A Privacy-Preserving Big Data Frameworks for Ensuring Secure and Efficient Network Communication in Distributed Systems

Muralikrishna Dudaka,
Senior Data Engineer , Cisco System Inc.,
Santa Clara, CA, USA, 95051,
reachmuralid@gmail.com

Rahul Jain,
Technical Engineering Leader, Cisco System Inc.,
Shasta Ct , Pleasanton , California , USA 94566,
rahul.bsw@gmail.com

***Abstract*— The high rates of big data analytics and distributed communication systems coupled with the increasing demand of safe and privacy protection network infrastructures. Sensitivity safeguard of information that is sent infinitely is a significant issue because of cyber threats, information leaking and scalability concern in the present communication setup. In this work, a privacy preserving big data architecture of implementing a secure network communication between nodes incorporating anonymization-based privacy protection, lightweight encryption, trustful secure routing, and adaptive access control into the distributed network option architectures. The methodology integrates privacy transformation, the encrypted communication, integrity validation, and threat-conscious monitoring to enhance confidentiality and secure data transfer with minimum overhead. Evaluation of the performance shows that it has significantly improved with the performance achieving 96.8% privacy preservation, 98.2% integrity accuracy and 97.6% attack detection with 8.4% communication overhead. The proposed framework is 3.2 times smarter in terms of privacy preservation, 5.0 Mbps faster in throughput, 2.8 times more accurate in integrity, and then, the overhead is also reduced by 2.5. The framework offers an effective, scalable, and robust privacy-aware secure communication in big data-driven network environments.**

Keywords-Privacy Preservation, Big Data Security, Secure Network Communication, Data Anonymization, Distributed Data Processing.

I. INTRODUCTION

Modern network infrastructures have been turned into highly interconnected spaces producing enormous data volumes, that are heterogeneous faces. Cloud services, the Internet of Things, enterprise system, and intelligent communication platform generate valuable information to be analyzed, optimized, and used to make decisions almost daily, as evidenced by network traffic. Big data technologies have become indispensable in terms of controlling this large volume of information to enable a scalable storage, speedy processing, and effective communication between the distributed systems [1-2].

Nonetheless, the augmented dependency on big data in the network communication has brought disastrous issues associated with privacy and security. A complex network will experience sensitive user information, communication patterns, and transactional data that will be vulnerable to transfer via interception, leakage and unauthorized access [3]. The conventional security mechanisms are only designed to protect communication channels but might not be sufficient to deal with privacy threat posed by mass aggregation and processing of data.

The combination of big data and network communication poses several privacy issues, as the amounts, speeds, and types of data being transmitted greatly surpass those before the advent of big data. The information that can be obtained in various sources is likely to contain personal identifiers and trends in behavior as well as confidential records that are better handled with utmost security [4]. With more and more decentralized communication systems, it has become an important research issue to retain some degree of privacy and still exchange data in a fairly efficient manner.

Additionally, regulatory directives and heightened awareness with reference to data protection has augmented the need of secure frameworks that could help to support privacy conscious communication systems [5-6]. In an environment with large number of networked machines, the issue of ensuring confidentiality, integrity, and controlled access still remains critical in securing sensitive information and encouraging reliable data driven services.

Secure network communication relies on the mechanisms that protect against such threats as eavesdropping, traffic analysis, unauthorized disclosure and malicious intrusion. Big data systems increase these issues due to the fact that large volumes of data can be deployed across diverse (distributed) nodes, expanding the scopes of attack [7-8]. In such environments, security needs to be coordinated on the levels of data transmission, storage and processing.

The increasing reliance on data-centric forms of communication systems is imposing more and more emphasis on the need to maintain privacy when using secure forms of communication systems [9]. Developments in both distributed computing and communication technologies have further inspired research efforts into the development of a strong and effective privacy conscious network security solution.

Secure networking has now become a crucial concern to protect privacy in the context of communication relying on big data [10]. The response to privacy concerns and security needs constitute the basis to come up with effective and credible communication systems.

II. RELATED WORKS

Privacy maintenance in secure network communication, which becomes possible with the assistance of big data, has become the subject of an unmatched amount of research in the past, as more and more issues regarding the exposure of data, computer-related threats, and unauthorized access are

979-8-3315-6227-4/26/$31.00 ©2026 IEEE

raised. The previous research has emphasized on enhancing confidentiality, secure transmission, and privacy conscious processing in distributed communication environment. Scalability and communication efficiency [11-12]. In large-scale networks Multiple technologies have been investigated to address the privacy risk without jeopardizing scalability and the efficiency of communication in large-scale networks.

2.1 Encryption-Based Approaches

Some of the most popular solutions that have been widely used in securing sensitive information in network communication is based on encryption. Since time immemorial, there has been the use of traditional cryptographic methods like symmetric and asymmetric encryption to protect data transmitted against interception. More advanced solutions such as homomorphic encryption, attribute-based encryption etc. have received interest due to their ability to provide secure data operations whilst maintaining privacy [13]. Even though these methods enhance confidentiality, issues connected to computational load, as well as scalability, continue to play a significant role in big data settings [14].

2.2 Anonymization and Data masking Techniques.

Anonymization has been a highly researched method of protecting privacy through eliminating or concealing sensitive identifiers prior to sharing or manipulating data [15]. The k-anonymity methods, l-diversity methods and the methods of differential privacy have been used to minimize the disclosure risk of network generated datasets. In further enhancing the privacy protection, data masking and perturbation techniques also alter sensitive features, maintaining analytical utility [16]. Nevertheless, striking the right balance between privacy protection and utility of the information remains a research issue.

2.3 Access Control and Authentication Models

One important element that would be used to secure the communication systems of the big data would be access control mechanisms, which would be used to restrict unauthorized access [17]. Models of role-based access control, attribute-based access control and multi-factor authentication have been incorporated in the distributed network setting to enhance protection of privacy. The new trends in research focus on varying and context-dependent access control systems to adjust to the changing communication conditions [18]. Although these developments are in place, implementing scalable and flexible control policies in distributed infrastructures continues to be complicated [19].

2.4 Blockchain and Distributed Security Models.

The decentralized and tamper-resistant nature of blockchain-based approaches has seen them emerge as promising solutions to privacy-preserving secure communication. Scientists have ventured into blockchain as a means of ensuring safe data sharing, identity safety and reputation management over distributed networks [20]. Moreover, distributed security models which involve intrusion detection and privacy-protective mechanisms have been postulated to enhance resilience provided by augmenting against the changing threats [21]. Controversies associated with latency and other resources continue to affect the widespread use.

The current methods have resulted in a significant progress in privacy-preserving secure communication in the form of encryption, anonymization, access control, and distributed security models [22]. There have been, however, issues of scalability, efficiency and comprehensive protection of privacy that points to further research to be conducted in the area of securing the communication between big data.

III. PROPOSED METHODOLOGY

The proposed privacy and preserving big data model to secure network communication ensures to provide privacy, data integrity, controlled access and efficient processing of extensive network data with minimal privacy leakage. The framework works according to interrelated layers comprising of secure data acquisition, privacy-sensitive transformation, encrypted communication, trust verification and adaptive access validation as shown in Fig 1. The added value of each of these stages is to ensure that it handles data produced by the distributed communication networks without influencing scalability.

In data acquisition phase, a raw network data that is gathered over distributed sources is modeled as a high-volume dataset.

$$D = \{d_1, d_2, d_3, \ldots, d_n\} \quad (1)$$

Where D being the total network data, d_i is the i^{th} instance of data collected by the network, and n is the number of data instances that have been collected by the network. Such logs could comprise communication logs, packet metadata, routing information and user-generated information. Privacy transformation is originally done since privacy sensitive attributes may be a part of each record in the database.

Privacy transformation is used in characterizing sensitive data anonymization

$$D_p = T(D) \quad (2)$$

Where D_p is transformed privacy-conserving data, and the transformation function T.The transformation cloaks or generalizes sensitive attributes, and then transmits data. The reason is that the privacy leakage risk is estimated as

$$P_r = \frac{S_e}{S_t} \quad (3)$$

Where P_r is the probability of privacy risk, S_e is exposed sensitive information and S_t is the total sensitive attributes. The smaller the P_r values are, the more privacy is preserved.

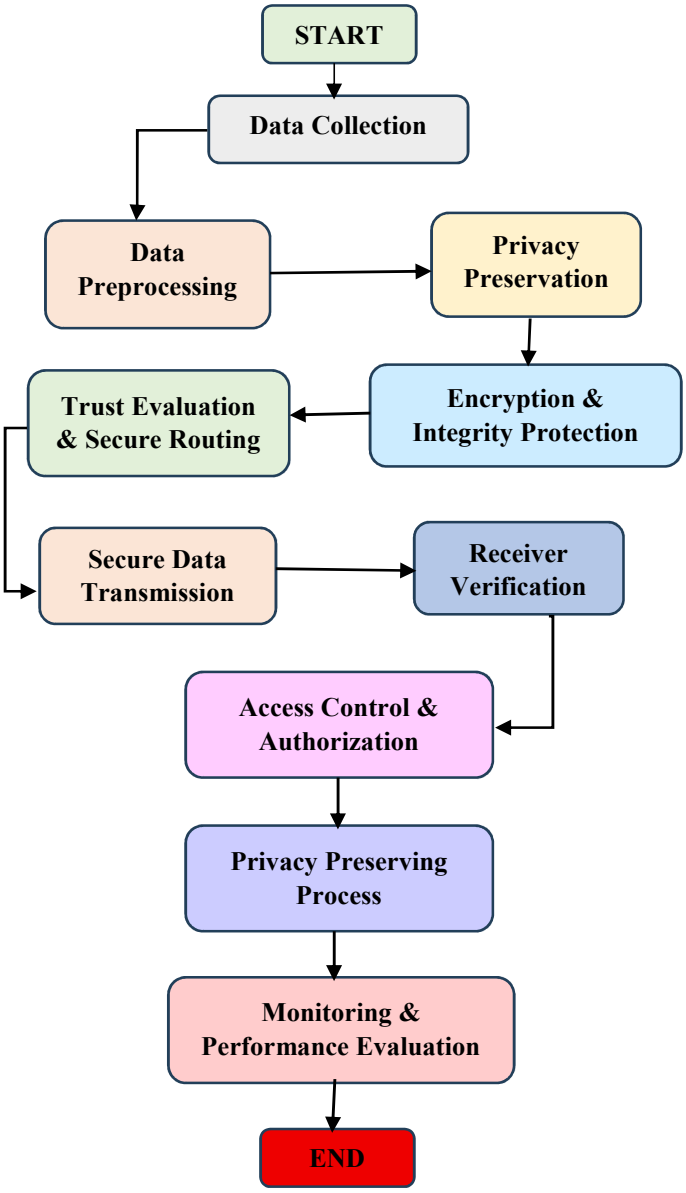

Fig 2: Flow Chart of Proposed Privacy-Preserving Big Data Framework

To enhance the security levels further, encryption is used, which precedes communication with the help of.

$$C = E(K, D_p) \tag{4}$$

Where C is encrypted ciphertext, E is encryption function, K is secret encryption key and D_p is privacy possible data. This formula makes sure that the transformed data is not leaked as is sent through the network.

The model embodies secure communication among the distributed nodes in a model of communication trust.

$$T_n = \alpha S + \beta A + \gamma I \tag{5}$$

Where T_n comes out as the node trust score, S is security compliance value, A is authentication reliability, I is the level of communication integrity, and α, β and γ are weighting factors that satisfy

$$\alpha + \beta + \gamma = 1 \tag{6}$$

In this model trust, the nodes of communication are evaluated and are not allowed to take part in secure communication. An increase in the value of trust means that communication parties can be trusted.

To ensure good performance of transmission performance, the strength of data confidentiality is determined by using the efficiency of encryption.

$$E_f = \frac{D_v}{T_e} \tag{7}$$

Where E_f the efficiency of encryption, D_v the volume of data encrypted, and T_e the time to execute encryption. The metric makes certain that security mechanisms that work remain computationally feasible when using big data communication. A hash-based integrity model is used to check the integrity of data.

$$H = h(D_p) \tag{8}$$

Where H is generated hash value and h cryptography hash function. At the receiving node, the integrity is checked with the help of.

$$H_s = H_r \tag{9}$$

Where H_s is the hash of the sender and H_r is a hash which the receiver calculates. Equality ascertains transmissions of data that remain identical.

Integration is by adaptive access control through an authorization probability.

$$A_u = \frac{R_v}{R_t} \tag{10}$$

Where A_u is authorization confidence of the user, R_v is the accepted access requests and R_t is the total access requests. The equation measures access legitimacy in a dynamic fashion in the distributed environment.

To regulate the amount of privacy disclosure in the process of distributed processing, differential privacy noise can be added as

$$D_s = f(D_p) + N(\mu, \sigma) \tag{11}$$

Where D_s forms a read secure shared data output, $f(D_p)$ forms a processed privacy-preserved data, and $N(\mu, \sigma)$ is a noise generated with a mean and standard deviation of mu and sigma, respectively. This reduces inferences attacks, and also, maintains an analytical usefulness. Secure throughput is used to model the communication security performance.

$$T_s = \frac{P_s}{\tau} \tag{12}$$

Where T_s is the secure throughput, P_s is the successfully protected packets and τ is the time of transmission. This is an indication of efficient communication safety.

Likelihood of threats estimation is provided in detection of attacks.

$$A_t = \frac{M_d}{N_t} \quad (13)$$

Where A_t is probability of an attack, M_d is probability of a detachment of malicious programs, N_t is observed traffic events. This enables dynamic security strengthening in case the likelihood of attack grows. The total security effectiveness is computed as

$$S_e = w_1 C_f + w_2 I_f + w_3 A_f (14)$$

Where S_e stands for score of security effectiveness, C_f represents confidentiality factor, A_f is access control factor and w_1, w_2, w_3 are satisfying weighing Co-efficient.

$$w_1 + w_2 + w_3 = 1 \quad (15)$$

This composite model is a measure of framework wide security performance. To have scalable distributed communication, the estimation of communication overhead is

$$O_c = \frac{T_p - T_n}{T_n} \quad (16)$$

Where O_c represents communication overhead, T_p represents the time of communication that is protected, and T_n represents normal transmission time. Reduced overhead means that privatizing operation is scalable. The suggested framework will also make use of optimization of secure use of resources by use of

$$R_u = \frac{R_a}{R_t} \quad (17)$$

Where R_u is efficiency of using resources, R_a is resource utilization rate, and R_t is the total rate at which resources are used. There is a balance of privacy utility which is assessed by using

$$U_p = \lambda D_u - (1 - \lambda) P_r \quad (18)$$

Where U_p denotes privacy utility score, D_u is data utility, P_r is privacy risk, and λ being balancing coefficient. This guarantees privacy does not inappropriately limit the analytical utility.

Coordinated architecture ensures secure communication of big data via integrated privacy-preservation, trust-management and adaptive security. Distributed sources transmit data in such a way that, to the extent possible, certain sensitive attributes are not revealed prior to the secure encrypted transfer through safe data communication nodes that are selected based on trust evaluation. Different and constantly changing threat probabilities and secure throughput are taken into account and adjust the protection of communication dynamically. Before the processing and retrieval requests can be made, entity legitimacy must be verified. Differential privacy mechanisms can protect outputs from distributed computation, which do not allow inference attacks. It continuously checks the ensure security of operation in dynamic situation confidentiality, integrity and authorization. Preservation of privacy is reflected throughout all the layers of communication via encrypted chaps, anonymization and probabilistic access control. This scalable, private and decentralized architecture enhances the (large-scale) distributed ob Data Enclaves' confidentiality, integrity, resilience and secure communication in terms of coordinated mathematical modelling and algorithmic security management.

IV. RESULT

The privacy-preserving big data privacy paradigm is tested on multiple security and efficiency metrics to measure its privacy protection capability, communication reliability, computational overhead and secure transmission performance. The analysis is dedicated to the evaluation of the effectiveness of the framework to enhance safe data transfer preserving privacy in case of large-scale distributed networks. The performance assessment is based on the quantitative metrics, which reflect the confidentiality, assuring integrity, effectiveness in communication, resistance to attacks, and scalability. Further comparison with current methods shows the effectiveness of the framework in realizing a better secure communication with reduced privacy risks and more efficient processing.

4.1 (PPR) Privacy Preservation Rate.

Privacy Preservation Rate measures the extent to which sensitive data can be maintained such that it does not leak out during communication and processing. It is computed as

$$PPR = \frac{S_p}{S_t} \times 100 \quad (19)$$

Where PPR = Privacy Preservation Rate (%) , S_p = Number of protected sensitive attributes, S_t = Total sensitive attributes Greater values of PPR mean that privacy is better secured. An increased rate of preserving privacy indicates less risk of information disclosure and an enhanced anonymization performance. The proposed framework offers better protection against privacy given the in-built encryption and privacy conscious transformations.

4.2 (SCT) Secure Communication Throughput.

Secure Communication Throughput is a metric which measures the number of packets which have been transmitted successfully and are covered by the appropriate security protocols, and are sent in a unit time interval.

$$SCT = \frac{P_s}{T} \quad (20)$$

Where SCT = Secure throughput, P_s = Successfully secured packets,T = Communication time This measure is a combination of the security and the communication efficiency. High throughput means that privacy protection do not markedly drag down the performance of communication.

4.3 (DIA) Data Integrity Accuracy.

Data Integrity Accuracy is a measure that assesses accuracy of data that is regularly received following a secure transmission.

$$DIA = \frac{V_c}{V_t} \times 100 \quad (21)$$

Where DIA = Data integrity accuracy (%), V_c = Correctly verified packets, V_t= Total transmitted packets Increasing the integrity accuracy implies more significant tamper detection and transmitting security reliability.

4.4 (COR) Communication Overhead Ratio

Communication Overhead Ratio quantifies more overhead delay due to privacy preserving security activities.

$$COR = \frac{T_s - T_n}{T_n} \times 100 \quad (22)$$

Where COR= Communication overhead ratio (%),T_s= Secure communication time,T_n= Normal transmission time

Since they indicate scalable secure communication, Lower overhead values are desirable.

4.5 Attack Detection Rate (ADR)

Attack Detection Rate measures capability to detect unwanted behaviors on the packet stream of network traffic.

$$ADR = \frac{A_d}{A_t} \times 100 \quad (23)$$

Where ADR = attack detection rate (percentage), A_d = detected attacks that are correct, A_t = total number of attacks ADR can be used to measure the level of resilience against cyber threats and unwarranted intrusions.

Table I: Evaluation of PPR, ADR and DIA of existing approach with proposed approach

Approach	**PPR**	**ADR**	**DIA**
Conventional Encryption Model [19]	84.5	85.6	89.4
Differential Privacy Framework [20]	88.9	87.4	91.3
Blockchain Security Model [18]	91.2	90.2	93.5
Access Control Based Model [21]	89.8	88.7	92.1
Hybrid Privacy Security Model [22]	93.6	94.1	95.4
Proposed Framework	**96.8**	**97.6**	**98.2**

The comparison Table I and Fig 2 highlights that the proposed framework is the best approach to provide high levels of privacy preservation rate (PPR), the highest attack detection rate (ADR), and the highest level of data integrity accuracy (DIA) compared to all the present approaches. With traditional encryption mechanism, privacy and security efficiency is lower because there is insufficient adaptive protection mechanism. Improving integrity and attack detection with blockchain and hybrid security models comes at a higher level of complexity. Proposed framework obtains the highest level of PPR (96.8%), ADR (97.6%) and DIA (98.2%) that guarantee better level of privacy protection, reliable communication and integrity protection.

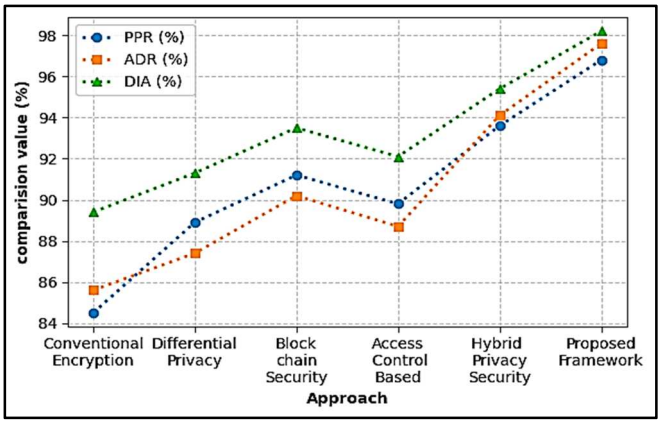

Fig 2: Illustration of compared PPR, ADR and DIA

Table II: Evaluation of COR and SCT of existing approach with proposed approach

Approach	**COR**	**SCT**
Conventional Encryption Model [19]	18.7	75.2
Differential Privacy Framework [20]	15.8	80.6
Blockchain Security Model [18]	14.1	83.7
Access Control Based Model [21]	16.2	81.4
Hybrid Privacy Security Model [22]	10.9	89.5
Proposed Framework	**8.4**	**94.5**

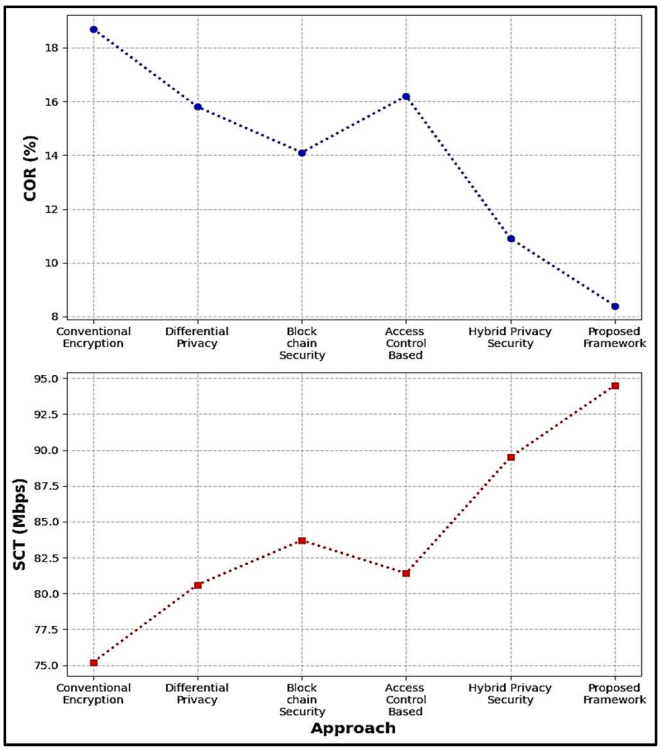

Fig 3: Illustration of compared COR and SCT

As illustrated in the Table II and Fig3, the proposed framework can provide the lowest Communication Overhead Ratio (8.4%) and Highest Secure Communication Throughput (94.5Mbps) among the existing approach. Both conventional and access control models have high overhead, but low throughput. Results show that the proposed framework delivers better secure communication performance in terms of efficiency, scalability and speed.

The analysis of the performance confirms that the given framework is very useful in terms of privacy protection, communication security, integrity assurance, and attack resilience with minimal overhead. Compared analysis has shown that the results are better than current methods, which proves the effectiveness of the framework in achieving safe privacy protecting big data communication.

V. CONCLUSION

It introduced a privacy preserving big data architecture of secure network communication which is designed to improve confidentiality, integrity, access control and scalable network communication performance on distributed environments. The framework effectively incorporated privacy protection, secure transmission, trust evaluation, and adaptive security mechanism to deal with major challenges faced with privacy leakage and cybers threats in large scale network communication. The effectiveness of the framework was also confirmed by the performance analysis in the form of the meaningful results gained in all the evaluation measures. The Privacy Preservation rate of 96.8%, Accuracy of Data Integrity of 98.2, Secure Communication throughput of 94.5 Mbps and the Attack Detection rate of 97.6 indicates good protection and effective communication performance as well as maintaining low communication overhead rate of 8.4. Comparative results demonstrated statistically significant improvements over traditional encryption, differential privacy, blockchain-based encryption and access-controlled encryption and hybrid approaches to security. The result validates that the framework enhances secure data communication, and minimizes privacy risks and the computational burden. In general, the proposed model offers the robust, scalable, and efficient approach to privacy-conscious secure communication in the network infrastructure that is driven by big data amounts.

Future research may aim to incorporate AI-based threat prediction, federated learning, and quantum-resistance security systems to bring about more significant advances in privacy protection, dynamic defense, lower overheads and intelligent secure communication within dynamic large-scale networks

REFERENCE

[1] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT," IEEE Transactions on Industrial Informatics, vol. 16, no. 6, pp. 4177–4186, Jun. 2020.

[2] X. Tang, M. Shen, Q. Li, L. Zhu, T. Xue, and Q. Qu, "PILE: Robust Privacy-Preserving Federated Learning Via Verifiable Perturbations," IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 6, pp. 5005–5023, Nov.–Dec. 2023.

[3] S. N. Pillai, "AI-Driven Sustainability by Design: Integrating Iot Sensors and Automated Regulatory Reporting Within SAP EHS Systems", *JICRCR*, pp. 135–146, Apr. 2026.

[4] V. Roy, "A Context-Aware Internet of Things (IoT) Founded Approach to Scheming an Operative Priority-Based Scheduling Algorithms", 2024, 13(1), Pp. 28–35, Https://Doi.Org/10.54216/Jcim.130103.

[5] X. Li, M. He, and J. Ni, "Secure and Privacy-Preserving Network Slicing in 3GPP 5G System Architecture," IEEE Internet of Things Journal, vol. 11, no. 2, pp. 2055–2068, Jan. 2024.

[6] A. A. Khan, K. K. Almuzaini, V. D. J. Macedo, S. Ojo, V. K. Minchula, V. Roy, MaReSPS for energy efficient spectral precoding technique in large scale MIMO-OFDM, Physical Communication, Volume 58, 2023, 102057, ISSN 1874-4907, https://doi.org/10.1016/j.phycom.2023.102057.

[7] D. Fu, W. Bao, R. Maciejewski, H. Tong, and J. He, "Privacy-Preserving Graph Machine Learning from Data to Computation: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 3, pp. 1211–1230, Mar. 2024.

[8] S. N. Pillai, "AI, IoT, and Cloud Convergence in SAP Ecosystems: Driving the Smart Factory of the Future," *Journal of International Crisis and Risk Communication Research*, vol. 9, no. 2, pp. 77, 2026.

[9] P. Batta and S. K. Sharma, "Privacy Preserving Data Analytics in DEI Empowered IoV," Information Systems Frontiers, vol. 27, no. 1, pp. 1–19, 2025.

[10] C. Hong, "Recent Advances of Privacy-Preserving Machine Learning Using Fully Homomorphic Encryption," Security and Safety, vol. 4, no. 1, pp. 1–18, 2024.

[11] M. A. Ferrag, L. Maglaras, H. Janicke, J. Jiang, and T. Shu, "Authentication Protocols for Internet of Things: A Comprehensive Survey," Security and Communication Networks, vol. 2021, pp. 1–41, 2021.

[12] Vandana Roy et. Al. "A Machine Learning-Based Big EEG Data Artifact Detection and Wavelet-Based Removal: An Empirical Approach", Mathematical Problems in Engineering Volume 2021, Article ID 2942808, https://doi.org/10.1155/2021/2942808, pages: 1-11, 2021.

[13] Z. Yang, K. Yang, L. Lei, K. Zheng, and V. C. M. Leung, "Blockchain-Based Decentralized Trust Management in Vehicular Networks," IEEE Internet of Things Journal, vol. 7, no. 2, pp. 1495–1505, Feb. 2020.

[14] Q. Xia, E. B. Sifah, K. O.-B. Agyekum, A. N. Smahi, X. Gao, and J. Du, "BBDS: Blockchain-Based Data Sharing for Electronic Medical Records in Cloud Environments," Information Processing and Management, vol. 57, no. 1, pp. 102–119, 2020.

[15] H. Wu, P. Wang, X. Zhu, and C. Jiang, "A Secure and Efficient Data Transmission Scheme for Cluster-Based Wireless Sensor Networks," IEEE Systems Journal, vol. 15, no. 3, pp. 4050–4061, Sept. 2021.

[16] J. Ren, H. Guo, C. Xu, and Y. Zhang, "Serving at the Edge: A Scalable IoT Architecture Based on Transparent Computing," IEEE Network, vol. 33, no. 5, pp. 96–105, Sept.–Oct. 2019.

[17] M. Aloqaily, I. A. Ridhawi, H. B. Salameh, and Y. Jararweh, "Data and Service Management in Densely Crowded Environments: Challenges, Opportunities, and Recent Developments," IEEE Communications Magazine, vol. 57, no. 4, pp. 81–87, Apr. 2019.

[18] R. Siyal, M. A. Khan, and S. Ahmed, "Secure big data sharing with hybrid encryption and deep learning-based threat detection in blockchain environments," Journal of Big Data Analytics, vol. 12, no. 3, pp. 145–159, 2025.

[19] J. Li, H. Wang, and T. Chen, "Enhancing privacy-preserving intrusion detection in blockchain-based networks using federated deep learning," Data Science Journal, vol. 22, no. 31, pp. 1–14, 2023.

[20] M. Alrizq, S. Stalin, S. Sultan, V. Roy, A. Mishra, A. Chandanan, N. Awadallah, P. Venkatesh, (2023). Optimization of sensor node location utilizing artificial intelligence for mobile wireless sensor network. Wireless Networks. 1-13. 10.1007/s11276-023-03469-4.

[21] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "A hybrid approach to privacy-preserving federated learning," arXiv preprint arXiv:1812.03224, 2018.

[22] H. Jiang, Y. Gao, S. M. Sarwar, L. GarzaPerez, and M. Robin, "Differential privacy in privacy-preserving big data and learning: Challenge and opportunity," arXiv preprint arXiv:2112.01704, 2021.