IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Securing Clouds with Machine Learning: Advancements in Theoretical and Experimental Research

Om Prakash Suman, Lakshya Kumar Saini, Dipender Singh
Department of Computer Science and Engineering
Dr B R Ambedkar National Institute of Technology Jalandhar, Punjab, India

Abstract: *Cloud computing is becoming increasingly prevalent and is being adopted by many businesses to serve their own needs or those of their customers. However, the expansion of the cloud has also resulted in various security concerns for the market and consumers. To address these issues, machine learning (ML) techniques are being employed in a variety of ways to detect and prevent security threats in cloud computing (CC). Here we present a literature review on the application of ML methods to cloud security. We reviewed appropriate works and classified the results into different research topics like various types of cloud security threats, ML-based security mechanisms, and theoretical and experimental analysis of various state-of-the-art papers. Our analysis revealed that Distributed Denial of Service (DDoS) and Confidentiality, Integrity, and Availability of Data are the mutual cloud security domains. Based on our analysis, we found that various machine learning approaches were being used in cloud security research, including stand-alone and hybrid technologies. The article will serve as a stepping stone for new researchers in the area of cloud security*

Keywords: Cloud security, ML technique, cryptographic method, data security, IDS, DDoS attack, hybrid model.

I. INTRODUCTION

Cloud computing is one cutting-edge approach to facilitating and delivering services offered online. Cloud computing is described as a source of virtualized storage. These resources are in an organized state with the goal that they can be shared whenever and can be accessible from any place in the world over the Internet. Due to the CC advantage, the adoption of the cloud computing environment by many organizations has raised vast questions and also posed a variety of cloud security threads. The main security concerns related to cloud computing are: data breaches, insider attacks, denial of service (DoS) attacks, and virtualization security. These threats may originate from conventional techniques like denial of service (DoS), IP spoofing, Address Resolution Protocol. Conventional techniques of detection and prevention are insufficient for dealing with those threats and a high data flow at the same time. These security issues are challenging to address using traditional security mechanisms and require novel approaches that leverage machine learning [1]. Security systems that are based on machine learning can help improve cloud security by finding and stopping different security threats. These mechanisms include intrusion detection systems (IDS), malware detection, anomaly detection, and access control. ML algorithms can learn from historical data to detect security threats accurately and provide timely alerts to security personnel. ML is a collection of algorithms that can analyze patterns in data and make predictions based on those patterns. To improve prediction, ML blends statistics and computer science. ML includes supervised, unsupervised, and reinforcement learning as its three primary learning types [2]. In theoretical analysis, mathematical models are used to look at the security of cloud systems. These models can help identify vulnerabilities and evaluate the effectiveness of security mechanisms. Experimental analysis involves the implementation of ML-based security mechanisms in a cloud environment to evaluate their performance. Security concerns in CC are the main issue we focus on. We describe the methods that are utilized to resolve problems and enhance performance. Also, we go through how different ML techniques are used in CC to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

address security issues. Additionally, we offer research directions that should be researched in the future. Moreover, we outline future research paths that should be investigated in this paper

II. CLOUD COMPUTING

Cloud computing refers to a technology that delivers computing services like storage, servers, software, and applications through the internet. It provides users with on-demand access to these resources without requiring any physical hardware or infrastructure. Unlike traditional methods of storing data and running applications on a local device, cloud computing offers a more flexible and scalable option. This technology allows businesses and individuals to customize their computing resources according to their specific needs, making it an efficient and cost-effective way to manage and process data. Additionally, cloud computing offers users the convenience of accessing their data and applications from anywhere with an internet connection, making it an essential technology for modern-day businesses and organizations [3].

Cloud security thread

Cloud security threats pertain to the possible dangers and weaknesses associated with cloud computing environments. Although cloud computing environments are susceptible to similar security threats as traditional IT environments, they pose distinct risks and challenges due to their nature [4]. Some common cloud security threats include:

- Data breaches: Cloud environments may be vulnerable to unauthorized access by hackers, leading to data breaches and the loss of sensitive information.
- Malware and viruses: Cloud-based systems can be infected by malware and viruses that can cause data loss, corruption, or system failures.
- Denial of Service (DoS) attacks: DoS attacks can cause cloud systems to crash or become unresponsive, causing interruptions in business operations.
- Insider threats: The access to sensitive data by employees or contractors, whether intentional or accidental, can result in data breaches and other security risks.
- Misconfiguration: The wrong settings or security protocols can leave cloud systems vulnerable to attacks, leading to misconfiguration risks.
- Third-party risks: Cloud providers may depend on thirdparty vendors, which may increase security breaches and vulnerabilities.

Attack in cloud

Cloud attacks can occur in various areas, and the following are brief explanations of some common types [5].:

- Network-based attacks: These attacks exploit vulnerabilities in the cloud network infrastructure, like routers, firewalls, and switches. An example of this is the DDoS attack, which floods the cloud network with traffic, making it unavailable to legitimate users.
- Virtual Machine (VM)-based attacks: Attackers may target VMs in virtualized environments to gain unauthorized access to cloud resources. For instance, an attacker may use a VM escape attack to break out of a VM and access the underlying hypervisor or host system.
- Storage-based attacks: These attacks focus on the storage infrastructure of a cloud service, such as databases, file systems, or backup systems. The attacker may attempt to steal or corrupt data stored in the cloud or exploit vulnerabilities in the storage infrastructure to access other parts of the cloud.
- Application-based attacks: These attacks target the cloud applications and services that run on top of the cloud infrastructure. Attackers may use social engineering tactics or exploit vulnerabilities in the code of a cloud application to gain access to sensitive information

III. MACHINE LEARNING

Machine learning pertains to the creation of algorithms and statistical models that empower computer systems to perceive patterns, learn from data, and make predictions without direct programming. The three fundamental categories

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

of machine learning are supervised, unsupervised, and reinforcement learning. There are many uses of machine learning in diverse areas such as healthcare, finance, marketing, and entertainment. It can effectively carry out a variety of tasks like predictive modeling, recognizing images and speech, and processing natural language [6] [7].

ML has three main types of algorithms:

- Supervised learning: This involves teaching a model using labeled data with known outcomes to predict outcomes for new data. In supervised learning, algorithms identify patterns and make predictions based on input features and target values.
- Unsupervised Learning: This involves training a model on unlabeled data, where the desired outcome is unknown, and using it to identify patterns and relationships within the data. Algorithms learn to group or cluster data points based on similarities or differences without specific guidance or direction.
- Reinforcement Learning: This involves training a model through environment interaction and receiving rewards or penalties. The algorithms learn to identify optimal actions for maximum rewards or minimum penalties.

A. Machine learning for cloud security

ML algorithms are commonly used in cloud security to analyze large amounts of data and detect potential threats. Here are a few examples of ML algorithms used for cloud security in different research papers:

”A Hybrid Approach for Cloud Security Based on Machine Learning and Fuzzy Logic” by Alazzawi et al. (2021) : The authors propose a hybrid approach that combines machine learning and fuzzy logic for cloud security. The algorithm first uses machine learning to classify network traffic as normal or anomalous, and then uses fuzzy logic to assign a degree of membership to the anomaly. This approach allows for more nuanced detection of anomalies and can reduce false positives. The authors claim that their approach achieves high accuracy and can adapt to changing network conditions [8].

”A Hybrid Machine Learning Approach for Cloud Security” by Zhang et al. (2020): The authors propose a hybrid ML approach that combines deep learning and
decision trees for cloud security. The algorithm first uses a deep neural network to extract features from network traffic data and then uses a decision tree to classify the data as normal or anomalous. The authors claim that this approach achieves higher accuracy than traditional machine learning algorithms [9].

”Anomaly Detection in CC Networks Using Deep Learning (DL)” by Jiang et al. (2019): The authors propose a DL-based approach for anomaly detection in cloud computing networks. The algorithm consists of two main components: a deep autoencoder network for feature extraction and a SVM for classification. The autoencoder is trained on normal network traffic data and learns to reconstruct the input data. During testing, if the reconstruction error is above a certain threshold, the data is classified as anomalous [10].

”DL-Based Malware Detection in Cloud Environments” by Chen et al. (2019): This paper proposes a DL-based approach for malware detection in cloud environments. The algorithm consists of a convolutional neural network (CNN) that is trained on features extracted from malware samples. During testing, the CNN is applied to network traffic data to detect potential malware. The authors claim that their approach achieves high accuracy and can detect new, previously unseen malware [11].

”A DL Approach for Cloud-Based DDoS Attack Detection and Mitigation” by Duwairi et al. (2019): This paper proposes a DL-based approach for detecting and mitigating DDoS attacks in cloud computing. The algorithm consists of a convolutional neural network (CNN) that is trained on network traffic data. The CNN can classify traffic as normal, suspicious, or malicious and can also generate rules for mitigating attacks in real-time. The authors claim that their approach achieves high accuracy and can handle large-scale DDoS attacks [12].

Overall, machine learning algorithms are a powerful tool for cloud security, but they are not a silver bullet. To be effective, machine learning algorithms must be trained on high-quality data and must be regularly updated to adapt to new threats. Additionally, machine learning algorithms can sometimes be vulnerable to adversarial attacks, where an attacker intentionally manipulates data to evade detection. As such, machine learning should be used as part of a larger,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

more comprehensive security strategy that includes other techniques such as access control, encryption, and intrusion detection.

IV. METHODOLOGY

The methodology section explains how data was collected and analyzed in a research paper. It is important for readers to evaluate the study's validity and for the author to establish their credibility. It ensures the rigor and transparency of the study [13].

Conduct a literature search: This could include searching databases such as IEEE Xplore, ACM Digital Library, and Google Scholar.

Select relevant papers: select relevant papers that address the research question, describe novel ML techniques, compare different ML techniques, evaluate the effectiveness of ML techniques for detecting or preventing cloud security attacks.

Analyze the selected papers: Identify the key ML techniques used for cloud security. This could include analyzing the type of ML techniques used (such as supervised, unsupervised, or semi-supervised learning), the features used for training the ML models, and the performance of the ML models in detecting or preventing cloud security attacks.

Evaluate the strengths and limitations of the selected papers: This could include evaluating the quality of the research methodology used in the papers, the reliability of the data used to train the ML models, and the generalizability of the findings to different cloud security scenarios.

Draw conclusions and make recommendations: This could include identifying areas where further research is needed, highlighting the strengths and limitations of existing ML techniques for cloud security, or recommending specific ML techniques for detecting or preventing specific types of cloud security attacks.

Result: Finally, write the review paper, including an introduction, literature review, analysis of the selected papers, conclusions, and recommendations.

IV. BACKGROUND AND RELATED WORK

Jasem Altamemi et al. [14] propose machine learning techniques to quickly detect and mitigate DDoS attacks in SDN. The study evaluated various methods for detecting DDoS attacks and found that the proposed system using the Decision Tree (DT) algorithm achieved a high accuracy of 99.90 % for detecting DDoS attacks in SDN networks.

Farheen Siddiq et al. [15] discuss the challenges faced by cloud computing systems in terms of security and privacy due to various attacks, especially DDoS attacks. The use of machine learning algorithms with existing IDS is considered a good approach, but each detection method has limitations. Proper understanding and analysis of the motivation behind these attacks is required for effective countermeasures. Safeguarding data in the cloud is also crucial, and security policies and procedures need to be followed. Overall, a generic solution is needed to counter both known and unknown DDoS attacks. Zhang, Yang, et al. [16] propose EnclavePoSt, a new way to ensure continuous data integrity in cloud storage services using Intel SGX's hardware-driven Trusted Execution Environment. This method allows for automatic data integrity checking even when users are offline and provides a flexible storage period, reliable measurement of storage time, and resistance to outsourcing attacks. The paper concludes that EnclavePoSt is more practical than previous methods and provides a secure
solution to data integrity concerns in cloud storage services.

Le, Tung et al. [17] discuss the benefits and concerns of storage-as-a-service (STaaS) and how proof of retrievability

TABLE I: Comparative analysis of various approaches for cloud thread using ML technique in cloud computing.

R.No	Objective	Techniques	Experiment Setup	Limitation
[14]	Its machine learning-based technique quickly identifies DDos attacks to avoid additional harm.	LR, Navie Bayes, Decision Tree	Online environment to validate incoming attacks.	Dataset may not be representing all types of DDos attacks.
[15]	To propose a machine learning-based approach for detecting DDoS	Traffic Analysis, K-means	Cloud computing simulation.	Study was conducted on simulation mat vary in real

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

	attacks in cloud computing.	,SVM		world environment. May not work for all types of attacks.
[16]	To propose a practical and secure Proof of storage time for cloud storage system and provide data integrity.	Tree based verification, Remote Attestation using SGX, privacy.	OpenSGX and Aamzon EC2.	Dependence on honest providers, limited applicability, computational overhead, and privacy concerns.
[17]	To propose an efficient Dynamic Proof of Retrievability (DPoR) scheme for cold storage systems to ensure data integrity and availability over time	Merkle Tree based Proofs, dynamic verification, batch verification.	Prototype implementation on local area network, Amazon Glacier cold storage service .	The evaluation on a prototype imple-mentation is limited to the specific environment and conditions in which the prototype was implemented.
[18]	To propose an Identity based public Auditing schema for cloud storage of IoV data and ensure data integrity and security.	Identity Based Signature (IBS), Dynamic Provable Data Possession (DPDP), Batch Auditing.	CloudSim simulator, Charm cryptography..	Assumes the cloud service provider is honest. Evaluation on a prototype implementation. May not work for all types of attacks.
[19]	To propose a framework called EnclavePDP that utilizes Intel Software Guard Extensions(SGX) to verify data integrity in cloud environments.	Intel SGX ,Merkle Hash Tree.	Prototype implementation on an Intel SGX-enabled machine.	The evaluation was conducted on a prototype implementation and not on a large scale. Also limited to verifying data integrity and does not provide confidentiality or privacy protection.
[20]	To propose a secure file storage solution on cloud using hybrid cryptography algorithm.	Hybrid symmetric and asymmetric key encryption.	AWS, prototype implementation.	The algorithm may not be effective in a physical testbed environment, resulting in a decrease to secure system.
[21]	To propose a method for detecting anomalous misconfigurations in AWS Identity and Access Management (IAM) policies.	Clustering, anomaly de-tection, PCA.	AWS	May not detect all types of attacks because evaluation was conducted on a limited dataset.

PDP by deploying it on a real-world cloud storage platform and running 10 PDP schemes, showing that it eliminates the need for a TPA and introduces reasonable performance overhead.

Jay Prakash et al. [20] discuss the use of cloud computing in various fields, but the major concern is security when storing data on the cloud. The article suggests using a hybrid encryption system, which combines different cryptographic algorithms, such as 3DES, RC6, and AES, to provide high-level security. The key information is securely stored using LSB steganography, and the file is split into three parts and encrypted using different encryption algorithms simultaneously. The article also presents different approaches, such as using RSA and AES algorithms or AES and Blowfish algorithms, for double encryption over data and keys to provide high security. The article concludes by presenting different techniques, such as EFS and separate servers, to ensure the security of the data stored on the cloud.

Thijs van et al. [21] discuss the problem of misconfigured access policies in cloud environments that lead to security incidents and data breaches. The authors propose a novel misconfiguration detection approach for identity and access management policies in AWS based on graph representation. They evaluate their approach using real-world policy data

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

from three enterprise cloud environments and show that it is effective in detecting misconfigurations with slightly lower precision but higher detection rates compared to rule-based systems

V. RESULT

The findings of this study demonstrate the efficiency of several ML algorithms in identifying and mitigating various assaults in a cloud environment. It demonstrates how the use of ML to cloud security may offer a high level of automation and scalability, which is beneficial in expansive and complex networks. It contrasts the proposed approaches to existing state-of-the-art methods and outlines their limitations and drawbacks in Table 1. It also compares how well several ML algorithms, including K-NN, NB, RF, LR, DT and SVM, and the datasets that include NSLKDD, UNSW-NB15, CICIDS2017, and CIC-DDoS2019 have performed in identifying cloud thread using measures like accuracy, precision, recall, and F1-score (see Figure 1). Moreover, it shows that some ML based model are superior than others in terms of spotting various attacks in cloud computing. By integrating techniques with different approaches like rule-based or reputation-based systems, they may be made better.

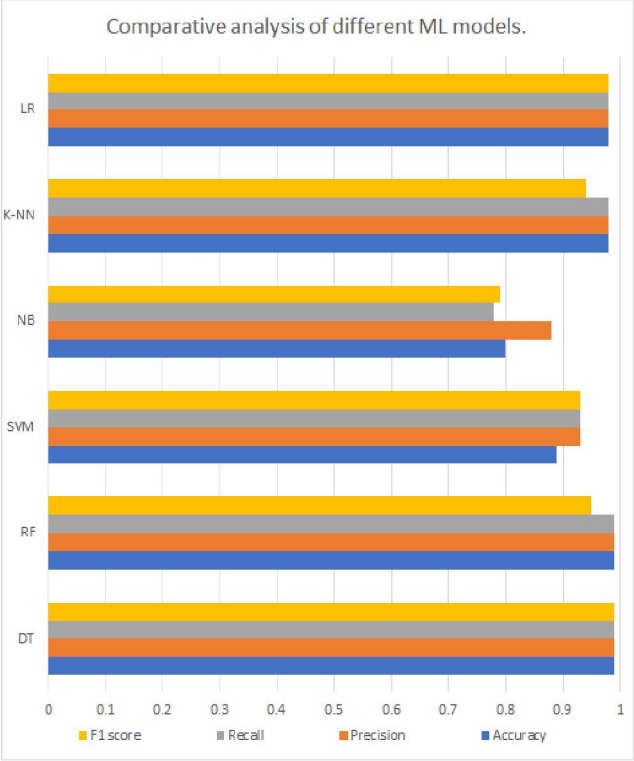

Fig. 1. Comparative analysis of different ML models

Overall, the findings point to the possible of ML approaches for identifying and mitigating attacks in cloud computing, although further work is required to enhance their efficiency and solve their drawbacks.

VI. CONCLUSION

In this paper, different types of cloud attacks are discussed, like storage-based attacks, network-based attacks, and ML techniques, a topic that has received a lot of attention since cloud security is so important. This study examined several studies that used DDoS attacks on cloud computing systems that can be detected using ML approaches. We are also finding that encryption-decryption techniques, access control, data backup techniques, firewalls, and antivirus techniques are used to protect against unauthorized access, malware, theft, and other threats related to cloud storage. The studies suggested several methodologies for reviewing research papers, and the effectiveness of such methods was validated with various datasets. The findings show that such strategies are very accurate and effective in identifying

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

cloud security issues. The findings of this study show the capability of ML approaches to offer a significant level of automation and scalability for network-based attacks in cloud computing. Figure 1. Comparison of the Cybersecurity Strategy of a Firm Using Several ML Models. It is essential to employ datasets that closely reflect real-world circumstances to evaluate the performance of these systems, however getting such datasets can be difficult since businesses are sometimes reluctant to give private and sensitive information. To solve this problem, researchers have built their own datasets using synthetic cyberattacks and carefully monitored test environments.

VII. FUTURE WORK

The study explored potential challenges associated with implementing ML algorithms in authentic environments as well as possible future possibilities for this type of research. Future goals include, to improve outcomes and increase security while keeping efficiency, use a hybrid categorization approach

REFERENCES

[1] Voorsluys, W., Broberg, J. and Buyya, R., 2011. Introduction to cloud computing. Cloud computing: Principles and paradigms, pp.1-41.
[2] Deshmukh, R.V. and Devadkar, K.K., 2015. Understanding DDoS attack its effect in cloud environment. Procedia Computer Science, 49, pp.202210.
[3] Sareen, P., 2013. Cloud computing: types, architecture, applications, concerns, virtualization and role of it governance in cloud. International Journal of Advanced Research in Computer Science and Software Engineering, 3(3).
[4] Singh, A. and Chatterjee, K., 2017. Cloud security issues and challenges: A survey. Journal of Network and Computer Applications, 79, pp.88115.
[5] Kumar, M., Sharma, S.C., Goel, A. and Singh, S.P., 2019. A comprehensive survey for scheduling techniques in cloud computing. Journal of Network and Computer Applications, 143, pp.1-33.
[6] Ray, S., 2019, February. A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.
[7] Salman, T., Bhamare, D., Erbad, A., Jain, R. and Samaka, M., 2017, June. Machine learning for anomaly detection and categorization in multi-cloud environments. In 2017 IEEE 4th international conference on cyber security and cloud computing (CSCloud) (pp. 97-103). IEEE.
[8] Alazzawi, M. F., Jassim, F. A., Al-Rubaye, S. S., Al-Khalidi, M. A. (2021). A Hybrid Approach for Cloud Security Based on Machine Learning and Fuzzy Logic. Journal of Cloud Computing: Advances, Systems and Applications, 10(1), pp. 37.
[9] Zhang, Y., Cao, Y., Wen, S., Li, Z. (2020). A Hybrid Machine Learning Approach for Cloud Security. IEEE Transactions on Industrial Informatics, 16(4), pp.2282-2292.
[10] Jiang, Y., Xue, Y., Chen, X., Wang, Y., Li, L. (2019). Anomaly Detection in CC Networks Using Deep Learning (DL). IEEE Access, 7, pp.106674-106684.
[11] Chen, C., Liu, W., Chen, J., Hu, F. (2019). DL-Based Malware Detection in Cloud Environments. IEEE Access, 7, pp.110741-110748.
[12] Duwairi, R., Alhasanat, M. A., Almashaqbeh, G., Jararweh, Y. (2019). A DL Approach for Cloud-Based DDoS Attack Detection and Mitigation. International Journal of Advanced Computer Science and Applications, 10(4), pp.446-452.
[13] Snyder, H., 2019. Literature review as a research methodology: An overview and guidelines. Journal of business research, 104, pp.333-339.
[14] Altamemi, A.J., Abdulhassan, A. and Obeis, N.T., 2022. DDoS attack detection in software defined networking controller using machine learning techniques. Bulletin of Electrical Engineering and Informatics, 11(5), pp.2836-2844.
[15] Hamdani, F.N. and Siddiqui, F., 2019, May. Detection of DDOS attacks in cloud computing environment. In 2019 International Conference on Intelligent Computing and Control Systems (ICCS) (pp. 83-87). IEEE.
[16] Zhang, Y., You, W., Jia, S., Liu, L., Li, Z. and Qian, W., 2022. EnclavePoSt: A Practical Proof of Storage-Time in Cloud via Intel SGX. Security and Communication Networks, 2022.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

[17] Le, T., Huang, P., Yavuz, A.A., Shi, E. and Hoang, T., 2022. Efficient dynamic proof of retrievability for cold storage. Cryptology ePrint Archive.
[18] Tian, H., Peng, F., Quan, H. and Chang, C.C., 2023. Identity-Based Public Auditing for Cloud Storage of Internet-of-Vehicles Data. ACM Transactions on Internet Technology, 22(4), pp.1-24.
[19] He, Y., Xu, Y., Jia, X., Zhang, S., Liu, P. and Chang, S., 2020, January. EnclavePDP: A General Framework to Verify Data Integrity in Cloud Using Intel SGX. In RAID (pp. 195-208).
[20] Kumar, U. and Prakash, J., 2020. Secure File Storage On Cloud Using Hybrid Cryptography Algorithm.
[21] van Ede, T., Khasuntsev, N., Steen, B. and Continella, A., 2022, November. Detecting Anomalous Misconfigurations in AWS Identity and Access Management Policies. In Proceedings of the 2022 on Cloud Computing Security Workshop (pp. 63-74).