Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777346

A NOVEL PRIVACY AND RELIABLE AGGREGATION FRAMEWORK IN WSN BASED IOT ENVIRONMENT

Nandini S[1], Dr. Kempanna M[2]
[1]Dept. of Computer Science & Engineering, Research Scholar, BIT, Bangalore, India
[2]Dept. of Computer Science & Engineering, Associate Professor, BIT, Bangalore, India

ABSTRACT
The widespread evolution of IoT over the last few years has resulted in the adoption of a crucial infrastructure. Therefore, many benefits have been extracted using a IoT structure that is short-termed because of the exponential increase in related security as well as threats associated with privacy. Associated adversaries in carrying out the attacks relating to privacy for accessing confidential as well as sensitive data for crucial infrastructure for different commercial, political, and self-centred gains. This research work presents PR-Aggregation (Privacy and Reliable Aware) framework for securing the environment. The proposed framework induces variables based on the designed constraint for privacy, reliability, and verifiability. PR-Aggregation framework is evaluated considering the number of deceptive nodes for securing the node in the network.
Keywords: WSN, Security ,Aggregation, IoT, Nodes.

INTRODUCTION
Presently, WSN (wireless sensory networks) are utilized at the forefront of communication systems that include environmental monitoring, smart home automation, industries, etc. This technology has a widely increasing potential concerning applications of real-time due to its small size, inexpensive and easy deployment [1]-[2]. The design of WSNs is dependent on various situations for their application. In applications of the industry as well as automation of the home, in which there is no constraint of energy, delivery packets are the most essential for designing a network. However, in an unsafe environment in which there is a battery that is not replaced or recharged which includes mining, which increased lifetime of the network, is the main factor in designing the network. Although, the deployed area size has an important role in the design of the WSN. Considering smaller areas, the transmissions of the packets is done directly from the sensory nodes to the base station or sink nodes, whereas, in large areas, the transmission of packets happens via various intermediate nodes to sink nodes [3].

Additionally, the system that is enabled using IoT has an ability for interconnection of various 'things' for effective communication as well as sharing of data for one network, there is a wide number of advantages that have attracted various technologists [4] [5]. In the short term, IoT systems are an important part of various sectors that include healthcare, manufacturing, logistics as well as transportation, which enable crucial infrastructure of IoT [6]. Considering the increased complexity in architecture, there are several uses of heterogeneous devices and threats towards privacy that are difficult to be identified, access and mitigated. However, the large-scale increased complexity of IoT systems introduces a deluge of data. Confidential and sensitive data have been constantly shared among the networks, and

privacy, as well as security, are huge concerns that are prevailing in the IoT's crucial infrastructure [7-9]. Any of the attacks on cyber systems that are vulnerable could compromise the integrity and privacy of huge amounts of data that is sensitive.

Approaches towards the preservation of privacy have gathered attention over the years due to progress towards information technology that has threatened individual privacy. Although, considering the rise towards the adoption of edge computing as well as fog paradigms for IoT crucial information has led to decreased latency, awareness of location, communication and data sharing in real-time and QoS [10] [11]. The devices used for edge computing in IoT crucial information are prone to attacks of privacy [12]. Considering this approach, many techniques for the preservation of privacy that include traditional as well as modern deep learning algorithms are proposed [13]. The existing deep learning methodologies that are implemented based on fog/edge computing for crucial information solutions need expensive computation.

In [14], a methodology of truth discovery for preserving privacy is proposed, although, the client overhead is high. In [15-18], a two-layer methodology is proposed for fulfilling the requirement of protecting the privacy of the user. In [19], a novel lightweight framework of truth discovery for preserving privacy is proposed that is used for the implementation of two cloud platforms that are non-colluding as well as adopting a homomorphic cryptosystem. In [20], a novel data poisoning disguise attack (DDPA) is proposed against private systems of crowd sensing that are empowered with the methodology of truth discovery. A stealth strategy is proposed that is, the malicious characteristics are disguised for avoiding the detection of methods for truth discovery [21-24]. Along with this strategy of stealth, the limitation of maximizing the effectiveness of the attack is naturally avoided using optimization problems at the bi-level via structuring that is resolved by an alternate optimization algorithm. In [25], a novel semantic awareness for the preservation of privacy for the trajectory of online location sharing method called Semantic-awareness information-theoretical Privacy (SEITP), for the protection of privacy of data as well as semantic during semantic awareness of data is utilized can be preserved completely.

There are various kinds of attacks on IoT crucial information that include attacks for Denial of Service, Sybil attacks, etc [8-9]. Access level-based cyber-attacks for IoT crucial networks are classified as passive and active attacks [10]. Active attacks, which are also termed attacks that are security oriented, derange the communication of the network by avoiding the available security protection. On the contrary, passive attacks are termed as attacks that are privacy-oriented such as networks with eavesdropping without any disturbance for gaining illicit access towards confidential sensitive information. The widely growing IoT crucial information is now prone to various attacks by hackers as well as organized crime syndicates. The growth in the count of threats towards privacy that are targeted towards IoT crucial information results in motivation for the development of various solutions. Although, most of the security approaches that are proposed have the absence of applicability, which may be the result of the complexity of computation, expense and other factors that are related. Motivated by the above issues, the contribution of the research is given as follows:

This research takes Privacy, security and reliability into account for aggregation in WSN based IoT environment and develops a PR-Aggregation framework.

PR-Aggregation framework is an integrated framework, which utilizes the designed constraint, time and event for preserving the nodes privacy.

PR-Aggregation is evaluated by inducing the dishonest nodes in the network for detection of dishonest nodes for classification and misclassification of nodes identification.
This research is organized as follows: First section start with a background of IoT, sensor nodes and issues of security and privacy of the nodes and the section concludes with research motivation and contribution. The second section presents the mathematical modelling of the PR-Aggregation model and the third section presents the performance evaluation.

PROPOSED METHODOLOGY

The data that has been gathered from IoT devices must be required for analysis as well as utilization. However, the gathered data could also have information that is sensitive and personal to the user, this hinders the privacy protection of the users. The unwillingness of the users to contribute their information affects the usability of the data. Privacy and reliability while aggregation has been one of the wide research using the consensus-based approach where nodes are required to exchange and disclose their state information to its adjacent node. However, this directly involves the privacy and reliability of nodes. This research work aims to develop a PR-aggregation mechanism to achieve the trade-off among privacy and aims to guarantee convergence. Moreover, secure data aggregation has been carried out in previous research work; this work utilizes a secured and efficient mechanism and further provide privacy and reliability of the user. Figure 1 presents the proposed PR-model that comprises various modules. At first, system model is designed based on the connected graph; further total number of variable is selected along with selection of adjacent nodes. Furthermore, considering the variable, nodes states are updated and constraints are checked and verified.

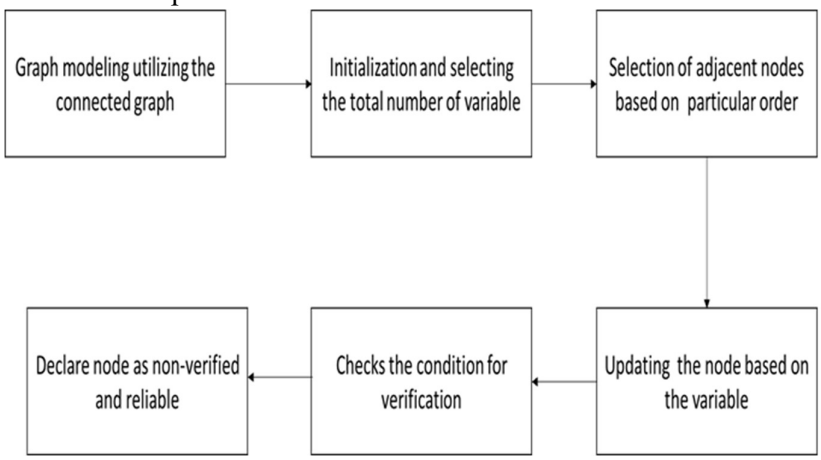

Figure 1 Proposed PR-model

System modelling and problem definition

Let us consider any particular network $Q(Q \geq 2)$ that communicates with adjacent nodes; also communication is established through the designed directed graph. In a designed graph,$J_g = (Y, H)$, where $Y = [y_1, y_2, y_3, \ldots\ldots y_q]$ and q is the particular node set with G as the edges. Furthermore, we consider particular time step $m \in C \geq 0$ with each node maintaining their particular state. Consider the particular strong connected graph as designed earlier, each node

has curtain stage denoted as $b_l[0]$, thus nodes are required to compute the following equation that satisfy the verifiable and reliable criteria.

$$B' = q(\sum_{o-1}^{q} b_o[o])^{-1} \quad (1)$$

The main aim of this research work is to design a particular strategy for verifiable and reliable strategy for nodes $y_k \in Y_s/Y_f$ while communicating with the nodes. Moreover, the problem is designed to compute the B' as in equation (1),

1.1 Designing constraints for privacy, reliability, and verifiability

In this section, we develop a constraint, which aims preserving the nodes information, and any malicious node, which aims to obtain information, is removed from the network.

Condition 1: Adding steps O_m in the variable of each node y_m required to be larger than or equal to y_m's degree such that the adjacent node receives at least one piece of information.

$$O_m \geq G_m'' \quad (1)$$

Condition 2: Moreover, the accumulated variable is infused in computation through the node $y_m$such that the node state can be computed without any error.

$$x_m = -\sum_{o_m}^{O_m} x_m[o_m] \quad (2)$$

Constraint 3: Variable $x_m[o_m]$ is induced to the network by each node based on the event, which needs to be non-negative.

$$x_m[o_m] \geq 0, forallo_m \epsilon [0, O_m] \quad (3)$$

Constraint 4: Node x_m stops inducing variables such that states can be calculated as follows:

$$x_m[o_m] = 0, forallo_m \notin [0, O_m] \quad (4)$$

1.2 PR-aggregation algorithm

Input is taken as the connected graph $J_g = (Y, H)$ with $q = \|H\|$ edges along with the initial state of $B_l\ [0] \in C$	
Step1:	Assigning a particular order R_{o_l} in a given set {0,1, ... G_m''} for each adjacent node $y_o \in Q_m''$
Step2:	Setting up the counter $f_m = 0$ along with index(priority based) h_m to f_m
Step3:	Setting up the counter $o_m = 0$, selects $O_m \in C > 0$ where $O_m \geq G_m''$
Step4:	Setting up $B_m' = B_m[0] + x_m$, $C_m[0] = 1, C_m^v[0] = 1$ &$C_m^v[0] = B_m'[0]$

Step5:	Choosing adjacent node $y_o \in Q_m''$ such that $R_{om} = h_m$ and transmit $C_m[0]$ and $C_m' + x_m[0]$ to a particular adjacent node. Furthermore, setting $B_m'[0] = 0$, $C_m'[0] = 0$, $n_m = n_m + 1$
Step 6:	Setting $f_m = f_m + 1$ and $h_m = f_m mod O_m''$
Step7:	Considering the iteration of $n = 0,1,2 \ldots . everynode y_m$ carries out following operation
Step8:	If it receives $B_l'[N], C_l[N]$ from adjacent node $y_m \in Q_m''$ and updation is carried out with equation 1 and equation 2
Step9:	If equation 1 and equation 2 hold then transmit the information about nodes for preserving privacy
Step10:	Output as $t_m^v[n]$ for each node $y_m \in Y$

In the above algorithm, the first o_m to zero and select the total number of variables adding with O_m steps such that $O_m \geq G_m''$ and with $(O_m + 1)$ positive variable$x_m[o_m] > 0$. Furthermore, considering $y_o \in Q_m'$ in accordance with order S_{om} and transmit $B_m'[0]$ and $B_m'[0]$. Furthermore, while execution of algorithm, at every step m, node $y_m$receive set of requests for packet transmission from each adjacent node, conditions are checked for verification and if holds then the data packets are transmitted and node remains else nodes are discarded out of the network.

PERFORMANCE EVALUATION

While performing the aggregation it is very important, that reliable and truthful data are being collected. PR-Aggregation aims to assure the privacy of sensor nodes as well the reliability of the data. Moreover, PR-Aggregation is evaluated considering the dishonest sensor nodes. It is evaluated on the system configuration that includes 2 TB of the hard disk loaded with 16 GB of RAM along with 2GB NVidia Cuda-enabled graphics. The proposed model here works on evaluating an incorrect identification of the node that results in unevenness in the network by accommodating various parameters like Identification of the correct node, identification of the wrong node and computing the throughput for 30, 40 and 50 nodes. Additionally, a comparative analysis is done between the proposed model with the existing model to ensure the model's security and efficiency and conclude that the proposed system performs better than the existing system.

Correct node Identification

In this section, the correct nodes are identified wherein a comparison is made between the existing system and the proposed system by evaluating the correct identification of nodes with 30, 40 and 50 nodes. Figure 2 shows the comparison of the stated above; in the context of 30 compromised nodes, the existing system detects 41 sensor nodes and the proposed model identifies 97 nodes. Consequently, in the context of 40 sensor nodes, the existing system identifies 75 nodes whereas the proposed model identifies 99 nodes. For 50 sensor nodes, the existing system identifies 66 nodes whereas the proposed model identifies 97 nodes.

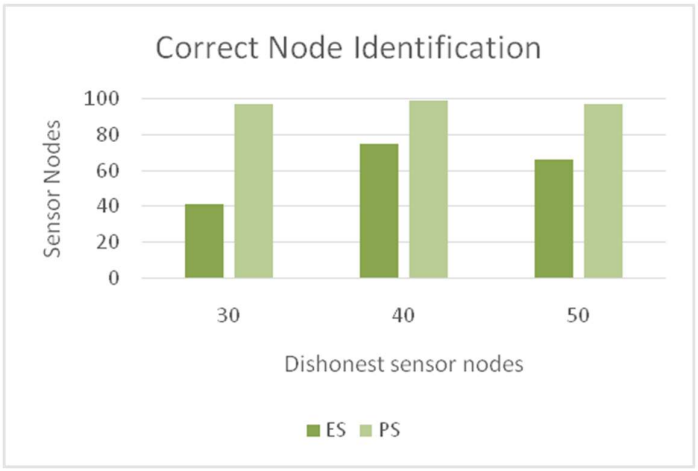

Figure 2 Correct Node Identification

Wrong Node Identification

Figure 3 depicts the wrong identification of nodes for 30, 40 and 50 sensor nodes. In 30 nodes context, the existing model identifies 34 wrong nodes whereas the proposed model wrongly identifies 4 nodes. In 40 nodes the existing model wrongly identifies 36 nodes whereas the proposed model wrongly identifies 2 nodes. In 50 nodes context, the existing model wrongly identifies 35 wrong nodes whereas the proposed model wrongly identifies 4 nodes.

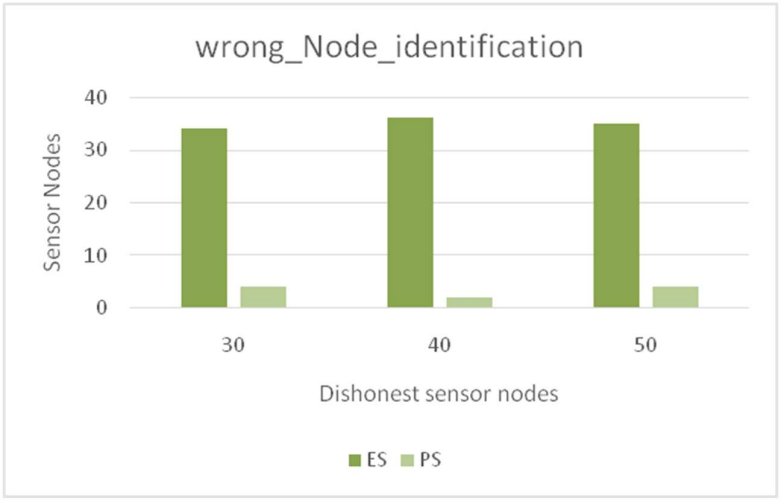

Figure 3 wrong node identification

Throughput

Throughput is defined as the amount of work done in a specific amount of time; it showcases the models' efficiency. In the case of 30 compromised nodes, the throughput of the existing model is 0.123 and for the proposed model, it is 0.9933. In the case of 40 compromised nodes, the throughput of the existing model is 0.3 and for the proposed model, it is 0.891. In the case of 50 compromised nodes, the throughput of the existing model is 0.33 and for the proposed model, it is 0.693 as shown in figure 4.

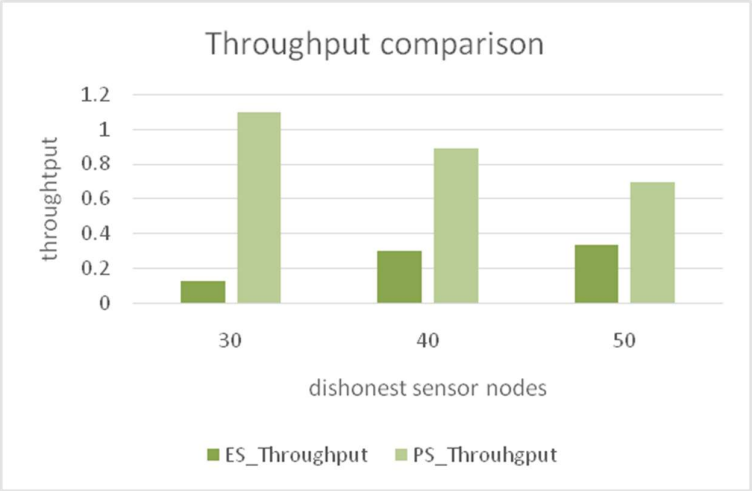

Figure 4 Throughput comparison

Comparative Analysis

This section displays the comparative analysis and shows the percentage improvisation for the proposed model from the existing model. The improvisation is carried out on correct node identification for 30 nodes is 8.159 % improvisation for 50 nodes the improvisation is 2.758 % for 50 nodes the improvisation is 3.803%. For wrong node identification, the improvisation for 30 nodes is 15.78% improvisation for 40 nodes is 17.89% for 50 nodes it is 15.89%. For throughput, the improvisation for 30 nodes is 15.59%, for 40 nodes the improvisation is 9.92%, and for 50 nodes it is 7.09%.

CONCLUSION

The enormous growth in IoT has led to huge amounts of sensor node deployment; moreover, the data are collected from the different sensors and aggregated for efficient energy utilization. This research work presents the PR-Aggregation (Privacy and Reliable Aware) framework, which aims for secured aggregation from any kind of attack as the sensor nodes, and data are verified before aggregation. In order to verify the conditions are designed for verification and if it satisfies then it keeps on transmission to adjacent node, else it stops transmission. Furthermore, the PR-Aggregation framework is evaluated by inducing the dishonest nodes of 30, 40 and 50 considering classified and misclassified identification of it. Thereafter throughput is computed; also, comparative analysis is carried out with the existing aggregation protocol and the PR-Aggregation model performs better. The future scope of research lies in the adoption of data integrity technologies such as block chain due to the rise of deep learning-based attack models.

REFERENCES

I. Lee and K. Lee, ``The Internet of Things (IoT): Applications, investments, and challenges for enterprises," Bus. Horizons, vol. 58, no. 4, pp. 431-440, 2015.

N. Ma, H. Zhang, H. Hu and Y. Qin, "ESCVAD: An Energy-Saving Routing Protocol Based on Voronoi Adaptive Clustering for Wireless Sensor Networks," in IEEE Internet of Things Journal, vol. 9, no. 11, pp. 9071-9085, 1 June1, 2022, doi: 10.1109/JIOT.2021.3120744.

W. Zhou, Y. Jia, A. Peng, Y. Zhang, and P. Liu, ``The effect of IoT new features on security and privacy: New threats, existing solutions, and challenges yet to be solved," IEEE Internet Things J., vol. 6, no. 2, pp. 1606-1616, Apr. 2019.

H. Attaullah, T. Kanwal, A. Anjum, G. Ahmed, S. Khan, D. B. Rawat, and R. Khan, ``Fuzzy-logic-based privacy-aware dynamic release of IoT-enabled healthcare data," IEEE Internet Things J., vol. 9, no. 6, pp. 44114420, Mar. 2022, doi: 10.1109/JIOT.2021.3103939.

Husnoo, Muhammad Akbar, et al. "Differential privacy for IoT-enabled critical infrastructure: A comprehensive survey." IEEE Access (2021).

Mehedi, SkTanzir, et al. "Dependable Intrusion Detection System for IoT: A Deep Transfer Learning-based Approach." IEEE Transactions on Industrial Informatics (2022).

Sei, Yuichi, and Akihiko Ohsuga. "Private true data mining: Differential privacy featuring errors to manage Internet-of-Things data." IEEE Access 10 (2022): 8738-8757.

Onesimu, J. Andrew, et al. "Privacy Preserving Attribute-Focused Anonymization Scheme for Healthcare Data Publishing." IEEE Access 10 (2022): 86979-86997.

J. Andrew and J. Karthikeyan, ``Privacy-preserving Internet of Things: Techniques and applications," Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 32293234, Aug. 2019, doi: 10.35940/ijeat.F8830.088619.

A. V. Dastjerdi and R. Buyya, ‘‘Fog computing: Helping the Internet of Things realize its potential,’’ Computer, vol. 49, no. 8, pp. 112–116, 2016.

R. Deng, R. Lu, C. Lai, and T. H. Luan, ‘‘Towards power consumptiondelay tradeoff by workload allocation in cloud-fog computing,’’ in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2015, pp. 3909–3914.

M. Mukherjee, R. Matam, L. Shu, L. Maglaras, M. A. Ferrag, N. Choudhury, and V. Kumar, ‘‘Security and privacy in fog computing: Challenges,’’ IEEE Access, vol. 5, pp. 19293–19304, 2017.

R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, ‘‘A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT,’’ IEEE Access, vol. 5, pp. 3302–3312, 2017.

C. Miao et al., “Cloud-enabled privacy-preserving truth discovery in crowd sensing systems,” in Proc. 13th ACM Conf. Embedded Networked Sensor Syst., 2015, pp. 183–196.

Y. Li et al., “An efficient two-layer mechanism for privacy-preserving truth discovery,” in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2018, pp. 1705–1714.

C. Miao, L. Su, W. Jiang, Y. Li, and M. Tian, “A lightweight privacypreserving truth discovery framework for mobile crowd sensing systems,” in Proc. IEEE Conf. Comput. Commun., 2017, pp. 1–9.

Z. Li, Z. Zheng, S. Guo, B. Guo, F. Xiao and K. Ren, "Disguised as Privacy: Data Poisoning Attacks against Differentially Private Crowdsensing Systems," in IEEE Transactions on Mobile Computing, doi: 10.1109/TMC.2022.3173642.

Z. Zheng, Z. Li, H. Jiang, L. Y. Zhang and D. Tu, "Semantic-Aware Privacy-Preserving Online Location Trajectory Data Sharing," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2256-2271, 2022, doi: 10.1109/TIFS.2022.3181855.

Z. Ma, L. Wang and W. Zhao, "Blockchain-Driven Trusted Data Sharing With Privacy Protection in IoT Sensor Network," in IEEE Sensors Journal, vol. 21, no. 22, pp. 25472-25479, 15 Nov.15, 2021, doi: 10.1109/JSEN.2020.3046752.

L. C. Mutalemwa and S. Shin, "Novel Approaches to Realize the Reliability of Location Privacy Protocols in Monitoring Wireless Networks," in IEEE Access, vol. 9, pp. 104820-104836, 2021, doi: 10.1109/ACCESS.2021.3099499.

S. Lata, S. Mehfuz and S. Urooj, "Secure and Reliable WSN for Internet of Things: Challenges and Enabling Technologies," in IEEE Access, vol. 9, pp. 161103-161128, 2021, doi: 10.1109/ACCESS.2021.3131367.

K. Haseeb, K. M. Almustafa, Z. Jan, T. Saba and U. Tariq, "Secure and Energy-Aware Heuristic Routing Protocol for Wireless Sensor Network," in IEEE Access, vol. 8, pp. 163962-163974, 2020, doi: 10.1109/ACCESS.2020.3022285.

C. Lyu, X. Zhang, Z. Liu and C. -H. Chi, "Selective Authentication Based Geographic Opportunistic Routing in Wireless Sensor Networks for Internet of Things Against DoS Attacks," in IEEE Access, vol. 7, pp. 31068-31082, 2019, doi: 10.1109/ACCESS.2019.2902843.

D. Yang, A. Mahmood, S. A. Hassan and M. Gidlund, "Guest Editorial: Industrial IoT and Sensor Networks in 5G-and-Beyond Wireless Communication," in IEEE Transactions on Industrial Informatics, vol. 18, no. 6, pp. 4118-4121, June 2022, doi: 10.1109/TII.2022.3142149.

R. Khan, J. Teo, M. A. Jan, S. Verma, R. Alturki and A. Ghani, "A Trustworthy, Reliable, and Lightweight Privacy and Data Integrity Approach for the Internet of Things," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 511-518, Jan. 2023, doi: 10.1109/TII.2022.3179728.