*Journal of Ubiquitous Systems & Pervasive Networks*
*Volume 9, No. 2 (2017) pp. 15-19*

# Evaluation of Privacy Preserving In-Network Aggregation for Different Routing Structures in WSNs

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

[a]*Indian Institute of Information Technology, Devghat, Jhalwa, Allahabad, 211012, Uttar Pradesh, India.*

## Abstract

Designing secure aggregation protocols for efficient data collection is driven mainly by preserving the privacy and integrity of the network data. Targeting the false data injection attack in wireless sensor networks (WSNs), a secure in-network aggregation scheme for preserving the privacy of the data, is proposed in this work. The proposed scheme is able to compute the SUM, MEAN and COUNT of the network data based on an encryption scheme which uses pallier cryptosystem. Extensive analysis with different routing structures (cluster, tree and chain) are used for validating the efficacy of the proposed scheme over existing approaches.

***Keywords:*** *In-network data aggregation, Network lifetime, Pallier cryptosystem, Privacy, Wireless sensor networks*

## 1. Introduction

Sensor nodes are deployed with the objective of collaborative reporting of events in the proximity to the sink. Owing to the constraints of WSNs, sending huge amount of data through the network is not reliable and moreover unnecessary [1]. In-network aggregation and processing schemes are therefore used for optimizing the network traffic and improve the network lifetime. Aggregate functions such as AVERAGE, SUM, PRODUCT etc. are used to obtain a summary of the network data at the sink. The intermediate nodes, in a typical multi-hop routing path, are responsible for processing partial results and obtaining the desired aggregate. For example, in a tree based routing structure each node transmits its data only after it has received the responses from all its children. However, such a system can be easily compromised by false injection attacks [2]. Such an attack is aimed to curb the reliability of the data by attacking the intermediate nodes and inducing false data. With the failure of each node, not only the confidentiality of the data is lost but each failure also results in the loss of a sub tree which results in data loss at the sink and thus generating erroneous aggregated values at the sink. A possible solution to this problem is the use of multipath routing for computing the aggregates [3]. But the reliability of the data can still not be ensured due to the high probability of nodes being compromised. Encryption techniques as such, can be used for maintain the secrecy of the data but with the compromised intermediate nodes, the decryption of the data for in-network aggregation, makes the data vulnerable to attacks [4]. The decryption and encryption of the data at every hop not only affects the confidentiality of the data but is also an energy exhausting task. The energy consumption during the encryption and decryption might not be reflected in small networks but with large scale networks, managing the networks energy consumption without affecting the confidentiality of the data is a difficult task. In our previous work addressing the issue of energy efficient and secure in-network aggregation [13], a privacy preserving scheme i.e. PPSDA, was proposed. This work, however is aimed at evaluating the performance of the already proposed PPSDA with different routing structures. The PPSDA is based on homomorphic encryption for ensuring the data confidentiality.
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

Secure data transmissions in WSNs is a difficult task because of the various inherent constraints of such networks such as, limited processing ability, limited memory and communication bandwidth. Additionally, the probability of the sensor nodes to be physically tampered is significantly high. As such, implementing security protocols for maintaining the data confidentiality for such wireless transmissions, are difficult. Existing security protocols allow encryption of the network data for maintaining the data confidentiality by hiding the original data through a cipher. However, such an approach is

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

DOI: 10.5383/JUSPN.09.02.002

prone to attack when the data is decrypted at intermediate nodes for aggregation. Thus, encryption fails to achieve an attack resilient scheme for securing the data privacy. The work presented in [5] discusses various routing protocols which are used in an attack prone environment along with their disadvantages. The work presents a detailed analysis on various possible attacks in a WSN such as false injection attack, wormhole attack etc. Most of the existing security protocols have failed to achieve end-to-end security and confidentiality of the data because of the necessity of decryption for calculation the aggregates at the intermediate nodes. Such methods not only fail in maintaining the confidentiality of the data but also are energy exhaustive because of the complex computations involved in the decryption at each hop. The TAG protocol, proposed in [6], aims at computing the COUNT and SUM of the network data. 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 improved aggregation scheme, proposed in [8], is aimed at decentralized aggregate computation. The scheme is based on a gossip-based protocol which assumes the nodes to form an overlay network. However, such a scheme is unrealistic in a WSN as it allows random nodes to be considered for pairing and being used as neighbors. Handling the privacy of the network data, the schemes proposed in [9], [10] and [11], aim at maintaining the secrecy of the network data by hiding the nodes readings from all its neighbors by using end-to-end encryption. The PPSDA proposed in [13] uses pallier cryptosystem to ensure homomorphic encryption of the data. The idea is to compute SUM, MEAN and COUNT of the network data without decrypting the data at any intermediate hop. The work presented here is an extension of the work presented in [13] and is aimed at analyzing the performance of the proposed PPSDA with various routing structures.

## 3. Privacy Preserving Data Aggregation

The Paillier encryption scheme, used in this work, is a homomorphic encryption technique which is based on the idea of asymmetric encryption using public key encryption scheme. The homomorphic encryption allows the proposed PPSDA to calculate the aggregated values of SUM, COUNT and MEAN at the sink in such a way that the encrypted data needs not to be decrypted at any intermediate node. Releasing the intermediate nodes from the complexities of endless decryption, the proposed PPSDA is able to significantly reduce the processing load of the network. Additionally, performing the aggregation operations on the encrypted data ensures end-to-end security allowing the proposed PPSDA to maintain the confidentiality and privacy of the data.

### 3.1. The working of the proposed PPSDA can be explained by the following steps:

i. Let the reading sensed by the i[th] node is given by $S_i$

ii. A predefined threshold 'k' is multiplied to the sensed reading $S_i$ .

iii. The product obtained (in step ii), is encrypted by the leaf node i.e. $e(kS_i)$

iv. The cipher is computed by: $C = rm.sn \bmod n^2$ such that, '$m$' is the original message to be encrypted and belongs to $Z_n$ .'s' is randomly selected such that $s \in Z_n$

v. Aggregation is performed by the parent node on the encrypted values $(\mathrm{Agg}(\mathrm{e}(\mathrm{S}_i)))$ and the resulting value is forwarded to the next hop in the communication hierarchy. This process is followed until the data reaches the sink.

vi. The aggregated value (X) is finally obtained at the sink by decrypting the data.

vii. At sink, the following operations are performed to obtain SUM, COUNT and MEAN:
- SUM is obtained by applying 'floor' operation to the decimal of $X/k$
- MOD of 'X' gives the COUNT
- MEAN is calculated by dividing the SUM by COUNT.

viii. Because of the homomorphic nature of Paillier cryptography, additions can be performed at intermediate nodes without decrypting the data at any point.

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

### 3.2. Pre-determined Threshold 'k'

The threshold 'k' is the most important factor required for calculating the aggregates at the sink, and is computed as:

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

Such that, $N$ represents the total number of nodes deployed in the network.

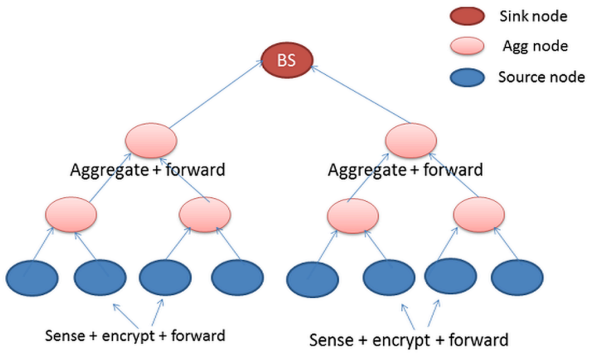


**Fig. 1. Privacy Preserving Secure In-network Data Aggregation (PPSDA)**

### 3.3. Energy Consumption Analysis

The major energy consuming tasks in a WSN involve data processing and data transmission. To analyze the energy dissipated in the process of reporting an event, the basic free space model ( $\mathrm{d}^2$ power loss model) and the multipath fading ( $\mathrm{d}^4$ power loss) channel models are considered in this analysis. The energy consumption is measured on the basis of the free space model given the distance is less than a predefined threshold "$C_0$", else multipath model is used. Thus, for a 'm'

bit message, to be transmitted to a distance "s", the energy dissipated by the radio is given by:

$$E_{tx}(m,s) = E_{tx-elec}(m) + E_{tx-amp}(m,s)$$

$$= \begin{cases} mE_{elec} + m\in_{FS} c^2 & s < C_0 \\ mE_{elec} + m\in_{MP} c^4 & s \geq C_0 \end{cases} \quad (1)$$

The energy dissipated in receiving this message is given by:

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

Such that:

- ✓ $E_{elec}$ represents the energy needed for running the electronics.
- ✓ $\in_{FS} c^2$ and $\in_{MP} c^4$, vary with the distance between the transmitter and the receiver.

## 4. Results and Discussion

The proposed PPSDA was tested and evaluated in different scenarios for identifying the best suited environment for practical implementation. Simulations were performed in MATLAB for a 400 node network, randomly deployed in an area of $100m \times 100m$ with the sink positioned at the center of the deployed area. A detailed comparative analysis of the proposed PPSDA along with the SDAP scheme proposed in [11] and SEEDA scheme proposed in [12], is presented for three different routing scenarios, specifically for:

- Cluster based routing
- Tree based routing
- Chain based routing

The initial energy of every node is considered to be 0.10 Joules and the message is considered to be of 1024 bytes. The objective of this analysis is to find the most suited routing environment for the proposed PPSDA. Therefore, based on the simulation results, the proposed PPSDA is implemented for a 23 node network such that nodes follow a clustered routing based on a standard clustering algorithm [14]. The details of the simulation setup are given in Table 1.

Table 1. Simulation Parameters.

| Parameters | Values |
|---|---|
| Deployment area | $100m \times 100m$ |
| Number of nodes | 400 |
| Position of sink | $50m \times 50m$ |
| Initial energy | 0.10 Joules |
| Message size | 1024 bytes |

### 4.1. Simulation Results

The proposed PPSDA is evaluated for the following parameters:

***Network Lifetime:*** The fig. 2 shows the stability period achieved by the proposed PPSDA along with SDAP and SEEDA approaches with different routing structures (a) clustered routing (b) tree based and (c) chain based. As shown in the fig. 2(a), the proposed PPSDA has the best performance in terms of stability period (number of rounds before the first node dies), as the number of nodes involved in transmission are fairly low in clustered routing structure. The effect is seen in the fig. 2(a), where the first node dies after 269 rounds of data collection as only the nodes with events in their proximity are responsible for transmitting the data. Since the homomorphic addition of the sensor data prohibits data decryption at intermediate nodes, a large amount of energy is conserved and the same is reflected in the fig. 2(b), where the proposed PPSDA achieves almost 25.39 % improvement in the stability period as compared to the SDAP and SEEDA schemes. The fig. 2(b) and 2(c) prove that the efficacy of the proposed PPSDA is maintained in other routing schemes as well. The proposed PPSDA is able to conserve significant energy by allowing decryption only at the sink because of which, the first node dies only after 269, 147 and 201 rounds of data collection in clustered, tree based and chain based routing respectively.

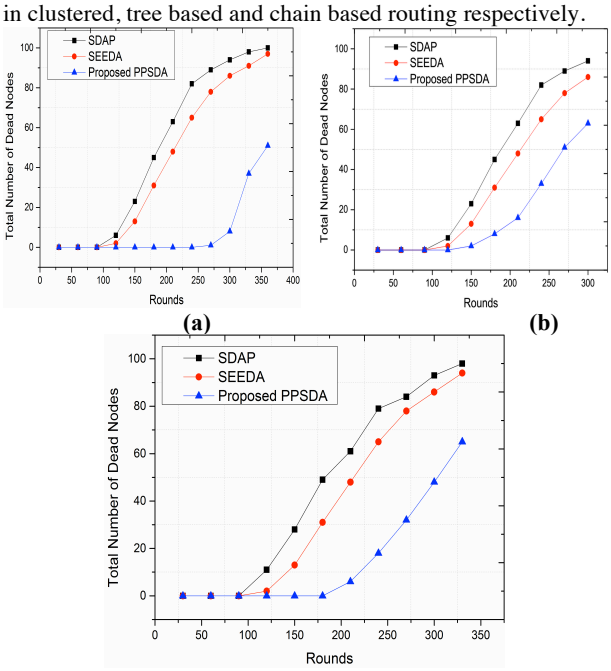


**Fig. 2. Performance Evaluation in terms of Network Lifetime (a) Clustered Routing (b) Tree Based Routing (c) Chain Based Routing.**

***In-Network Traffic:*** The in-network traffic is evaluated on the basis of total number of packets transmitted within the network with varying number of data collection rounds. The fig. 3 shows the performance of the proposed PPSDA, SDAP and SEEDA approaches with different routing schemes. As shown in the fig. 3, the proposed PPSDA outperforms both the existing schemes with clustered routing structure with a significant margin. However, with tree based and chain based routing schemes, the performance of the proposed PPSDA is marginally better than SDAP and SEEDA. The reason for this increase in the in-network traffic with tree and chain based routing is that, in a clustered routing structure, all the nodes send their encrypted data to the cluster head (CH), which performs the aggregation of the data. Therefore, only the nodes which have events in their proximity are bound to send the data packets, and thus very few nodes are involved in the data transmission. However, with tree based and chain based routing, the nodes are forced to forward the data packets even if there is no event in their proximity, resulting in increased in-network traffic.

The fig. 3(b) marks the packet transmission in a tree based structure for 300 data collection rounds. As shown in the figure 3, the proposed PPSDA has significantly high transmissions (up to 7128 packets) as compared to the 5341 packets in SDAP and 5867 packets in SEEDA. A reasonable cause of this high

in-network traffic is the improved data transmission with the nodes remaining alive for a longer duration as compared to the SDAP and SEEDA. The in-network traffic remains constant for about 150 data collection rounds and only increases beyond this point only because with the proposed PPSDA the networks connectivity remains unaffected with increasing number of data collection rounds. However, the nodes die frequently and the networks connectivity is lost with SDAP and SEEDA as the data collection rounds go beyond 150.

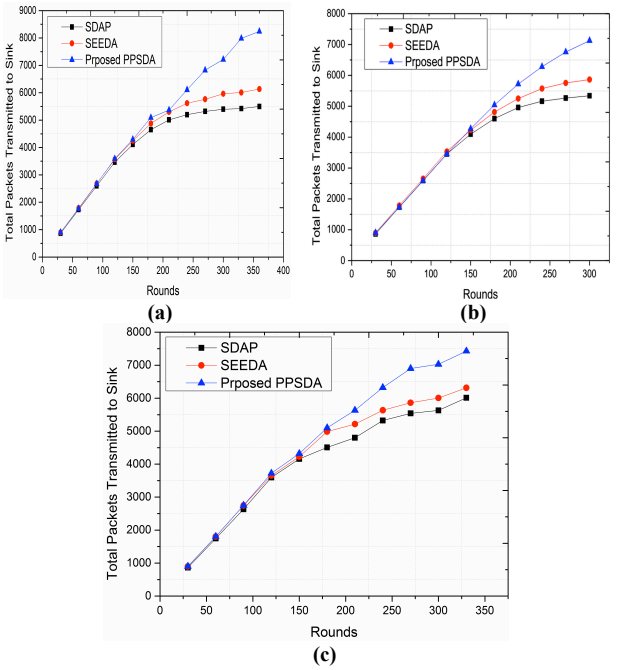


**Fig. 3. Performance Evaluation in terms of In-Network Traffic (a) Clustered Routing (b) Tree Based Routing (c) Chain Based Routing.**

***Accuracy and Efficiency Analysis:*** The data lost in transmission is the termed as the measure of the accuracy and efficiency of the algorithms compared in this section.

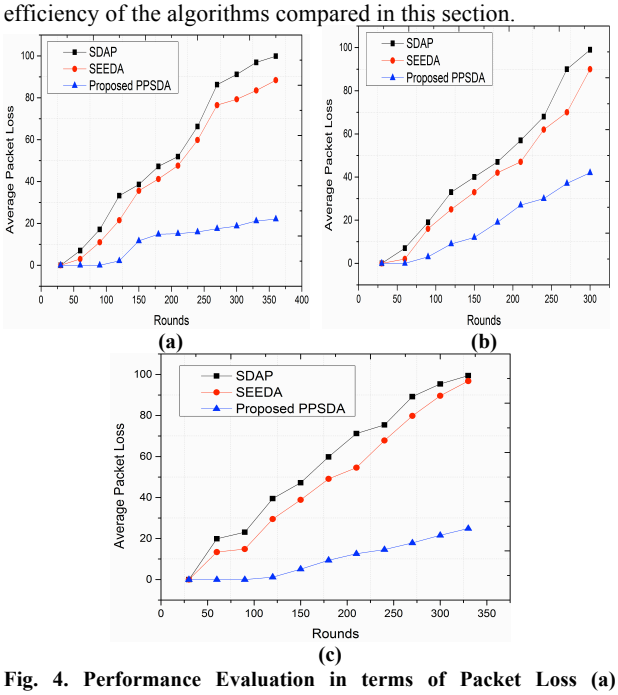


**Fig. 4. Performance Evaluation in terms of Packet Loss (a) Clustered Routing (b) Tree Based Routing (c) Chain Based Routing.**

The fig. 4 shows the average packets lost during transmission as the number of data collection rounds are increased. The accuracy of the proposed PPSDA is maintained best in clustered routing where the average packet loss hardly reaches 22.1 even after 360 data collection rounds. The effect of network lifetime is seen in the connectivity of the network which in turn dictates the average packet loss of the network. As validated from the fig. 4, the average packet loss is minimum with clustered routing structure as compared to the tree based routing (fig. 4(b)) and chain based routing (fig. 4(c)) is the proof of efficacy of the proposed PPSDA in cluster based networks.

***Data Privacy:*** Data privacy in the SDAP scheme, proposed in [11], is maintained at every hop by following the divide-and-conquer and commit-and-attest principles. The analysis of the group aggregates at the sink is used to identify the malicious data which in turn is followed by the attestation process which proves the validity and privacy of the data from the corresponding group. The proposed PPSDA, however, ensures data privacy by allowing the decryption only at the sink. The homomorphic encryption allows addition at intermediate nodes to the encrypted values removing all possibility of malicious data to be injected at any intermediate hop.

### 4.2. Experimental Results

Validated by the simulations, the proposed PPSDA was practically tested for a 23 node network of e-motes for 40 rounds of data collection.

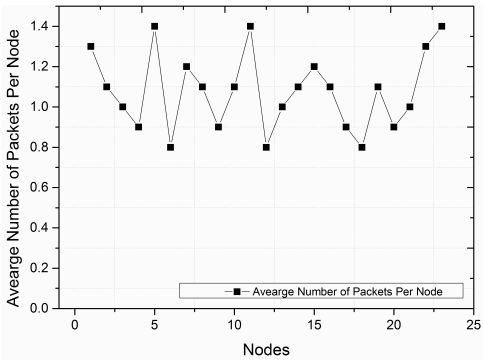


**Fig. 5. Average Packet Transmissions per Node**

The fig. 5 shows the average transmissions for all the deployed nodes when the nodes follow a standard clustering scheme [14] with the proposed PPSDA. As evident from the fig. 5, almost uniform data transmission load is monitored for the network. Clustered routing allows a fault tolerant data transmission resulting in minimal packet loss and thereby improving the overall accuracy of the aggregation process. Practical implementation of the proposed PPSDA proved the efficacy of the proposed PPSDA with clustered routing structure.

## 5. Conclusion

This work aimed at identifying the best suited environment for the PPSDA scheme which targets the end-to-end security in WSNs. The proposed PPSDA scheme is tested for parameters such as network lifetime, in-network traffic and accuracy over three different routing structures. The proposed PPSDA is able to achieve maximum lifetime with a clustered routing structure as compared to the SDAP and SEEDA schemes proposed in the literature. The efficacy of the

proposed PPSDA is also validated over a tree based and chain based routing structure, where the proposed PPSDA outperforms the existing privacy preserving schemes with significant margin. Simulations and practical implementation of the proposed PPSDA prove that the proposed scheme performs optimally in a clustered routing structure as compared to tree based and chain based routing schemes.

## 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. https://doi.org/10.1109/TIFS.2014.2307197

[2] Shaheen, Ahmad, Awadh Gaamel, and Abdulqader Bahaj. "Comparison and analysis study between AODV and DSR routing protocols in vanet with IEEE 802.11 b." J. Ubiquit. Syst. Pervasive Netw 7, no. 1 (2016): 07-12.

[3] Chaudhry, Shafique Ahmad, Weiping Song, Muhammad Habeeb Vulla, and Cormac J. Sreenan. "EMP: A Protocol for IP-Based Wireless Sensor Networks Management." JUSPN 2, no. 1 (2011): 15-22. https://doi.org/10.5383/JUSPN.02.01.002

[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. https://doi.org/10.1109/TDSC.2014.2316816

[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. https://doi.org/10.1016/S1570-8705(03)00008-8

[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. https://doi.org/10.1145/1055558.1055597

[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. https://doi.org/10.1145/1082469.1082470

[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. https://doi.org/10.1109/icc.2005.1494953

[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. https://doi.org/10.1109/MOBIQUITOUS.2005.25

[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. https://doi.org/10.1145/1380564.1380568

[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. https://doi.org/10.1109/wocn.2010.5587353

[13] Singh, Vishal Krishna, Saurabh Verma, and Manish Kumar. "Privacy Preserving In-network Aggregation in Wireless Sensor Networks." Procedia Computer Science 94 (2016): 216-223. https://doi.org/10.1016/j.procs.2016.08.034

[14] Sahoo, Rashmi Ranjan, Moutushi Singh, Abdur Rahaman Sardar, Sharmilla Mohapatra, and Subir Kumar Sarkar. "TREE-CR: Trust based secure and energy efficient clustering in WSN." In Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013 International Conference on, pp. 532-538. IEEE, 2013. https://doi.org/10.1109/ice-ccn.2013.6528557