International Journal of Data Science and Artificial Intelligence (IJDSAI)
Volume 02, Issue 03, May – June (2024)

RESEARCH ARTICLE

IOT-CENTRIC DATA PROTECTION USING DEEP LEARNING TECHNIQUE FOR PRESERVING SECURITY AND PRIVACY IN CLOUD

C. Senthil Singh [1,*], Sameena Naaz [2] and G. Saranya [3]

[1] Department of Electronics and Communication Engineering, Shadan Women College of Engineering and Technology, Hyderabad, India.
[2] Department of Computer Science, Roehampton University, London, SW15 5PH, United Kingdom.
[3] Department of computer science and Engineering, SA Engineering college, Poonamallee, Tamil Nadu 600077 India.

[*]*Corresponding e-mail: senthilsingh@gmail.com*

Abstract – The Internet of Things (IoT) describes a system where interconnected physical objects are connected online. As the collection and sharing of vast amounts of personal data grow, so do concerns over user privacy within IoT environments. While IoT devices offer significant advantages in terms of productivity, accuracy, and financial benefits by minimizing human intervention and providing exceptional flexibility and convenience, they also face challenges related to communication overhead, security, and privacy. To address these issues, a novel Internet of Things-based Cloud Information Security Preservation (IoT-CISP) has been proposed. This approach enhances the model's effectiveness and ensures security by first separating sensitive data from non-sensitive data using an SVM classifier, and then employing this data for partial decryption and analysis. Sensitive data is protected through Okamoto-Uchiyama encryption, ensuring that data storage, analysis, and sharing are conducted securely to maintain the system's safety and privacy. The effectiveness of this novel method was assessed against existing methodologies using parameters like precision, accuracy, F1 score, and recall, revealing its superior security and efficiency compared to other schemes. Results demonstrate that the IoT-CISP approach offers encryption times that are 31.24%, 23.12%, and 33.03% shorter than those of the CP-ABE, GDBR, and HP-CPABE algorithms, respectively.

Keywords *– cloud computing, Internet of things, Support vector machine, Okamoto Uchiyama.*

1. INTRODUCTION

IoT devices, generate enormous amounts of data but lack the storage and the processing capability to process it, which are the foundation of the cyber-physical system [1, 2]. The cloud platform has established itself as a top-tier method of storing, analyzing, and exchanging data with numerous stakeholders for the best results [3, 4]. However, it is not advisable to put your faith in a third-party cloud platform, particularly for preserving private information, as renting data to the cloud results in the gadget losing ownership of it [5].

An average data loss expense worldwide will be $4.35 million in 2022, up to 2.6% and 12.7% from 2021 and 2020, respectively, according to research by IBM and Ponemon Institute [6]. These factors have made data protection a major problem, which has inspired researchers to provide strategies for maintaining data privacy [7]. Most of the models that are now in use are based on encryption techniques [8], and differential privacy, but they have a limit According to research by IBM and the Ponemon Institute, the average cost of a data breach globally will be $4.35 million in 2022, an increase of 2.6% and 12.7% from 2021 and 2020, respectively.

These reasons have made data protection a significant issue, prompting researchers to offer techniques for preserving data privacy. The majority of models now in use are based on encryption methods as well as differential privacy, although their accuracy, utility, and efficiency might be improved. As far as the authors are aware, there are currently no models that adequately balance the accuracy and privacy of the data that is being outsourced [9,10]. To overcome these challenges Internet of Things based Cloud Information Security preservation (IoT-CISP) has been proposed. The following are the paper's main contributions:

- The proposed IoT-CISP approach enhances the model's effectiveness and ensures security by first separating sensitive data from non-sensitive data using an SVM classifier, and then employing this data for partial decryption and analysis.
- Sensitive data is protected through Okamoto-Uchiyama encryption, ensuring that data storage, analysis, and sharing are conducted securely to maintain the system's safety and privacy.

- The effectiveness of this novel method was assessed against existing methodologies using parameters like precision, accuracy, F1 score, and recall, revealing its superior security and efficiency compared to other schemes.

The remaining sections are arranged as follows: Section 3 presents the proposed framework. The experimental setting is demonstrated in Section 4. The collected results are then detailed in Section 5. With a conclusion and future work, the paper is completed.

2. LITERATURE SURVEY

In 2021, Zhang et al. [11] proposed the HP-CP-ABE hidden access policy to safeguard data security and authenticate authorized users. One potential issue is the possibility of attackers launching attribute values to reveal. The parameter data is included in several HP-CP-ABE methods' access controls, employing guessing attacks (AVGA). To address the computational PBDHE assumption as a means to establish the selective IND-AVGA security of the proposed scheme, this is the first time this assumption has been applied.

In 2023 Vaidya S [12] Suggested a hidden approach framework paired with a based encrypted approach to improve security associated with healthcare IoT information. The system ensures precise control over access to encrypted data and safeguards the privacy of clinical clients. The proposed system far surpasses existing systems in terms of, security, storage load, and computing efficiency, especially when the approach structure is concealed. It presents a new approach for strongly transmitting data in the context of IoT.

In 2023 Wang. C et al. [13] Proposed a robust data encryption scheme called Attribute Hiding and Multiple Authorization Centers-based Data Hierarchical Encryption Scheme (AH-MAC-DHE). This strategy Maintains private data by disguised access controls and user features. to handle the problem of private data being leaked. Assuming judgmental q-parallel the coefficient of the Bilinear Diffie-Hellman, they have shown that AH-MAC-DHE is reliable and offers security for privacy as well as anti-collusion. According to experimental findings, AH-MAC-DHE performs better than current methods.

In 2022 Li, M., Xiao, D., et al. [14] proposed based on compressive sensing, utilizing private cloud for three different levels of cloud service users. This enables the sensor-cloud system to offer a variety of multimedia service levels and security assurances. From the standpoint of consumers of cloud-based services, ensuring the privacy of essential data is a challenge. The suggested approach successfully balances the relationships between cloud service providers, sensor network suppliers, and cloud service customers, according to theoretical studies and experimental simulations.

In 2022 Wei et al. [15], proposed Scourge modeling techniques for addressing privacy of data threats in independent systems, and we have analyzed these techniques in the context of GDPR compliance. Additionally, discussed the challenges and identified gaps, offering suggestions for a new modeling technique. This technique not only models' conventional risks to data privacy but also efficiently does GDPR compliance checks.

In 2023 Huang, B. et al [16] Proposed Ciphertext-Policy Attribute-Based Encryption (CP-ABE) as a lattice-based encryption technique. Compared to techniques built on the Learning with Error problem, this particular approach is more efficient because it is based on the Ring Learning By Error issue. Due to unreliable cloud service providers, the approach addresses security and access control concerns for critical data. Evaluations and experimental simulations demonstrate the scheme's excellent applicability and efficiency.

3. PROPOSED METHODOLOGY

A malicious utility provider might take data that has been outsourced from the cloud, store it, analyze it, and share it with the parties involved to gather private information that could be abused. As a result, protecting data has grown to be a difficult undertaking that must be handled carefully. This paper provides a safe data protection technique for preserving confidentiality in a cloud context to address this crucial and difficult problem. It does so by effectively separating the data into sensitive and non-sensitive categories using an SVM classifier, partially decrypting the data, and performing data analysis that increases the model's effectiveness while maintaining security. Okamoto Uchiyama encryption has been used to protect sensitive data. By carrying out safe data storage, analysis, and exchange, the model guarantees the security and privacy of the system. Specific criteria, including precision, accuracy, F1-score, and recall, have been used to compare the suggested method to the existing methodologies. Figure 1 shows the overall framework for the suggested work.

3.1 Support Vector Machine

Support vector machines (SVMs) are used to collect data from IoT devices, which is divided into two categories: sensitive and nonsensitive data. SVMs are a type of supervised machine learning technology that converts complex, highly nonlinear circumstances into binary classification models [5]. The SVM must construct the decision surface, which is a hyperplane, utilizing data samples to maximize the margin around it. During the training stage of the algorithm, each data sample is assigned a class designation and the projected value. The data sample contains what are referred to called characteristics, these are the variables in the data that specify the data sampling vector's activity. The weights applied to each input feature and a collection of support vectors that construct the ideal hyperplane are used to forecast the results of the SVM training phase. In contrast to other neural networks, the SVM maximizes the number of nonzero weights while lowering the overall number of nonzero weights by maximizing the margin. These only match the important traits that provide information that is useful for selecting the hyperplane.

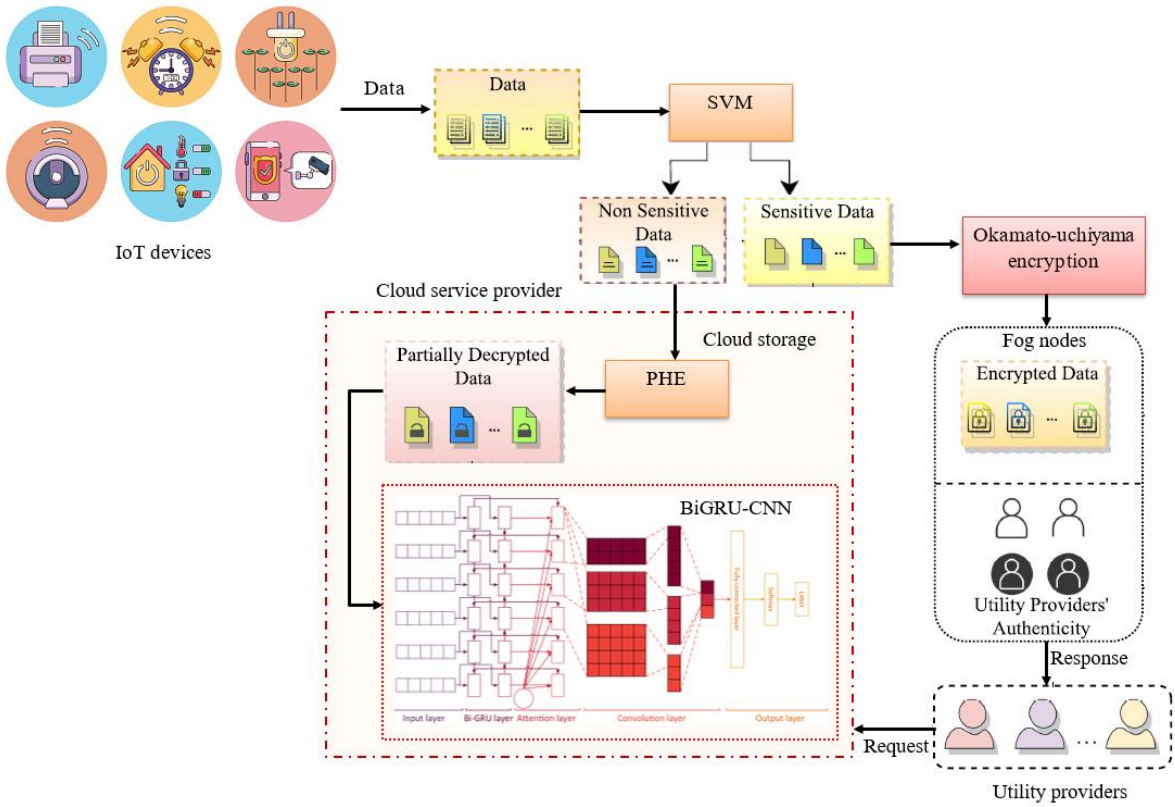

Figure 1. Block diagram of the proposed method

The kernel function modifies the data dimensions to define the hyperplane's form, which is a crucial step in the SVM. Simply put, the kernel function increases the hyperplane's size to help distinguish between the classes. The employment of many kernel types is possible, including the polynomial, linear, both sigmoid and the Gaussian radial basis function. The type of data sample affects how each kernel performs. The simplest kernel, the linear kernel, performs better when applied to linear problems. The supplied characteristics are combined by the polynomial, RBF, and sigmoid kernels to produce support vectors. They work best with non-linear data, but their complexity depends on how many additional features they find. In Figure 2, the SVM flowchart is displayed.

3.2 Okamoto Uchiyama Encryption

Numerous concepts from number theory, discrete mathematics, and abstract algebra are used in the Okamoto-Uchiyama cryptosystem. Numerous of these ideas are fundamental and are applied in various areas of cryptography. However, even though beyond integer computations, there is no need to provide detailed or rigorous treatments, those essential notions are still important. They are not sufficiently covered in mathematics curricula in underdeveloped and developing countries, by examining the fundamental concepts and mathematics used in the Okamoto-Uchiyama algorithm.

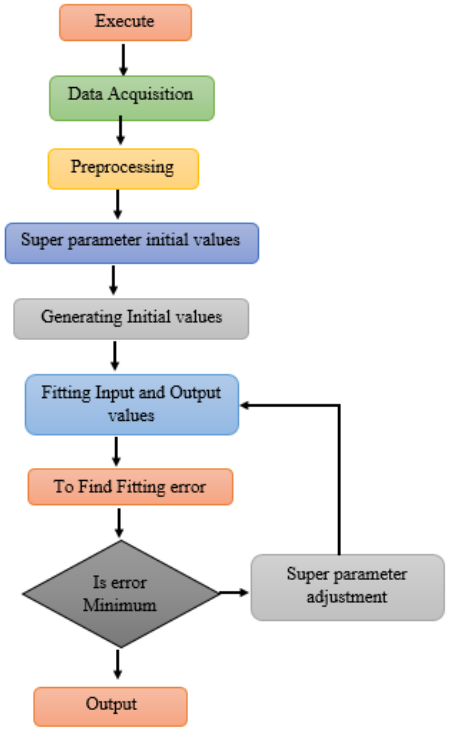

Figure 2. Flow chart of Support Vector Machine (SVM)

3.2.1 Key Generation

The following procedure generates a public/private key:

1. Create the two big primes, A and B.
2. Compute N=B^2A
3. Select an integer with a random value G2 n-1 such that G(B-1)
4. Computes H=GN mod N.

Next, we have (N, G, H) as the public key and (B, A) as the private key.

3.2.2 Encryption

Using the public key (B, A), the following can be done to encrypt a message M < B.

1. A random number R1n-1 should be chosen so that G(B-1)
2. Compute C= GN HR mod N.

The value C is the encryption of M

3.2.3 Decryption

An encrypted message C can be decrypted with the private key (B, A) as follows:

1. Compute P= l (C^{B-1}mod B^2
2. Compute Q= l (G^{B-1}mod B^2 p and Q will be integers.
3. Using the extended Euclidean technique, compute the opposite value of Q modulo B.
   $Q' = Q^{-1}\ mod\ B$
4. Compute M=$PQ'mod\ B$

The value M is the decryption of C.

3.3 Partially Homomorphic Encryption Using Sensitive Data

Schemes for partial homomorphic encryption (PHE) provide the performance of certain accurate processes on scrambled data. The PHE cryptosystem uses public keys. It comprises three stages: encryption, decryption, and key creation. Specifically, since it possesses the public key and secret key throughout the phase of generating keys, every device receives the public key, which is used for encrypted data. The system operator or just one agent has access to the private key needed to decrypt messages. Specifically, there is a time variation in both the public and private keys.

$$Enc(L_1)°Enc(L_2) = Enc(L_1 + L_2) \quad (1)$$

In particular the ciphertext of $L_1 + L_2$) can be obtained from the ciphertexts of L_1 and L_2 As seen in Algorithm 1, the three roles execute all three PHE encryption system features.

3.4 BiGRU-CNN

Figure 3 shows the process of entering current load data into the CNN network's computational layer from the data source layer. The convolutional layer is used to record data correlation. By applying pooling operations to reduce data dimensionality, the pooling layer improves the efficiency of network learning.

Algorithm 1: PHE Cryptosystem
Function Keygen ()
Output: Public key K_P:(N, G), private key:(λ, μ)
Choose two large prime numbers P and Q of equal bit-length and compute N=P.Q;
G←N+1;
λ=$\emptyset(N) = (P-1).(Q-1)$, where $\emptyset(\cdot)$ denotes the Euler's totient function;
$\mu = \emptyset^{-1}$(N)mod N, which is the modular multiplicative inverse of $\emptyset(N)$;
Function $\mathcal{E}(L)$
Output: Ciphertext c
the random $r \in \emptyset^*_{-1}$={$z|z \in Z, 0 \le z < N, gcd(z,n) = 1$};
element the ciphertext by c=$G^L.r^N$mod N^2,where m∈ Z_N= {z|z ∈ $Z, 0 \le z < N$}, $c \in Z_N{}^*2$;
Function D(C)
Output: Message L
Define the integer division function M (μ)=$\frac{\mu-1}{N}$;
Generate the plaintext as L=m ($c^{\lambda} mod\ N^2$). modN;

Following data entry into the BiGRU network, the processed data undergoes full learning, which enhances the accuracy of temporal feature extraction even further. The completely connected layer ultimately produces the final forecasting results.

3.4.1 Convolutional Neural networks

The deep design and integration of convolution processing define convolutional neural networks (CNNs), a subclass of feed-forward neural networks. They usually deal with spatial data loss, inefficiencies, and overfitting. Combining both convolutional and pooling layers in the CNN model framework enables efficient feature-learning and classification activities by automatically obtaining features at different scales and levels.

3.4.2 Gated Recruitment Unit (GRU)

One model that processes sequence data using the deep learning technique is called BIGRU. It is predicated on enhancing GRU with the addition of a bidirectional loop structure, which improves the collection of contextual information in sequence data. Two gated loop units are included in every BiGRU unit; one is used to process sequence data in the direction that is forward, and the other is used to process data in the opposite direction.

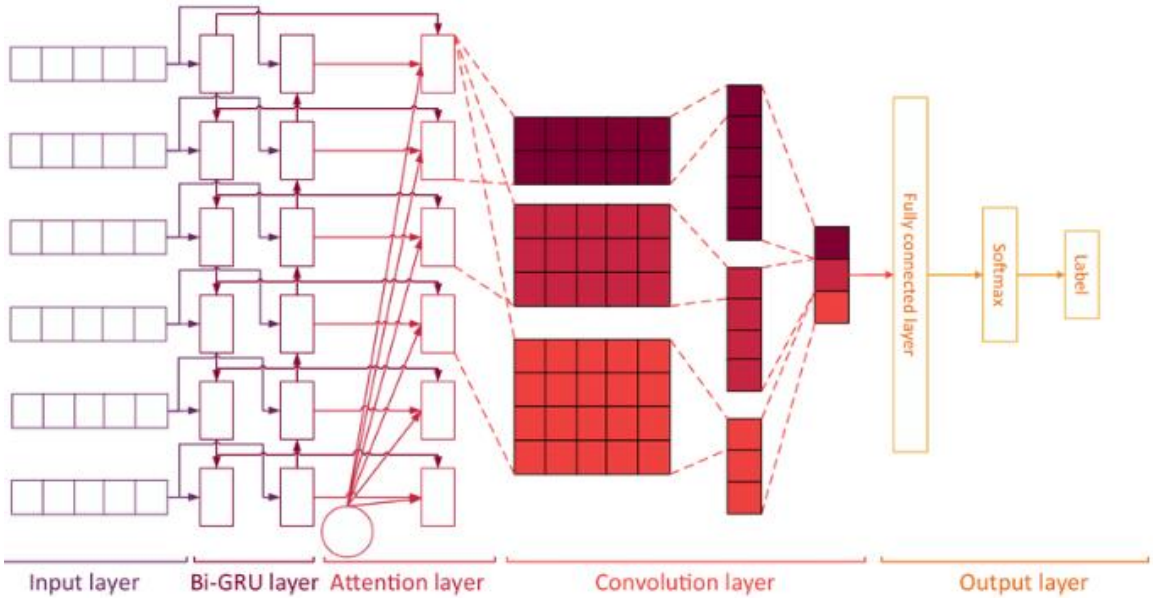

Figure 3. Bi-GRU CNN Neural networks

It lowers the model's training complexity by combining an LSTM's input and remembering gates into just one updating gate complexity and convergence time, as well as its number of parameters, and achieving faster training convergence.

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (1)$$

$$z_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (2)$$

$$\tilde{h_t} = \tanh(W x_t + U(\boldsymbol{r_t} \odot h_{t-1})) \quad (3)$$

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h_t} \quad (4)$$

Here, Wt, Wz, W, Uk, U are the GRU weight matrix values. σ indicates the reasonable sigmoidal function; $tanh$ indicates the tanh function; $\odot$ indicates the element multiplication operation; zt indicates the update gate, which determines the degree of informing of the GRU unit's activation value based on the state of the input and the state of the earlier hidden layer in tandem; rt indicates the rearrange gate, whose informing procedure is comparable to that of zt; The candidate hidden layer I indicated by $h\tilde{t}$, while the hidden layer indicated by ht .

4. RESULT AND DISCUSSION

As shown in the next sections, we tried several experiments in this work to address the privacy issue using deep learning algorithms. By carrying out safe data storage, analysis, and exchange, the model guarantees the security and privacy of the system. Specific criteria, including precision, accuracy, F1-score, and recall, have been used to compare the suggested method to the existing methodologies.

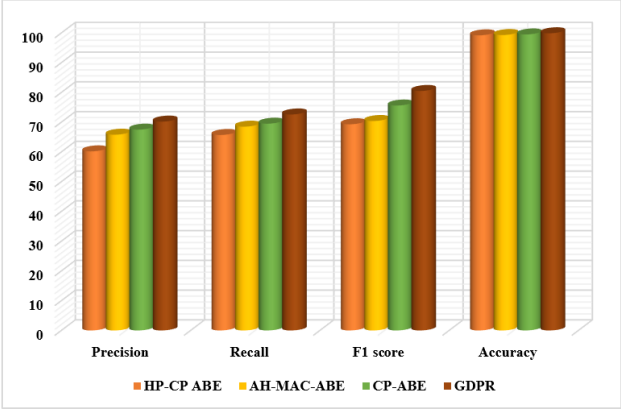

Figure 4. Performance Evaluation

Figure 4 illustrates the findings, which demonstrate that all deep learning methods offer high evaluation metrics in contrast, the averages for benign are 99.92%, 98.85%, and 99.90% for precision, recall, and F1-score.

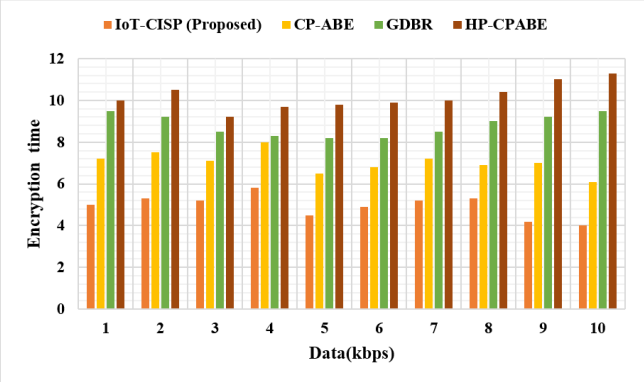

Figure 5. Encryption Time

Figure 5 illustrates the encryption time for the IoT-CISP technique, defined as the duration an encryption algorithm takes to generate ciphertext from plaintext. The encryption time depends on the number of images used in the process.

The proposed method demonstrates faster encryption compared to existing techniques like CP-ABE, GDBR, and HP-CPABE, which require more time. The proposed IoT-CISP technique not only achieves quicker encryption but also performs better with larger datasets. According to the results, IoT-CISP achieves encryption times that are 31.24%, 23.12%, and 33.03% faster than CP-ABE, GDBR, and HP-CPABE, respectively.

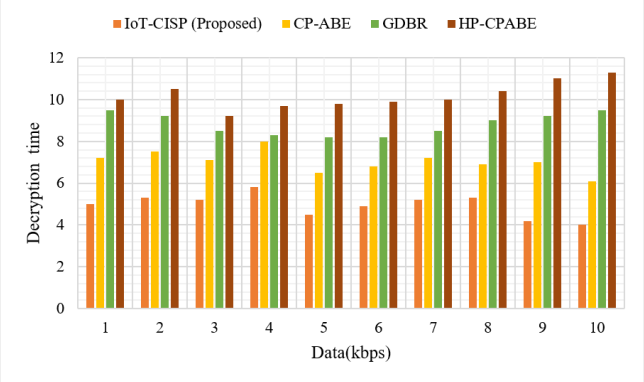

Figure 6. Decryption time

Figure 6 shows the decryption time for the IoT-CISP technique, which refers to the time it takes for the decryption algorithm to convert ciphertext back into plaintext. This decryption time varies based on the number of data involved in the process. The proposed IoT-CISP method achieves faster decryption compared to existing techniques like CP-ABE, GDBR, and HP-CPABE, which require more time. Furthermore, the IoT-CISP technique performs better with larger datasets. The results indicate that IoT-CISP is 21.24%, 13.12%, and 23.03% faster in decryption times compared to CP-ABE, GDBR, and HP-CPABE, respectively.

5. CONCLUSION

In this research a novel Internet of Things based Cloud Information Security Preservation (IoT-CISP) has been proposed. This approach enhances the model's effectiveness and ensures security by first separating sensitive data from non-sensitive data using an SVM classifier, and then employing this data for partial decryption and analysis. Sensitive data is protected through Okamoto-Uchiyama encryption, ensuring that data storage, analysis, and sharing are conducted securely to maintain the system's safety and privacy. The effectiveness of this novel method was assessed against existing methodologies using parameters like precision, accuracy, F1 score, and recall, revealing its superior security and efficiency compared to other schemes. Results demonstrate that the IoT-CISP approach offers encryption times that are 31.24%, 23.12%, and 33.03% shorter than those of the CP-ABE, GDBR, and HP-CPABE algorithms, respectively. In the future, the current model will be tested on additional benchmark datasets as part of our research plans.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING STATEMENT

Not applicable.

ACKNOWLEDGEMENTS

The author would like to express his heartfelt gratitude to the supervisor for his guidance and unwavering support during this research for his guidance and support.

REFERENCES

[1] K. Singh and I. Gupta, "Online information leaker identification scheme for secure data sharing", *Multimedia Tools Appl.*, vol. 79, no. 41, pp. 31165–31182, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] I. Gupta, A. K. Singh, C. N. Lee, and R. Buyya, "Secure data storage and sharing techniques for data protection in cloud environments: A. systematic review, analysis, and future directions", *IEEE Access*, vol. 10, pp. 71247–71277, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] D. Saxena, I. Gupta, A. K. Singh, and C. N. Lee, "A fault-tolerant elastic resource management framework toward high availability of cloud services", *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 3, pp. 3048–3061, Sep. 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] R. Gupta, D. Saxena, I. Gupta, A. Makkar, and A. K. Singh, "Quantum machine learning driven malicious user prediction for cloud network communications", *IEEE Netw. Lett.*, [CrossRef] [Google Scholar] [Publisher Link]

[5] I. Gupta and A. K. Singh, "SELI: Statistical evaluation-based leaker identification stochastic scheme for secure data sharing", *IET Commun.*, vol. 14, pp. 3607–3618, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] IBM Ponemon, "Cost of a data breach study", 2022. [Online] Available: https://www.ibm.com/security/data-breach[CrossRef] [Google Scholar] [Publisher Link]

[7] I. Gupta and A. K. Singh, "Dynamic threshold-based information leaker identification scheme", *Inf. Process. Lett.*, vol. 147, pp. 69–73, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] X.Ma, J.Ma, H. Li, Q. Jiang, and S. Gao, "PDLM: Privacy-preserving deep learning model on cloud with multiple keys", *IEEE Trans. Serv. Comput.*, vol. 14, no. 4, pp. 1251–1263, Jul./Aug. 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] W. Zhao, W. Yang, H. Wang, T. Zhang, D. Man, T. Liu, J. Lv, and M. Guizani, "Privacy-Preserving Outsourcing of K-Means Clustering for Cloud-Device Collaborative Computing in Space-Air-Ground Integrated IoT", *IEEE Internet of Things Journal*, vol. *10*, no. 23, pp.20396-20407, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] S. Fugkeaw, L. Wirz, and L. Hak, "Secure and Lightweight Blockchain-Enabled Access Control for Fog-Assisted IoT Cloud Based Electronic Medical Records Sharing", *IEEE Access*, vol. *11*, pp.62998-63012, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Z. Zhang, W. Zhang, and Z. Qin, "A partially hidden policy CP-ABE scheme against attribute values guessing attacks with online privacy-protective decryption testing in IoT-assisted cloud computing", *Future Generation Computer Systems*, vol. 123, pp. 181-195, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] S., Vaidya, A., Suri, V., Batla, I. Keshta, S.S.M. Ajibade, and G. Safarov, "A computer-aided feature-based encryption

model with concealed access structure for medical Internet of Things", *Decision Analytics Journal*, vol. 7, pp. 100257, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] C. Wang, J. Pan, J. Lu, and Z. Wu, "A Data Hierarchical Encryption Scheme Based on Attribute Hiding under Multiple Authorization Centers", *Electronics*, vol. 13, no. 1, pp. 125, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] M. Li, D. Xiao, H. Huang, and B. Zhang, "Multi-level video quality services and security guarantees based on compressive sensing in sensor-cloud systems", *Journal of Network and Computer Applications*, vol. 205, pp.103456, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] N. Azam, L. Michala, S. Ansari, and N.B. Truong, "Data privacy threat modeling for autonomous systems: A survey from the GDPR's perspective", *IEEE Transactions on Big Data*, vol. 9, no. 2, pp.388-414, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] B. Huang, J. Gao, and X. Li, "Efficient lattice-based revocable attribute-based encryption against decryption key exposure for cloud file sharing", *Journal of Cloud Computing*, vol. 12, no 1, pp.37, 2023. [CrossRef] [Google Scholar] [Publisher Link]

AUTHORS

C. Senthil Singh received his B. E, MTech, PhD degree in Information and Communication from Anna University, Chennai, India. He has a great flair for teaching and research and has a total experience of 22 years in teaching in Engineering. His professional interests include VLSI, Wireless Communication, Embedded Systems and Telemedicine.

Sameena Naaz is currently working as a Senior Lecturer at the University of Roehampton, London, United Kingdom. She holds the post of Professor at the Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India (On leave). She has a total experience of more than 23 years. She received her B.Tech (Computer Engg.) from Aligarh Muslim University, in 1998 and the M.Tech. Degree in Electronics with Specialization in Communication and Information Systems from Aligarh Muslim University, in 2000. She completed her Ph. D from Jamia Hamdard in the field of distributed systems in year 2014. Sameena Naaz has published several research articles in reputed International Journals and Proceedings of reputed international conferences published by IEEE and Springer. Her research interests include Distributed Systems, Cloud Computing, Big Data, Machine Learning, Data Mining and Image Processing.

G Saranya received her B.E degree in Computer Science and Engineering from Anna University, Chennai and M.E degree in Computer Science and Engineering from Hindustan University, Chennai. She started her career as an Assistant Professor and has 9 years and 6 months of experience. Currently she is working as an Assistant Professor in S.A. Engineering College, Chennai. Her research interests include Deep Learning and Cloud Computing. She is a lifetime member of ISTE.

Arrived: 29.04.2024
Accepted: 12.06.2024