Turkish Journal of Computer and Mathematics Education Vol.12 No.13(2021), 2217-2224

Research Article

Trust Aware Data Aggregation mechanism for malicious node identification in WSN based IoT Environment

Swathi S.[a], , Yogish H. K.[b]

[a]Research scholar, Department of Computer Science and Engineering, M. S. Ramaiah Institute of Technology, Bengaluru, India
[b]Professor & Head, Department of MCA, M. S. Ramaiah Institute of Technology, Bengaluru, India
Corresponding author. [a]Email:s.swathieswar@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract: As a promising paradigm, the increase in the productivility of Internet of Things (IoT) has contributed largely to the design of modern technology. WSN is an integral part of IoT and founds its application almost in every area of human life such as healthcare, agriculture. Moreover, data collected through these sensors is vulnerable for few application such as health domain , defence domain etc. Hence data collection and analysis is a major challenge. Data Aggregation is considered to be influential and effective mechanism for avoiding the issue of data redundancy and efficient designing of IoT. Despite of such successful implementation and plethora of work in data aggregation, security remains the top priority. Hence, in this research work we design and develop TADA (Trust Aware Data Aggregation) mechanism to provide the efficient and secure environment for data aggregation. In this mechanism, in order to achieve the trade-off between privacy and accuracy, noise are added to data, accuracy parameter and malicious node identification parameter is introduced; further considering these two general constraint is designed and optimization is carried out for malicious node identification and. Furthermore, TADA is evaluated considering the three important parameter i.e. malicious packet identification rate, throughput and packet misclassification rate

Keywords: TADA-mechanism, Secure Data aggregation, malicious node Identification

1. Introduction

Internet of Things (IoT) comprises of smart devices in order to exchange information with each other [1]. The IoT have been developed through multiple intelligent sensory elements and wearable smart devices [2] and play major role in many areas like healthcare, buildings, transportation, mining, agriculture, cities, Industries and automated systems [3]. Sensor networks consist of large number of sensor nodes, which are connected to each other wirelessly. The information obtained by the sensor nodes are transferred to a specific node, called the sinkhole. Lot of work has been done with regard to the construction of sensor networks and the solutions offered to resolve the issues related to the layers and the protocols connecting them some of them being, error tolerance, optimum use of energy resources in sensors, hardware, scalability, environment, energy consumption, low cost, change in network regulations and so on. Being the problems of utmost priority, these issues are to be addressed by the researchers [4]. One such issue is data aggregation in wireless sensor networks [5], wherein the sensor device senses certain physical parameters of a system or region and converts it into electrical signals. These signals are then transmitted to the base station by means of wireless radio. Wireless sensors use small batteries, for which recharging in remote areas becomes tedious. Hence, they have limited energy, unlike the conventional sensors. Thus, the amounts of data that can be sent are reduced by using data aggregation techniques, which in turn increases the lifetime of the network.

Usually, the sensors in the wireless sensor are scattered randomly in spaces, with less human intervention. Using the multi-hop system architecture, the sensors collect the data and send them to the base station or the sink. The main concern in such situations is the conservation of energy and hence proper management of packets. This can be achieved either by merging data through middle sensors or data compression [6]. This leads to effective energy utilization of the sensors, thereby increasing the lifetime of the network and increased bandwidth. Thus, data aggregation (collecting the data from multi-sensors) is considered as effective technique for combining data. Another advantage of using data aggregation is that, the data are delivered in a very efficient manner with minimal data latency. Different data aggregation algorithms to improve the lifetime of sensor network exist.

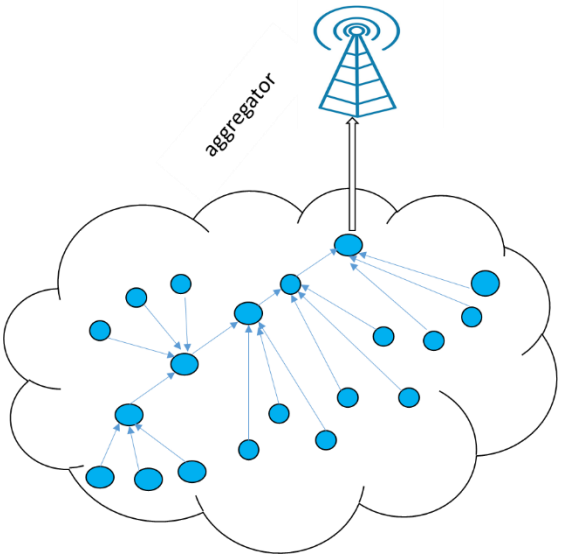

Figure 1. Data aggregation mechanism

Enormous amount of heterogeneous data involved in IoT like wireless sensor network, results in consumption of more energy while sending data from different sources [7]. Figure 1 shows the data aggregation mechanism. Energy consumption in such cases can be reduced by processing and aggregating the summarized data prior to sending it. However, the major concern in data aggregation is the heterogeneity of data in the network. The data may be either an image or audio or sensory. Hence, there is a need for modelling and compressing the data. However, WSNs are prone to attacks like replay attack, tempering attack, injection attack etc, since they usually exist in hostile and unattended environments. Since WSNs are resource-constrained, existing abundant security algorithms are not found suitable for them. Hence, there is a strong need for ensuring security for data aggregation [8]-[10]. Lot of work has been done in literature to address this issue of data security and privacy in IoT [11]-[14]. Most of them deal with ensuring the data protection in IoT sensors, through various cryptographic solutions, such as homomorphic encryption, wherein the data collected through various wireless sensor devices are aggregated and do not require decryption. Homomorphic properties of partially homomorphic public key systems, like BGN cryptosystem were initially improved [15] by performing addition and multiplication on two cipher texts. Some researchers have considered fully homomorphic cryptosystems using the assumptions of learning with errors (LWE) [16]. Homomorphic data aggregation scheme (by preserving privacy) for wearable devices in WSN cloud computing [17] was proposed by using fully homomorphic encryption, secure multi-party computation and one-way trapdoor function. Privacy of patients in WSN of IoT devices in the implementation was ensured [18] by employing [19] and [20] cryptosystems. For ensuring confidentiality and efficient authentication in healthcare systems and IoT, confidentiality and authenticity of information need to be done. However, in lightweight IoT devices, this results in increase in communication overhead and also the computational complexity. The reduction in the computational overhead of encryption and digital signature can be achieved using non-homomorphic signcryption algorithm [21]. Many such signcryption schemes have been proposed since then. To name a few. Identity-Based Signcryption [22], Certificate less [23] etc. Further, both attribute and identity based signcryption context-aware access control and authentication [24] have also been suggested to achieve adaptive contextual privacy and security in WBAN. Certificate less Sign cryption have also been proposed in dealing with the problem of key escrow and to avoid impersonation attack, by ensuring generation of only partial private key, by the key generation centre (KGC).

2. Motivation and contribution of research work

The Internet of Things (IoT) has a major role in the modern era for use in different applications. The applications comprise generating, exchanging, aggregating and analyzing a large amount of security-critical and privacy-sensitive data, which make them prone to various attacks. Thus, there is a need for IoT devices to protect themselves from risks associated with their security and privacy, while meeting the functional requirements and services. Hence, IoT devices need to implement privacy, at the same time preserving data manipulation. In this regard, it is intended to process privacy-sensitive data for extracting information that is expected, without any disclosure of privacy. Further, many factors related to privacy, like risk assessment and privacy valuation have an impact on private and sensitive data trading between owners of the data and those who request the data. The security behaviour of owners of the data also has a key role in protecting the privacy in IoT applications.

Motivated by all the above phenomena, this research work proposes TADA-mechanism for secure data aggregation; further contribution of research work is given as:

1. In this research work, we design and develop TADA(Trust Aware Data Aggregation)-mechanism to identify the malicious packet in WSN .
2. In TADA-mechanism At first we add noise to the original data, where noise are designed through random initialization.
3. Further DA-accuracy and malicious packet identification parameter are computed and general constraint is designed considering these parameter.
4. Further, this is optimized and trade-off among these two are achieved.
5. TADA-mechanism is evaluated considering the three metrics i.e. malicious packet identification, packet misclassification and throughput.

This research is carried out as follows: First section starts with background IoT, WSN based IoT and its application; further in the same section we discuss the importance of data aggregation and security concern is discussed. Further in same section few existing mechanism is reviewed followed by that motivation and contribution of research work is highlighted. In second section TADA-framework is designed along with mathematical formulation and design. Third section evaluates and performs comparative analysis of existing and TADA-mechanism.

3. Proposed Methodology

In this section of the research, we design and develop proposed methodology for secure data aggregation; Hence TADA mechanism aka Trust Aware Data Aggregation mechanism is designed for gaining the trust through absolute identification of malicious nodes. Identification of malicious modes has been focused by several research work, however they identifier the incorrect as the malicious node and considering the importance of privacy if the genuine node are identified as the malicious node then by default it would be discarded and further it leads to more cost, energy and model is considered as inefficient.

Figure 2 presents the TADA-mechanism workflow which comprises the five blocks; first block represents the adding original data with noise, second block is computation of aggregation accuracy parameter which shows the how well data is aggregated further malicious node identification computation block formulated the parameter of malicious node to be identified. Further in next and fourth block generic constraint is designed with respect to second and third block. At last fifth block indicated the optimization model of designed objective.

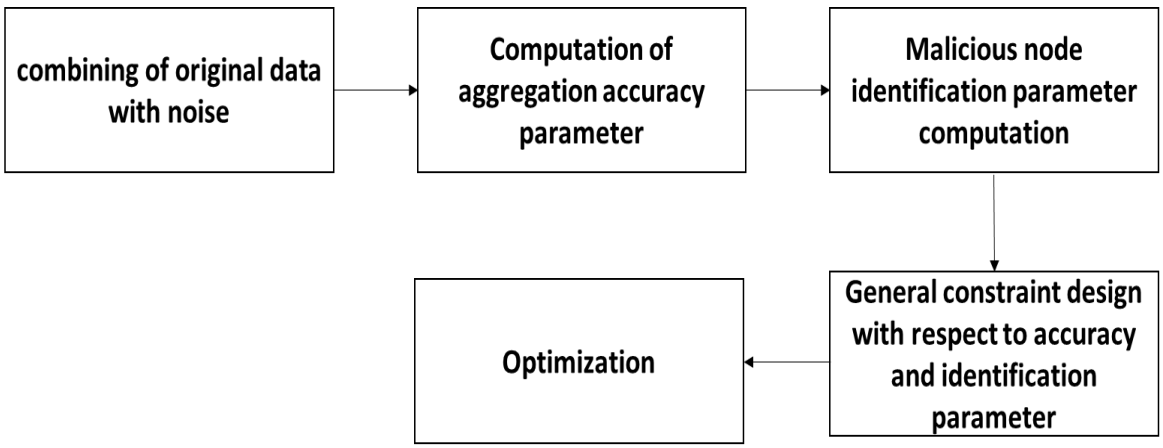

Figure 2. TADA mechanism workflow

System model:

Let's consider a model which comprises a number of users i.e. $U = \{u_1, u_2, \ldots, u_m\}$ and DIC (Data information Centre); all m have data as $Y = \{y_1, y_2, \ldots, y_m\}$ to DIC. In here, each node are rated as malicious node, normal node or the honest node with respect to $u_j$defined reputation; DIC computes the average as given

$$z = O^{-1} \sum_{j=1}^{O} y_j \quad (1)$$

Further, we consider the untrusted DIC(Data Information Center) and two type of nodes are considered named honest node and malicious nodes and considers two distinctive threats, first threat is aggregated data quality and second one is comprising the privacy.

3.1. Initialization

Since privacy is primary concern each node uploads the unsettled data and not raw one. Further we perform parameter initialization

Each member w_j adds the data y_j along with initialised noise δ_j and formulated as:

$$y_j' = y_j + \delta_j \quad (1)$$

In above equation, δ_j is nearly equivalent to $O(0, \tau^2)$ and noise consideration is based on designed random function $N(.): s \text{ to } S$ to given data y_j which can be indicated as y_j' above equation can be written as:

$$z_j'^{l} = \sum_{j=1}^{o'} x_j^l y_j^l \quad (2)$$

In above equation, O' indicates hones number of nodes among the uploaded data.

As we adopt the random noise perturbation, it develop another issue which is getting the trade-off among privacy and accuracy of data aggregation, thus to achieve that we introduce a accuracy parameter denoted as μ.Accuracy parameter is computed through below equation

$$\mu = [z - z_j'] \quad (3)$$

the above equation clearly indicates that smaller value of parameter will provide the better accuracy i.e. for each user w_j if the defined weight which is x_j with given distribution ρ_j then the accuracy parameter is computed as below;

3.2. Parameter Computation and general constraint design

Lets consider any aggregated data without any additional noise which is given as below.

$$z = \sum_{j=1}^{o'} x_j y_j \quad (4)$$

And the unsettled data are given as

$$z' = z + \sum_{j=1}^{o'} x_j \delta_j \quad (5)$$

Further above equation can be optimized and given as:

$$\aleph(\delta_j) = \sum_{j=1}^{o'} (x_j)^2 \varphi_j^2 \quad (6)$$

Since $\aleph(\delta_j) = 2\varphi_j^2$ and for all δ greater than zero, equation can be further formulated as:

$$\frac{2}{\delta^2} \sum_{j=1}^{o'} (x_j)^2 \varphi_j^2 = 1 - q \quad (7)$$

And accuracy parameter is computed as:

$$\delta = \omega (1-q)^{-1/2} \left(\sum_{j=1}^{o'} (x_j)^2 \varphi_j^2 \right)^{-1/2} \quad (8)$$

On the other side adding the noise provides more security and other factors like cost and rational constraint have fair impact on the accuracy and privacy, also the identification rate of malicious node plays major role in trust aware Data Aggregation Hence the identification can be given as:

$$\text{identification}_j = \tau_j \varphi_j^2 \quad (9)$$

General constraint can be defined as the relation among the total cost and the security parameter φ_j and the constraint is given through the below equations:

$$\text{total}_{\text{idintifation}_j} - \tau_j(\varphi_j)^2 \text{ is greater than or equavelent to } 0 \quad (10)$$

In above condition for all j exist in $[1, O]$.

3.3. Optimization

Further, as observed earlier in this section, that accuracy parameter δ should be minimum, hence additional mechanism is designed to minimize with respect to malicious packet and general constraint ; objective can be formulated as:

$$\min_{\varphi_j} \text{identification}_j \sum_{j=1}^{O'} \frac{(y_j^l)^2}{\varphi_j} \quad (11)$$

Considering the above problem solution can be given through two distinctive equation i.e. optimal identification and optimal accuracy which is given as

$$\hat{\varphi_j} = \left(\frac{(x_j^l)^2 (\tau_j)^{2/3}}{\sum_{j=1}^{O'} (x_j^l)^2 (\tau_j)^{1/3}} \right)^{-1/2} C \quad (12)$$

Optimal identification is given as:

$$\hat{\text{total\_identification}_j} = \left(\frac{(x_j^l)^2 (\tau_j)^{1/3}}{\sum_{j=1}^{O'} (x_j^l)^2 (\tau_j)^{1/2}} \right)^{-1/2} C \quad (13)$$

Moreover these both equation i.e. equation 12 and equation 13 compliment each other despite of mentioned in different equation. Moreover the main intention is to achieve the absolute identification malicious node such that trust can be guaranteed.

4. Performance Evaluation

Herein, it is intended to evaluate the proposed methodology considering the different metrics such as malicious packet identification rate, packet misclassification rate and throughput. Further details is given through below table

parameter	Value
Initial energy	5mj
Total number of sensor nodes	100
Area of the network	50m *50m
Induced malicious nodes	10, 20, 30 and 40

Moreover, to achieve that sensoria simulator is used in system configuration of 1 TB SSD hard disk packed with 4GB Nvidia graphics along with 16GB RAM. VS17(Visual Studio 2017) is used as IDE and c sharp is used as the programming language.

4.1. Malicious Packet Identification

Figure 3 shows the comparison of malicious packet identification between the existing and proposed mechanism; herein, x-axis represents the malicious nodes induced, while y axis represents the malicious packet; moreover higher packet identification rate indicates the efficiency of the model. In here for 10 malicious nodes, existing mechanism identifies the 62 malicious nodes whereas TADA-mechanism identifies 96 malicious packet. Similarly for 20 and 30 malicious nodes, existing mechanism identifies 58 and 46 whereas TADA mechanism identifies 94 and 93 respectively. At last 40 malicious nodes were induced where existing mechanism identifies 62 malicious packet and TADA identifies 96 malicious packet.

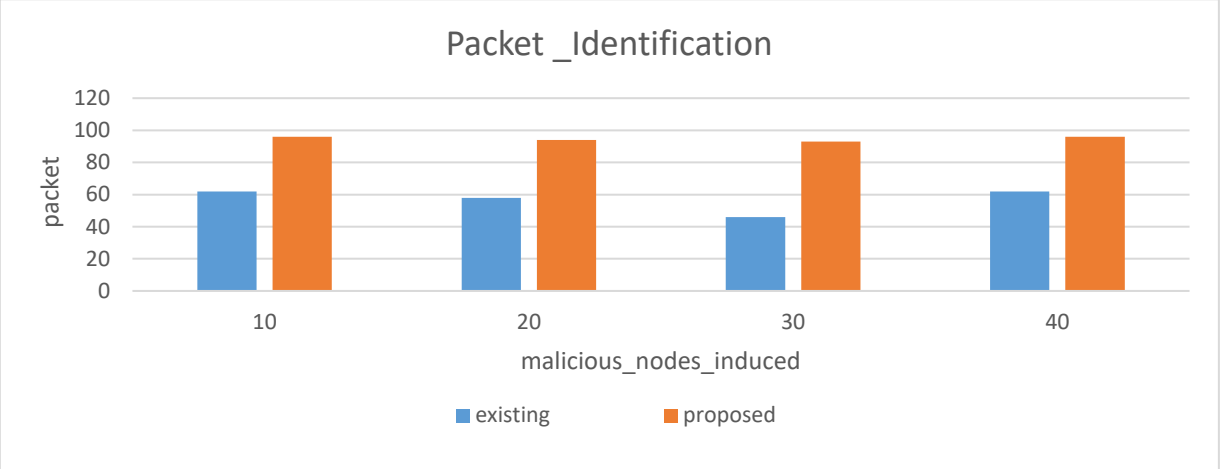

Figure 3.

4.2. Packet Misclassification

Figure 4 shows the comparison of malicious packet misclassification between the existing and proposed mechanism; herein, x-axis represents the malicious nodes induced, while y axis represents the misclassified packet; moreover lower packet misclassification rate indicates the efficiency of the model. In here for 10 malicious nodes, existing mechanism identifies the 40 malicious nodes whereas TADA-mechanism misclassifies 4 malicious packet. Similarly for 20 and 30 malicious nodes, existing mechanism misclassifies 42 and 50 whereas TADA mechanism misclassifies 6 and 7 respectively. At last 40 malicious nodes were induced where existing mechanism misclassifies 38 malicious packet and TADA identifies 4 malicious packet

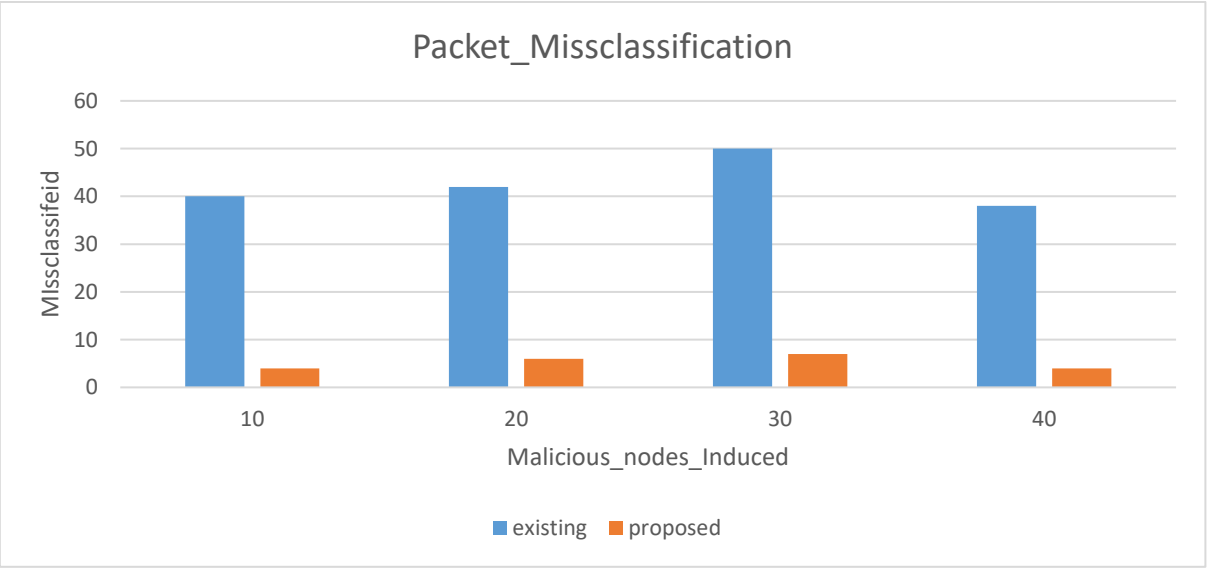

Figure 4.

4.3. Throughput

Throughput is defined as the rate at which work is being done in efficient manner; higher throughput indicates the higher efficiency of model.

Figure 3 shows the comparison of throughput between the existing and proposed mechanism; here, x-axis represents the malicious nodes induced and y axis represents the throughput; moreover higher higher throughput indicates the efficiency of the model. In here for 10 malicious nodes, existing mechanism achieves throughput of 0.4588 here as TADA-mechanism achieves throughput value of 0.7104. Similarly for 20 and 30 malicious nodes, existing mechanism achieves throughput of 0.3199 and 0.1866 whereas TADA mechanism achieves throughput of 0.5177and 0.3813 respectively. At last 40 malicious nodes were induced where existing mechanism achieves throughput of 0.1364 malicious packet and TADA-mechanism achieves throughput of 0.2112.

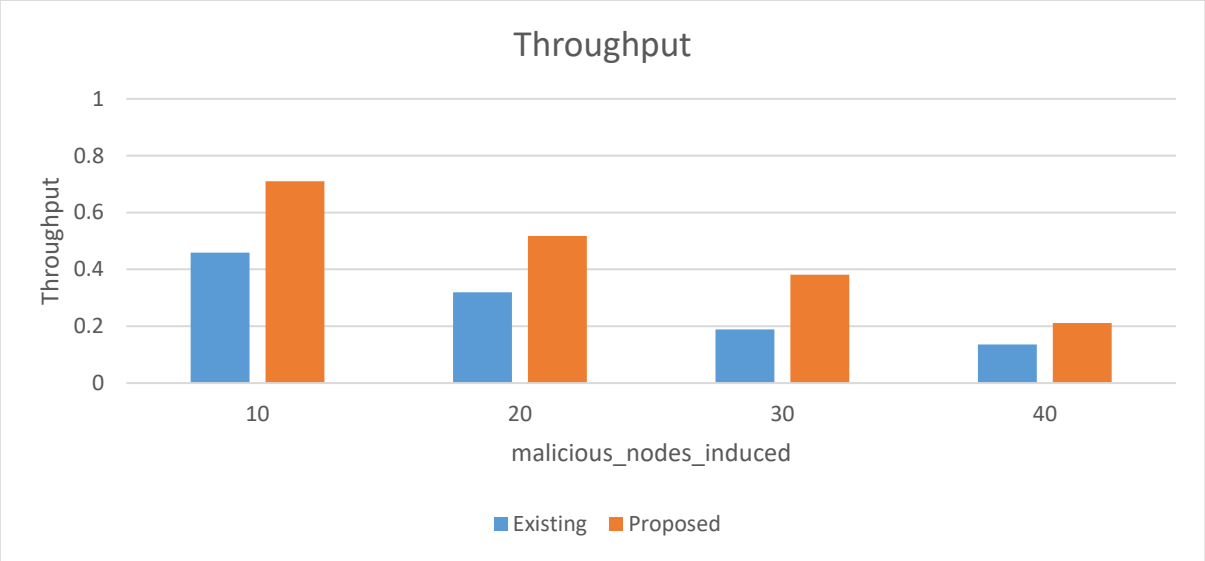

Figure 5.

4.4. Comparative Analysis

The comparative analysis of existing and proposed mechanism has been carried out herein. At first we observed that in terms of packet identification rate, proposed methodology achieves 54.83%, 62.06, double, and 54.83 % improvisation in comparison with existing mechanism. Further, in terms of misclassified packet identification rate TADA improvises 90%, 85.75%, 86% and 89.47% improvisation than existing methodology. Further in terms of throughput TADA-mechanism achieves 54.83%, 62.069%, double and 54.83% better throughput than the existing methodology.

5. Conclusion

In this research work, we design and develop a secure data aggregation mechanism named TADA (Trust aware Data Aggregation)-mechanism for secure and efficient data aggregation; TADA mechanism designs the general constraint with respect to the data aggregation accuracy and identification rate; further considering these two general constraint were designed and optimized for efficient modelling. Furthermore, TADA-mechanism is evaluated considering the three different mechanism malicious packet identification which identifies the number of malicious packet correctly identified and malicious packet misclassified i.e. number of malicious packet incorrectly identified. Further, throughput is considered as another evaluation metrics, also comparative analysis with existing model is carried out for all three metrics and TADA mechanism simply outperform the existing methodology. Although TADA-mechanism outperforms the existing model, security in data aggregation is wide research area and there are several gaps which need to be still fixed such as data reliability which will be carried out in future and consideration of energy as evaluation parameter.

References

1. G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, “A new architecture of internet of things and big data ecosystem for secured smart healthcare monitoring and alerting system,” Future Generation Comp. Syst., vol. 82, pp. 375–387, 2018.
2. R. Casadei, G. Fortino, D. Pianini, W. Russo, C. Savaglio, and M. Viroli, “Modelling and simulation of opportunistic IoT services with aggregate computing”, Future Generation Computer Systems, vol. 91, pp. 252-262, 2019.

3. Z. Xu, L. Chen, C. Chen, and X. Guan, "Joint clustering and routing design for reliable and efficient data collection in large-scale wireless data collection position networks", IEEE Internet of Things Journal, vol. 3, no. 4, pp. 520-532, 2016.
4. K. Das, S. Zeadally, and D. He, ''Taxonomy and analysis of security protocols for Internet of Things,'' Future Gener. Comput. Syst., vol. 89, pp. 110–125, Dec. 2018.
5. Pourghebleh and N. J. Navimipour, ''Data aggregation mechanisms in the Internet of Things: A systematic review of the literature and recommendations for future research,'' J. Netw. Comput. Appl., vol. 97, pp. 23–34, Nov. 2017.
6. M. Aazam, S. Zeadally, and K. A. Harras, ''Fog computing architecture,evaluation, and future research directions,'' IEEE Commun. Mag., vol. 56, no. 5, pp. 46–52, May 2018.
7. Rezazadeh, J., Moradi, M., Sandrasegaran, K., & Farahbakhsh, R. (2018). Transmission power adjustment scheme for mobile beacon-assisted sensor localization. IEEE Transactions on Industrial Informatics, 15, 2859–2869.
8. Lashkari, B., Rezazadeh, J., Farahbakhsh, R., & Sandrasegaran, K. (2019). Crowdsourcing and sensing for indoor localization in IoT: A review. IEEE Sensors Journal, 19(7), 2408–2434.
9. X. Wang, J. He, P. Cheng, and J. Chen, "Privacy Preserving Collaborative Computing: Heterogeneous Privacy Guarantee and Efficient Incentive Mechanism," IEEE Transactions on Signal Processing, vol. 67, no. 1, pp. 221-233, 2019.
10. M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, "Privacy-Preserving Support Vector Machine Training over Blockchain-Based Encrypted IoT Data in Smart Cities," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 7702-7712, 2019.
11. F. Rezaeibagha, Y. Mu, X. Huang, W. Yang, and K. Huang, "Fully secure lightweight certificateless signature scheme for IIoT," IEEE Access, vol. 7, pp. 144 433–144 443, 2019.
12. J. Hong, B. Liu, Q. Sun, and F. Li, "A combined public-key scheme in the case of attribute-based for wireless body area networks," Wireless Networks, vol. 25, no. 2, pp. 845–859, 2019.
13. W. Tang, J. Ren, K. Deng, and Y. Zhang, "Secure data aggregation of lightweight e-healthcare iot devices with fair incentives," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8714–8726, 2019.
14. Z. Wang, "Blind batch encryption-based protocol for secure and privacy-preserving medical services in smart connected health," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9555–9562, 2019.
15. D. Boneh, E. Goh, and K. Nissim, "Evaluating 2-dnf formulas on ciphertexts," in Theory of Cryptography, Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005, Proceedings, 2005, pp. 325–341.
16. C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009, 2009, pp. 169–178.
17. J. Zhou, Z. Cao, X. Dong, and X. Lin, "Security and privacy in cloud-assisted wireless wearable communications: Challenges, solutions, and future directions," IEEE Wireless Commun., vol. 22, no. 2, pp. 136–144, 2015.
18. X. Yi, A. Bouguettaya, D. Georgakopoulos, A. Song, and J. Willemson, "Privacy protection for wireless medical sensor data," IEEE Trans. Dependable Sec. Comput., vol. 13, no. 3, pp. 369–380, 2016.
19. P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Advances in Cryptology - EUROCRYPT '99, May 2-6, 1999, Proceeding, ser. Lecture Notes in Computer Science, vol. 1592. Springer, 1999, pp. 223–238.
20. T. El Gamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," in Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings, ser. Lecture Notes in Computer Science, vol. 196. Springer, 1985, pp. 10–18.
21. Y. Zheng, "Digital signcryption or how to achieve cost(signature & encryption) << cost(signature) + cost(encryption)," in Advances in Cryptology - CRYPTO '97, 17th Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 1997, Proceedings, ser. Lecture Notes in Computer Science, vol. 1294. Springer, 1997, pp. 165–179.
22. J. Malone-Lee, "Identity-based signcryption," IACR Cryptology ePrint Archive, vol. 2002, p. 98, 2002.
23. M. Barbosa and P. Farshim, "Certificateless signcryption," IACR Cryptology ePrint Archive, vol. 2008, p. 143, 2008.
24. Arfaoui, O. R. M. Boudia, A. Kribeche, S. Senouci, and M. Hamdi, "Context-aware access control and anonymous authentication in WBAN," Computers & Security, vol. 88, 2020..