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

RESEARCH ARTICLE

DEEP LEARNING BASED LSTM-GAN APPROACH FOR INTRUSION DETECTION IN CLOUD ENVIRONMENT

Rajendran Arulappan Mabel Rose [1, *], Aarthi Gopalakrishnan[2] and J. Vasuki [3]

[1] Department of Computer Science and Business System, Panimalar Engineering College, Thiruverkadu, Tamil Nadu 600077, India.
[2] Department of Computer Science and Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India.
[3] Department of Computer Science and Engineering, Research Point India private limited, Nagercoil, India.

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

Abstract – Cloud computing is a rapidly growing technology paradigm with enormous potential. While cloud computing has many advantages, it also poses new security risks. Cloud computing security vulnerabilities have been identified as the most significant impediment to reaping its many benefits. When sensitive data and business applications are outsourced to a third party, these security concerns become critical. In an existing study, research has found that cloud-based intrusion detection systems (IDS) difficult, more time-consuming, and less secure are major issues. The paper proposes Intrusion Detection Systems and fully homomorphic elliptic curve cryptography (IDS-FHECC). LSTM-GAN is a deep learning algorithm that detects intrusion and non-intrusion data. To reduce computing complexity and improve security, FH-ECC is utilized to encrypt the input data. The encrypted data is processed using homomorphic operations such as implicit additions and multiplications. In the process of evaluating the proposed mechanism, different metrics like as accuracy, encryption, and decryption time are utilized. The proposed improves the overall accuracy by 8.5%, 9.8%, and 12.5% better than LOA, AODV, RLWE, and WOA respectively. The encryption time of the proposed method is 21.8%, 35.7%, and 39.2% decreased compared to existing LOA, AODV, RLWE, and WOA methods.

Keywords *– Cloud computing, Intrusion detection systems, fully homomorphic encryption (FHE), Deep learning.*

1. INTRODUCTION

Cloud computing has quickly become a globally known virtual machine in recent years, with on-demand delivery of apps, platforms, and computing infrastructures likely to play a vital role in the upcoming Internet of Services [1]. Users can keep data and run programmes on top of a virtual computing infrastructure provided by cloud computing [2]. Computing in the cloud allows easy, on-demand access to computer resources [3]. Cloud-based resources can be accessed by organizations on a need-to-know basis. Cloud computing facilitates the analysis of large volumes of data by making use of shared computer resources while being able to easily adapt to changes in data volume and variety.

Cloud service providers (CSPs) provide users with on-demand access to these services while they invest in the infrastructure and computational backbone [4]. Data storage and computing are the two categories of cloud computing services. Cloud service users [5] don't need to know where their data is stored in the cloud or which machines in their network are performing their computations. It is crucial to ensure that cloud data is secure while communicating and uncheckable when building up the cloud architecture. To do this, a wide variety of authentication and encryption methods must be incorporated, which are constantly evolving to deal with new risks and problems.

While encryption protects data at rest, the data can be lost if the encryption key is lost. It is essential to develop cryptographic solutions that can conduct calculations on encrypted data without decryption to prevent malicious attacks on the cloud. Cryptographic algorithms keep communication between individuals, groups, and organizations as private as possible [6]. Cryptographic algorithms play a critical role in data or information security, lowering or eliminating the risk of data leaks [7, 8].

A growing number of factors are contributing to data transmission security in cloud computing. Increasingly, people are concerned about the security of the data they store in the cloud because of frequent hacking incidents. There are several security limitations and difficulties that arise during data transmission in a cloud environment. It is therefore essential to establish a strong intrusion detection system (IDS) to detect and avoid attacks at an early stage. Therefore, intrusion detection systems (IDS) are now an essential part of computer and network security.

The key contribution of this paper is to secure communication in the cloud using Intrusion Detection Systems and fully homomorphic Elliptic Curve Cryptography (IDS-FHECC).

- In this section, Intrusion Detection Systems and fully homomorphic elliptic curve cryptography (IDS-FHECC) have been proposed.
- The proposed system has three phases namely pre-processing, intrusion detection classification, and data encryption.
- Initially, the UNSW-NB15 dataset is pre-processed using tokenization, Repeated Word Removal, and dimension reduction. Intrusion detection is classified using a deep learning framework namely LSTM-GAN. The classified non-intrusion data is encrypted using a Fully Homomorphic-Elliptic Curve Cryptography.
- In the process of evaluating the proposed mechanism, different metrics like as accuracy, encryption, and decryption time are utilized. The proposed improves the overall accuracy by 8.5%, 9.8%, and 12.5% better than LOA, AODV, RLWE, and WOA respectively.

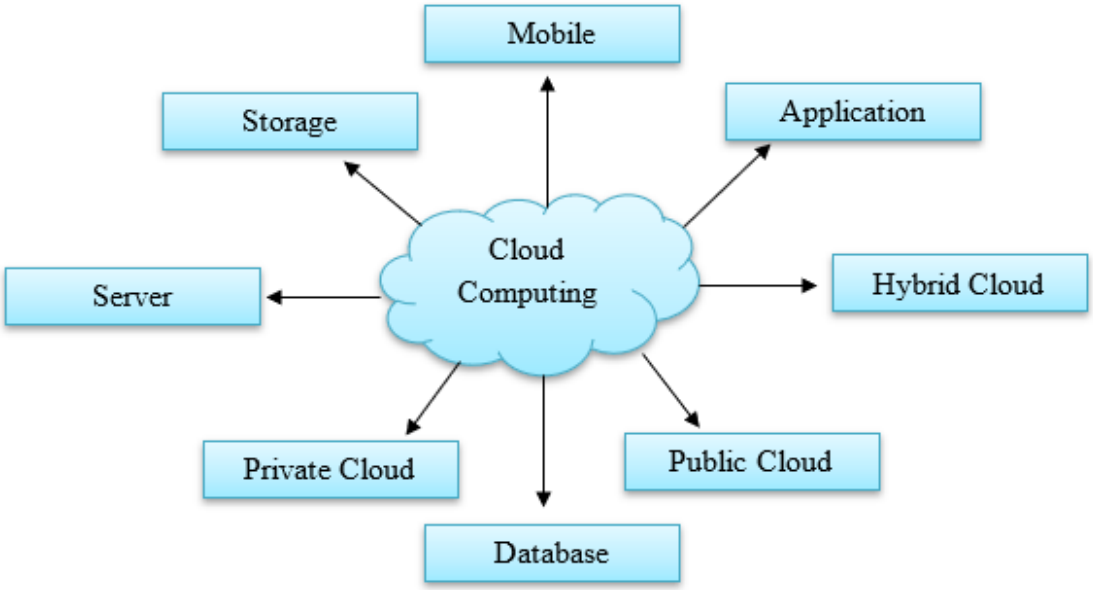

Figure 1. Cloud Computing Models

The research's remaining portion is illustrated in the manner that follows. Section II explains the literature review. Section III included an explanation, the corresponding algorithm, and examples of the suggested techniques. Section IV contains the performance results and analyses. In Section V, conclusions and additional work were finished.

2. LITERATURE SURVEY

In 2023, Kasongo, S.M., [9] A new Intrusion Detection System (IDS) utilizing the Tree-CNN hierarchical algorithm and featuring the Soft-Root-Sign (SRS) activation function has been introduced. The model reduces the training time of the generated model for detecting DDoS, Infiltration, Brute Force, and Web attacks. Additionally, they test whether the proposed model performs well in binary and multiclass classifications by examining the effects of the number of nerve cell and erudition rates. RNN-IDS performs better in binary and multiclass classification than classical machine learning approaches in the experiments, showing that it can model a classification model with high accuracy.

In 2022, Alghamdi, S.A., [10] proposed a novel framework named Trust-Aware Intrusion Detection and Prevention System (TA-IDPS) aimed at safeguarding networks from potential adversaries. Combining symmetric and asymmetric techniques will provide great security, and every user will have an encrypted private key that may be used by multiple people simultaneously. In addition to hybrid cryptography's security the key's security will be increased by this method.

In 2020, Adil M. et al [11] A novel lightweight anonymous authentication technique using MAC-based AODV was introduced to address the identified problem, specifically mitigating the threat of black-hole attacks. It is capable of identifying hostile nodes. In addition, it generates all probable pathways with their trust lengths, and the most trusted path is chosen for routing using AODV as the safest and optimal path. Simulated results show that this tactic is effective for identifying black hole attackers, calculating trust length ratios, and analyzing performance metrics. Consequently, the proposed method is more important than relying on neighbour nodes for trust.

In 2023, Salama. R, et al. [12] presented the SecureLR framework for protecting sensitive patient data by using logistic regression and hybrid cryptographic techniques incorporating secure hardware and homomorphic encryption. By using this technique, logistic regression can be used in the real world to analyze data related to patients, such as health data.

In 2023, Riya K.S, et al [13] proposed safe access in smart city applications, and mobile cloud computing was used. Using an efficient anonymous mutual authentication mechanism the suggested approach enables a mobile user to utilize a unique private key to access a variety of services via distinct service providers. Compared to conventional systems, the suggested technique is more efficient in terms of computational overhead.

In 2023, Patil.S et al. [14] proposed a homomorphic encryption protocol based on RLWE for message management and user authentication. In the proposed encryption protocol, identity values, user information, cloud identification values, and gateway identification values are specified. The investigation utilized a comparison analysis of computation space complexity and time to ensure that the suggested communication protocol delivers robust security and comparable efficiency. Confidentiality and decoding were used in the research.

In 2022, Balashunmugaraja, B, et al [15] proposed Whale Optimization (JWO) is a hybrid of the Whale optimization algorithm (WOA) and algorithm that uses homomorphic encryption to start safe cloud data transmission. A suggested JWO beats the existing techniques with scores of 0.720, 0.822, and 0.722 in well-being, BD, and accuracy, correspondingly.

3. PROPOSED METHODOLOGY

In this section, the Intrusion Detection System and fully homomorphic elliptic curve cryptography (IDS-FHECC) have been proposed. The proposed system has three phases namely pre-processing, intrusion detection classification, and data encryption. Initially, the UNSW-NB15 dataset is pre-processed using data normalization, label encoding, and removing normal traffic. Intrusion detection is classified using a deep learning framework namely LSTM-GAN. Using fully homomorphic-elliptic curve cryptography, the classified non-intrusion data is encrypted. Figure 2 illustrates the general suggested process.

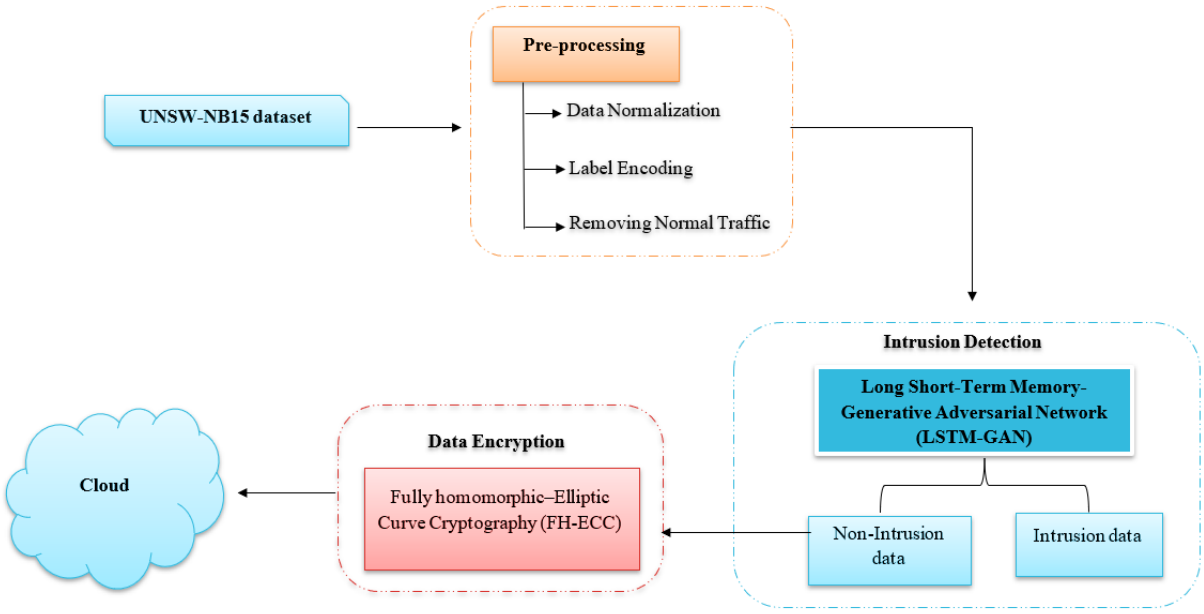

Figure 2. Overall IDS-FHECC Block diagram

3.1. Pre-processing

In this paper, the UNSW-NB15 dataset is used as input data. Different kinds of procedures, including data normalization, label encoding, and removing normal traffic, are performed on the incoming data during the pre-processing step. To achieve data normalization, each attribute's values must be normalized and should have 0 as a minimum number and a maximum number of 1. Attack labels are provided as string values for the multi-class labels in the dataset. Information packets reflecting typical network usage from both datasets are ignored for multiclass classification since they make up the majority of the data. Only attack data streams**.**

3.2. Intrusion Detection using LSTM-GAN

Recurrent neural networks (RNNs) of the artificial variety are similar to the LSTM. This network receives different inputs than conventional recurrent neural networks. It can analyze entire streams of data at a single point. Since there may be unidentified lags between significant events in data from time series, LSTM networks are a good fit for categorization, analysis, and forecasting. Problems with the breakdown and disappearance of gradients have been addressed using LSTMs. This indicates that it ought to take into account every value in the signal and that a memory-based model ought to be employed. Thus, new ECG signals can be created using Long Short-Term Memory GAN (LSTM-GAN). Because of its structure, LSTM attempts to correlate the upcoming values with the past values, remembering previous knowledge. This allows for the processing of data such as frequency variations, amplitude values, and signal rise and fall. Figure 3 below depicts the LSTM-GAN model's architecture, which was used in this investigation.

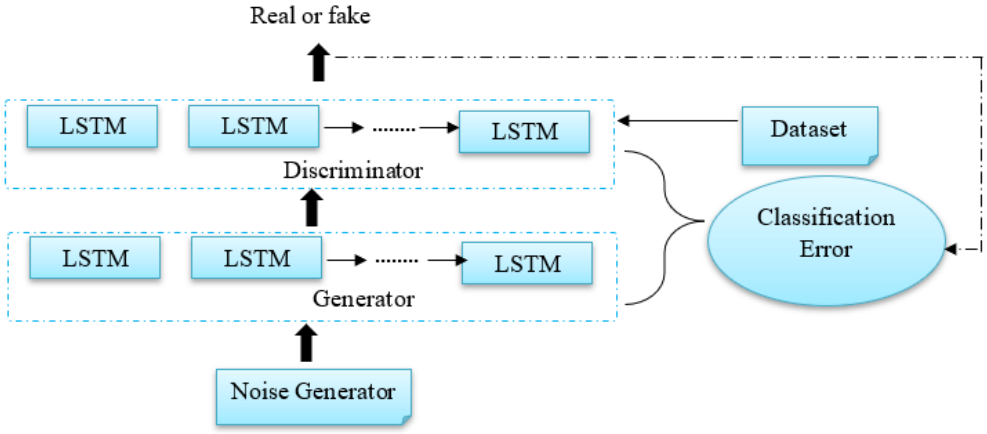

Figure 3. LSTM-GAN Architecture

There are two distinct loss parameters defined for the generator and discriminator, and these functions are employed in tandem.

$$l_d = E\big(d_{(X)}, 1\big) + E\big(d\big(g(z)\big), 0\big) \quad (1)$$

$$l_g = E\big((d\big(g(z)\big), 1\big) \quad (2)$$

$$\mathrm{H}(\mathrm{P}, \mathrm{Q}) = E_{X\sim p(x)}[-logQ(X)] \quad (3)$$

The presence of LSTM modules within the Generator and Discriminator units distinguishes the LSTM-GAN model from the traditional GAN. LSTM-GAN aims to reduce generator loss, increase discriminator loss, and achieve Nash equilibrium, similar to classical GAN.

3.3. Data encryption using FH-ECC:

For secure data storage, a technique called fully homomorphic-elliptic curve cryptography (FH-ECC) is created. The FH is used as an asymmetric method in this study, and the functionalities of the FH are changed by the ECC.

Algorithm for Encryption in FH-ECC:

Input: Private Key (P_r)

Output: Encrypted data (E_d)

1. Create Public key (P_u)
2. Public key =P_r* G_f, where G_f denotes generator function
3. Create the cipher text
4. Cipher text=P_t* G_f
5. Create encrypted data
6. E_d= (P_t*P_u) + (d_j, P_c), where P_c denotes the point of the curve
7. Consider E_d is an encrypted data for FH scheme is applicable
8. E_d = (P_t+P_u) $\sum_{j=1}^{f}$.(d_j, P_c)
9. Implement Homomorphic scheme using multiplication property FHE→F $_{(*/)}$
10. For encrypted data, Calculate 3 coefficient a, b, c
11. F_a = FHE (a1, a2, a3)
12. Calculate F_b = FHE (b1, b2, b3)
13. Create F_c = FHE (d_{Fb}, c)
14. Finally, the encrypted data F_c (E_d) is uploaded to the cloud.

Encryption: It remains the process of turning plain text into the encrypted text to improve security. An often-used cloud security mechanism called the ECC offers protection according to the complexity of specific challenges. One advantage of this technique is that it just depends on the listed table and key. Moreover, it provides improved data solutions and enables safe key exchange across communication devices. Multiplication and addition on encrypted data, as well as arbitrary data calculations, are made possible via FH operations. To safely share data in the cloud, this study uses the FH-ECC technique.

Decryption: Data is decrypted when it is received by the user (e.g., employees or students). The cipher text is extracted from the encrypted data using FH processes, and the cipher text is then converted back to the primary text by decrypting it with the ECC algorithm.

4. RESULT AND DISCUSSION

The suggested IDS-FHECC performance is estimated using several known approaches in this part. The proposed job is carried out with the help of Java-running software and cloud computing. The text documents are obtained experimentally from the NSL-KDD data set, which includes intrusion and non-intrusion data.

4.1. Performance Analysis

The performance evaluation in this study is determined utilizing performance metrics including sensitivity (Sen), accuracy (A), and specificity (Spec), as stated in the equations below.

$$Sensitivity = \frac{T_P}{T_P + F_N} \quad (1)$$

$$Specificity = \frac{T_N}{T_N + F_P} \quad (2)$$

$$Accuracy = \frac{T_P + T_N}{Total\ no.of\ samples} \quad (3)$$

For Equations (1), (2), and (3), TP, TN, FN, and FP, each stand for genuine positives, phony positives, and phony negatives. In a TP result, intrusion documents have been correctly classified, whereas in a TN result, nonintrusive documents have been classified appropriately.

4.2. Comparative Analysis

The efficiency of the suggested technique is demonstrated in the graph below, which compares training and testing accuracy over numerous epochs for intrusion detection. Figure 4 shows how, after 45 epochs, the training accuracy gradually approaches 100%. In 50 epoch, numbers, convergence is faster and more efficient, as shown.

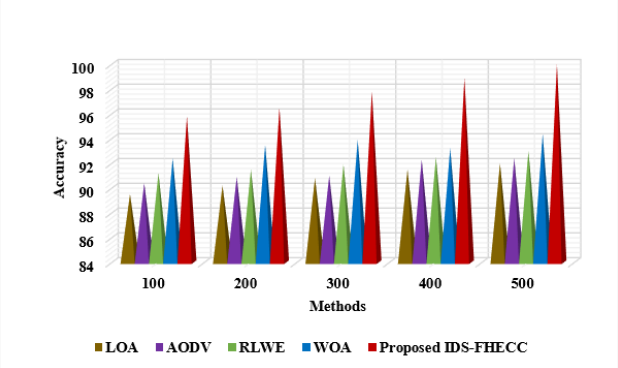

Figure 4. Comparative Analysis of proposed Vs Existing

In terms of accuracy, the presented approach outperforms all other algorithms shown in figure 4. As a consequence, the suggested model achieves accuracy of 99.87 percent.

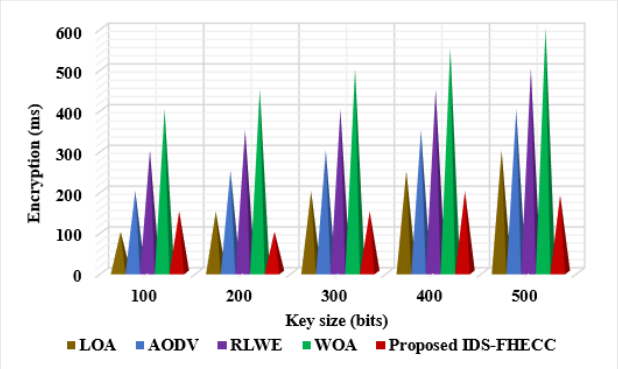

Figure 5. Comparative Analysis of Proposed Vs Existing Methods

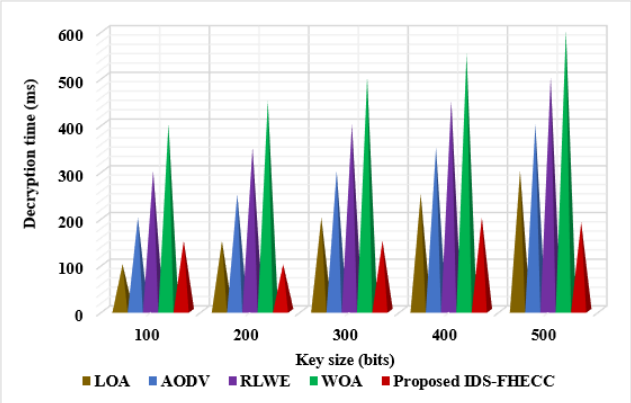

Figure 6. Comparative Analysis of Proposed Vs Existing Methods

Figure 5 shows the encryption times of the existing LOA, AODV, RLWE, WOA, and suggested IDL-FHECC algorithms. An increase in key size (bits) can enhance encryption time, according to this study. The analysis indicates that the IDL-FHECC methodology requires shorter encrypting duration than alternative methods. As a result, homomorphic operations are performed on separated formatted data that was encrypted.

According to Figure 6, the decryption times of existing LOA, AODV, RLWE, WOA, and suggested IDL-FHECC algorithms vary with key size (bits). By increasing the key size linearly, the decryption time can be enhanced.

5. CONCLUSION

The paper proposes Intrusion Detection Systems and fully homomorphic elliptic curve cryptography (IDS-FHECC). LSTM-GAN is a deep learning algorithm that detects intrusion and non-intrusion data. In this section, Intrusion Detection Systems and fully homomorphic elliptic curve cryptography (IDS-FHECC) have been proposed. The proposed system has three phases namely pre-processing, intrusion detection classification, and data encryption. Initially, the UNSW-NB15 dataset is pre-processed using tokenization, Repeated Word Removal, and dimension reduction. Intrusion detection is classified using a deep learning framework namely LSTM-GAN. The classified non-intrusion data is encrypted using a Fully Homomorphic-Elliptic Curve Cryptography. In the process of evaluating the proposed mechanism, different metrics like as accuracy, encryption, and decryption time are utilized. The proposed improves the overall accuracy by 8.5%, 9.8%, and 12.5% better than LOA, AODV, RLWE, and WOA respectively. The encryption time of the proposed method is 21.8%, 35.7%, and 39.2% decreased compared to existing LOA, AODV, RLWE, and WOA methods. The proposed methodology achieves the best results and is extremely safe, decreasing the complexity of encryption and decryption.

CONFLICTS OF INTEREST

This paper has no conflict of interest for publishing.

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] S.A. Bello, L.O. Oyedele, O.O. Akinade, M. Bilal, J.M.D. Delgado, L.A. Akanbi, A.O. Ajayi, and H.A. Owolabi, “Cloud computing in construction industry: Use cases, benefits and challenges”, *Automation in Construction*, vol. 122, pp. 103441, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] B. Alouffi, M. Hasnain, A. Alharbi, W. Alosaimi, H. Alyami, and M. Ayaz, “A systematic literature review on cloud computing security: threats and mitigation strategies”, *IEEE Access*, vol. 9, pp. 57792-57807, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] T. Alam, “Cloud Computing and its role in the Information Technology”, *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 1, no. 2, pp. 108-115, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] R. Maeser, "Analyzing CSP Trustworthiness and Predicting Cloud Service Performance", *IEEE Open Journal of the Computer Society*, vol. 1, pp. 73-85, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] F. Nawaz, M.R. Asadabadi, N.K. Janjua, O.K. Hussain, E. Chang, and M. Saberi, "An MCDM method for cloud service selection using a Markov chain and the best-worst method", *Knowledge-Based Systems*, vol. 159, pp. 120-131, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] R. Abid, C. Iwendi, A.R. Javed, M. Rizwan, Z. Jalil, J.H. Anajemba, and C. Biamba, "An optimised homomorphic CRT-RSA algorithm for secure and efficient communication", *Personal and Ubiquitous Computing*, pp. 1-14, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] S. Tayal, N. Gupta, P. Gupta, D. Goyal, and M. Goyal, "A review paper on network security and cryptography", *Advances in Computational Sciences and Technology*, vol. 10, no. 5, pp. 763-770, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] I. Bhardwaj, A. Kumar, and M. Bansal, "A review on lightweight cryptography algorithms for data security and authentication in IoTs", In *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, pp. 504-509, 2017. IEEE. [CrossRef] [Google Scholar] [Publisher Link]

[9] S.M. Kasongo, "A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework", Computer Communications, vol.199, pp.113-125, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] S.A. Alghamdi, "Novel trust-aware intrusion detection and prevention system for 5G MANET–Cloud", International Journal of Information Security, vol. 21, no. 3, pp. 469-488, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] M. Adil, R. Khan, M.A. Almaiah, M. Al- Zahrani, M. Zakarya, M.S. Amjad, and R. Ahmed, "MAC-AODV based mutual authentication scheme for constraint-oriented networks", Ieee Access, vol. 8, pp. 44459-44469, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] R. Salama, S. Al-Turjman, C. Altrjman, F. Al-Turjman, R.O. Prakash, S.P. Yadav, and S. Vats, "Authentication using Biometric Data from Mobile Cloud Computing in Smart Cities", In 2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE) 445-448 (2023). IEEE. [CrossRef] [Google Scholar] [Publisher Link]

[13] K.S. Riya, R. Surendran, C.A. Tavera Romero, and M.S. Sendil, "Encryption with User Authentication Model for Internet of Medical Things Environment", Intelligent Automation & Soft Computing, vol. 35, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] S. Patil, and R. Patil, Jaya-based, "CAViaR: Hadamard product and key matrix for privacy preservation and data sharing in cloud computing environment", International Journal of Grid and Utility Computing, vol. 14, no. 4, pp. 389-399, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] B. Balashunmugaraja, and T.R. Ganeshbabu, "{Privacy preservation of cloud data in business application enabled by multi-objective red deer-bird swarm algorithm", Knowledge-Based Systems, vol. 236, pp.107748, 2022. [CrossRef] [Google Scholar] [Publisher Link]

AUTHORS

Rajendran Arulappan Mabel Rose received her B.E. degree and M.E. degree in Computer Science and Engineering from Anna University, Chennai, India. She started her career as Lecturer and has 13 years of experience. Currently she is working as Assistant Professor in S.A. Engineering College, Chennai. Her research interests include Wireless Networks and Cloud Computing. She is a lifetime member of MISTE.

Aarthi Gopalakrishnan Completed her B.Tech in Information Technology and M.Tech in Computer and Communication from the Anna University in 2011. At present, she is pursuing Ph.D from B.S.Abdur Rahman Crescent Institute of Science and Technology. She has published/presented several research papers in National/ International Conferences.

J. Vasuki she was born in Kanyakumari District, Tamil Nadu, India in 1999. She received her BE degree in computer science and engineering from Arunachala college of engineering, Manavilai, Anna University, India in 2021. Currently she is a Research and Development Engineer in Research Point India private limited, Nagercoil, India. Her interested research area is image processing, machine learning and deep learning.

Arrived: 15.04.2024
Accepted: 02.06.2024