 ", the free space model is used, else multipath model is used. Therefore, for a "m" bit message, to be transmitted to a distance "s", the energy dissipated by the radio is given by:

$$E_{tx}(m,s) = E_{tx-elec}(m) + E_{tx-amp}(m,s) = \begin{Bmatrix} mE_{elec} + m\in_{FS} c^2 & s < C_0 \\ mE_{elec} + m\in_{MP} c^4 & s \geq C_0 \end{Bmatrix} \quad (1)$$

In order to receive this message, the energy consumed by the radio can be given by:

$$E_{rx}(m) = E_{rx-elec}(m) = mE_{elec} \quad (2)$$

Where:-

- ✓ E_{elec} is the energy required to run the electronics.
- ✓ $\in_{FS} c^2$ and $\in_{MP} c^4$, depends upon the distance between the transmitter and the receiver.

4. Results and Discussion

For testing the proposed CPSDA, a network of 23 e-motes was deployed randomly in an area. The communication range of the nodes was set to 10 meters.

In order to test the method over large networks, the proposed scheme was also simulated in MATLAB with 400 nodes in an area of $100m \times 100m$ with sink placed at the center. For a comparative analysis of the proposed PPSDA, two different secure aggregation based schemes i.e. SDAP[11] and SEEDA[12] were also implemented with the same set of parameters. A detailed description of the simulation parameters is given in Table 1.

Table 1. Simulation Parameters.

Parameters	Values
Deployment area	$100m \times 100m$
Number of nodes	100
Position of sink	$50m \times 50m$
Initial energy	0.10 Joules
Message size	4000 bits

4.1 Experimental Results

The proposed PPSDA was implemented over 24 e-motes for 10 round of data collection. The average number of packets transmitted by each node is shown in fig. 2.

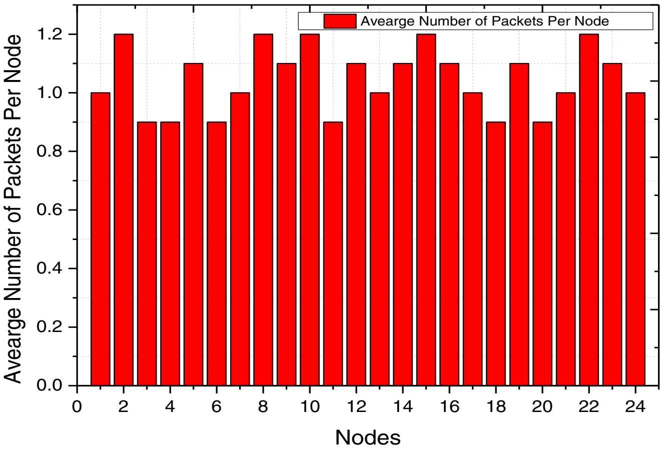

Fig. 2. Average number of packets transmitted per node

As shown in fig.2, the average number of packets transmitted by every node in the experiment, are almost equivalent. Apart from the leaf nodes, the intermediate nodes also add their data after encryption to the received encrypted value. Homomorphic property of the paillier algorithm allows addition of the two encrypted values without any decryption at the intermediate node. The homomorphic addition results in an encrypted aggregate at the intermediate node. Hence, as seen in the fig. 2, the number of data packets at every node remain almost the same.

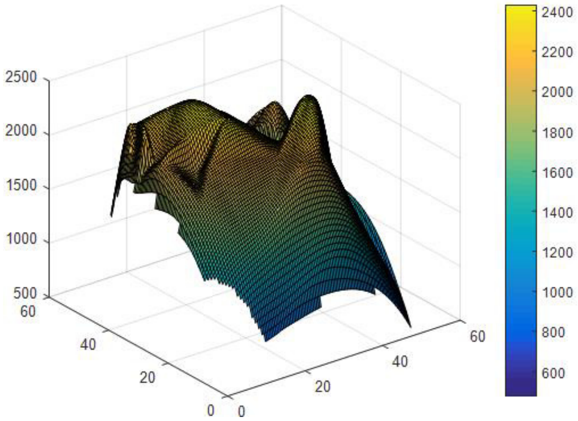

Fig. 3. Total sending and receiving activity of the nodes for a randomly deployed network.

The fig. 3 shows the node activity (total number of transmitted and received bits) for one data collection round. As seen in the fig. 3, the node activity is almost uniform throughout the network. The nodes performing the aggregation show a slightly high activity as the energy consumption of such nodes is relatively higher.

4.2 Simulation Results

The accuracy of the results obtained from a deployed sensor network, relies upon the number participating nodes after a certain number of data collection rounds. Hence, in this work the network is considered dead after 50 % of the total deployed nodes die. The performance of the proposed PPSDA is verified and compared for the following evaluation parameters:

Stability Period and Network Lifetime Analysis: As shown in fig. 4, the stability period of the proposed PPSDA is relatively much better as compared to the SDAP[11] and SEEDA[12].

The homomorphic property of the paillier algorithm allows the intermediate nodes to add their data to the encrypted values obtained from the leaf nodes. Thus, the need of decrypting the data for adding information at the intermediate nodes is no longer an issue and hence significant amount of energy is preserved. As a result, the first node dies in the 139th round in the proposed PPSDA. The stability period of the network improves up to 25.39 % as compared to

SDAP and SEEDA approaches where the first node dies in 108th and 112th round respectively. The effect of improved stability period is seen in the lifetime of the network. Almost 50 % of the deployed nodes die between 183rd and 209th round of data collection in the SDAP and SEEDA approaches. However, with the proposed PPSDA, the network remains alive until the 270th round after which the 51st node dies.

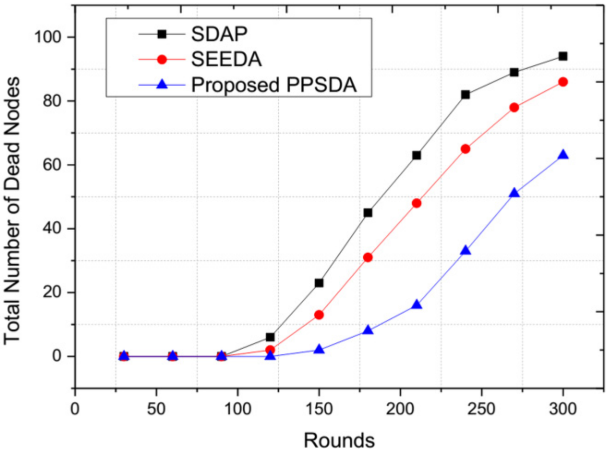

Fig. 4. Total number of dead nodes over data collection rounds.

Data Transmission and Network Traffic Analysis: The fig. 5 shows a comparative distribution of the total number of packets sent to the sink after the completion of 300 data gathering rounds for the proposed scheme as well as for the two compared approaches.

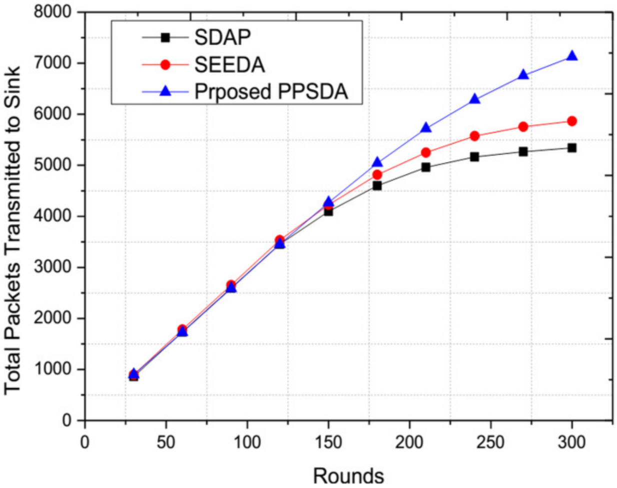

Fig. 5. Total packets transmitted to sink over rounds.

As shown in the fig. 5, for 300 rounds of data collection rounds, the proposed PPSDA achieves high transmission (up to 7128 packets) as compared to the 5341 packets in SDAP and 5867 packets in SEEDA. The reason for this improved data transmission and collection is the use of homomorphic encryption, which not only ensures the privacy of the data but also allows the aggregation of the data to be done easily with minimal energy consumption. The homomorphic encryption algorithm allows homomorphic addition of the data values at intermediate nodes as a result, the network remains secure and alive for a longer period of time and hence the total number of packets sent to the sink is relatively higher in the proposed PPSDA. After the 132nd round of data gathering, the connectivity in the network of the SDAP and SEEDA protocols, is compromised because of the dying nodes. As the data collection rounds are increased to 300, more nodes die and the connectivity gets poorer resulting in per node data transmission to fall up to 17.80 packets per node in SDAP and 19.55 packets per node in SEEDA (fig. 6). However, the proposed PPSDA maintains 23.47 packets per round even at the 300th round of data collection, which is relatively much higher as compared to the SDAP and SEEDA.

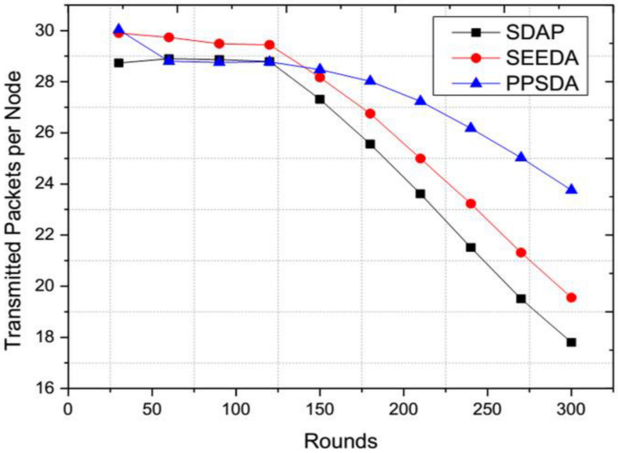

Fig. 6. Total packets transmitted per node over rounds.

Accuracy and Efficiency Analysis: The accuracy and efficiency of the proposed PPSDA is tested on the basis of total packets lost in the data transmission process.

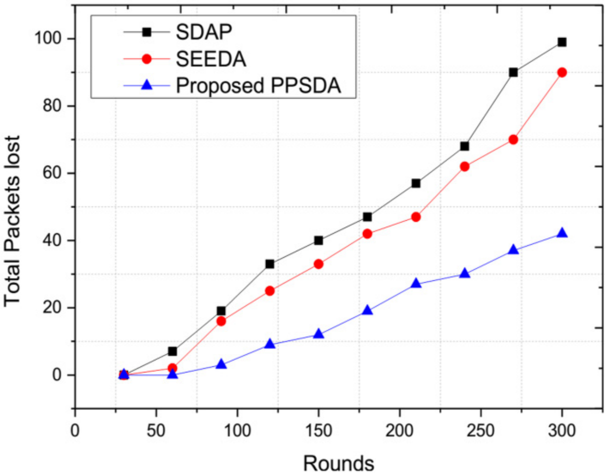

Fig. 7. Total packets lost over rounds.

As shown in the fig. 7, the number of packets lost within the data transmission process is significantly low in the proposed PPSDA and reaches a maximum of 42 packets at 300th round of data collection. A significant high packet loss is seen in the SDAP and SEEDA where at 300th round of data collection the number of lost packets is very high at 99 and 90 packets respectively. The significant difference in the number of packets lost in the data transmission, between the three approaches, is because the proposed PPSDA has better lifetime resulting in better connectivity between the nodes and that too for longer period of time. The nodes remain alive for a longer period and hence provide better communication within the network resulting in reduced number of lost packets. Also, the encryption process ensures the safety and privacy of the data resulting in improved data transmission.

Privacy-preservation: The "slicing and assembly" technique[5] ensures the privacy of the data by the nodes slicing their data and sending encrypted slices of the data to the aggregators until the data reaches the sink. Another approach, SDAP[11], presented provides hop-by-hop security of the data. The design is based on divide-and-conquer and commit-and-attest principles. The group aggregates are identified by the sink for any suspicious value, resulting in the attestation process of the identified group to prove the correctness of its group aggregate. However, the proposed PPSDA ensures the privacy of the data in a different way. The nodes, on detecting an event, multiply their reading with the predefined threshold and encrypt the data using the paillier algorithm. The algorithm is a homomorphic algorithm and is completely secure because of its public key encryption method. The data once encrypted, is only decrypted at the sink and not before. The intermediate nodes encrypt their readings and perform the aggregation on the encrypted data received from their children nodes. The homomorphic addition enables the addition at the intermediate nodes without decrypting the data in between the communication channel.

5. Conclusion

The proposed PPSDA aims at achieving end-to-end security within the constraints of WSNs. Paillier cryptosystem is used for ensuring the privacy of the sensed data and use of a threshold based aggregation mechanism that allows computation of aggregates at the sink with minimum energy consumption. The proposed scheme is able to achieve approximately 25 % efficiency in terms of stability period of the network as compared to the other existing approaches. The proposed PPSDA achieves better transmission results as compared to existing tree based secure aggregation approaches owing to the improved lifetime and reduced data loss. The homomorphic addition at the intermediate nodes for aggregation allows the data to be decrypted only at the sink, ensuring the end-to-end privacy of the data.

References

1. S. Roy, M. Conti, S. Setia, and S. Jajodia, "Secure Data Aggregation in Wireless Sensor Networks: Filtering out the Attacker's Impact," IEEE Trans. Inf. Forensics Secur., vol. 9, no. 4, pp. 681–694, Apr. 2014.
2. L. C. Jain, S. Patnaik, and N. Ichalkaranje, Eds., Intelligent Computing, Communication and Devices, vol. 309. New Delhi: Springer India, 2015.
3. C.-X. Liu, Y. Liu, Z.-J. Zhang, and Z.-Y. Cheng, "High energy-efficient and privacy-preserving secure data aggregation for wireless sensor networks," Int. J. Commun. Syst., vol. 26, no. 3, pp. 380–394, Mar. 2013.
4. M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks," IEEE Trans. Dependable Secur. Comput., vol. 12, no. 1, pp. 98–110, Jan. 2015.
5. C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," Ad Hoc Networks, vol. 1, no. 2–3, pp. 293–315, Sep. 2003.
6. S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG," ACM SIGOPS Oper. Syst. Rev., vol. 36, no. SI, p. 131, Dec. 2002.
7. M. B. Greenwald and S. Khanna, "Power-conserving computation of order-statistics over sensor networks," in Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '04, 2004, p. 275.
8. M. Jelasity, A. Montresor, and O. Babaoglu, "Gossip-based aggregation in large dynamic networks," ACM Trans. Comput. Syst., vol. 23, no. 3, pp. 219–252, Aug. 2005.
9. J. Girao, D. Westhoff, and M. Schneider, "CDA: concealed data aggregation for reverse multicast traffic in wireless sensor networks," in IEEE International Conference on Communications, 2005. ICC 2005. 2005, 2005, vol. 5, pp. 3044–3049.
10. C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2005, pp. 109–117.
11. Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP," ACM Trans. Inf. Syst. Secur., vol. 11, no. 4, pp. 1–43, Jul. 2008.
12. A. S. Poornima and B. B. Amberker, "SEEDA: Secure end-to-end data aggregation in Wireless Sensor Networks," in 2010 Seventh International Conference on Wireless and Optical Communications Networks - (WOCN), 2010, pp. 1–5.