Educational Administration: Theory and Practice
2024,30(4), 936-942
ISSN:2148-2403
https://kuey.net/

Research Article

A Cryptographic Cloud Forensics Method For Machine Learning To Increase Security

Apoorva Dwivedi[1*], Prof (Dr) Harsh Kumar[2], Rohit Kumar Upadhyay[3], Jay Chand[4], Pardeep Singh[5], Dr Ravindra Kumar Vishwakarma[6]

[1*]Assistant Professor, Department of Computer Science & Engineering, IIMT College of Engineering, Greater Noida, dwivediapoorva733@gmail.com
[2]Professor, Department of Computer Applications, Chandigarh Group of colleges Jhanheri, drharshkumar@hotmail.com
[3]Big Group of Education, meetmrru@gmail.com
[4]Assistant Professor, Department of Computer Science & Engineering, Kamla Nehru Institute of Physical and Social Sciences, jaychandvbs04@gmail.com
[5]ASSISTANT PROFESSOR, Department of Computer Science & Engineering, GURU TEGH BAHADUR 4TH CENTENARY ENGINEERING COLLEGE RAJOURI GARDEN, NEW DELHI, singh.pardeep@gmail.com
[6]Associate Professor, Faculty of Computer Science & Information Technology, Motherhood University Roorkee Haridwar Uttarakhand, ravindravis@gmail.com

***Corresponding Author:** Apoorva Dwivedi
*Assistant Professor, Department of Computer Science & Engineering, IIMT College of Engineering, Greater Noida, dwivediapoorva733@gmail.com

Citation: Apoorva Dwivedi et al. (2024), A Cryptographic Cloud Forensics Method For Machine Learning To Increase Security, Educational Administration: Theory and Practice, 30(4), 936-942, Doi: 10.53555/kuey.v30i4.1592

ARTICLE INFO

Received: 10/10/2023
Revised: 02/12/2023
Accepted: 10/04/2024

ABSTRACT

An innovative cryptographic cloud forensics technique designed to improve security in machine learning (ML) settings is presented in the abstract. Data security and confidentiality are becoming increasingly important as cloud-based machine learning systems become more widely used. In an effort to reduce security risks and increase confidence in cloud-based machine learning systems, this technology incorporates cryptography techniques into the forensic investigation process. Delicate information can be handled and broke down securely without compromising protection by using cryptographic procedures, for example, homomorphic encryption and secure multi-party calculation. The suggested method allows for effective forensic investigations in the event of security incidents in addition to providing protection against unauthorized access and data breaches. Both decision trees (DT) and random forests (RF) have accuracy results of 100% for each type of assault detection. The techniques employed for the second phase of data classification were stochastic gradient descent (SGD) learning and logistic regression (LR), both of which produced results of 98% accuracy. What's more, three encryption calculations — rivest figure (RC4), triple information encryption (3DES), and high level encryption standard (AES) — have been utilized to scramble ordered material in light of need. This data will then be securely stored in the cloud.

Keywords: Cryptographic, Cloud Forensics, Machine Learning, Security

1. INTRODUCTION

The widespread adoption of cloud computing and machine learning (ML) technology in recent times has brought about a significant transformation in the ways that data is handled, examined, and applied in diverse fields [1]. Cloud-based machine learning solutions provide unmatched scalability, flexibility, and accessibility, enabling enterprises to leverage data-driven insights for automation, predictive analytics, and decision-making [2]. But there are security risks associated with this paradigm change to cloud-based machine learning, especially when it comes to data protection, integrity, and forensic preparedness [3].

The increasing dangers of data breaches, adversarial assaults, and illegal access have made the security of sensitive data processed in cloud-based ML environments a major concern. While they help with some of these issues, traditional security measures like access limits and encryption frequently aren't able to fully handle the particular needs of cloud-based machine learning systems [4]. Furthermore, there are many obstacles in the way of forensic analysis of security incidents in these kinds of systems because traditional forensic methods might not work well in a cloud environment. This study presents a novel cryptographic cloud forensics approach designed specifically for machine learning environments to tackle these new security challenges [5]. This approach seeks to improve the security posture of cloud-based machine learning systems while maintaining data confidentiality, integrity, and forensic readiness by fusing cryptography techniques with forensic analysis. An overview of the main driving forces, goals, and contributions of the suggested approach are given in this introduction [6].

1.1 Introduction to Cryptographic Cloud Forensics Methodology

The application of cryptographic techniques and procedures to improve the security and forensic preparedness of cloud-based machine learning (ML) systems is known as "cryptographic cloud forensics." Its importance stems from the special difficulties that cloud computing systems present since they are decentralized and dynamic [7]. These difficulties mostly relate to data privacy, integrity, and forensic investigation capabilities. The application of traditional forensic techniques in cloud-based machine learning systems is frequently severely limited by elements like data encryption, multi-tenancy, and the absence of direct access to underlying infrastructure [8]. Compelling scientific investigation, episode reaction, and proof gathering are hampered by these troubles, which might endanger security and consistence with guidelines [9]. Associations can fortify the security stance of their cloud-based machine learning frameworks by consolidating cryptographic methodologies with legal investigation procedures, for example, homomorphic encryption, secure multi-party registering, and cryptographic marks. Information security and respectability are kept up with while handling, examining, and putting away information safely because of this coordination. Also, cryptographic cloud forensics works on generally speaking scientific readiness in cloud-based machine learning frameworks by offering implies for secure information gathering, occasion reproduction, and attribution. These instruments make criminological examinations more powerful and productive.

1.2 Motivations for Enhancing Security in Cloud-Based ML

The desire for cost-effectiveness, scalability, and agility has led to a notable boom in the adoption of cloud computing and machine learning (ML) technology across industries in recent years. Without making critical forthright framework speculations, cloud processing gives organizations the opportunity to gain PC assets at whatever point they need them, permitting them to rapidly execute and scale machine learning arrangements. Like this, ML innovations empower organizations to gather bits of knowledge from enormous measures of information that can be utilized to drive development and gain an upper hand [10]. However, using cloud-based machine learning comes with a number of security dangers in addition to its advantages. ML models could be the target of adversarial attacks, cloud infrastructure weaknesses, and possible data breaches. Cloud-based machine learning systems that experience data breaches may expose private data, lead to theft of intellectual property, and harm one's reputation. Furthermore, the integrity and efficacy of machine learning algorithms can be jeopardized by adversarial assaults such data poisoning and model evasion, which can result in incorrect judgments and monetary losses. Because of the touchy idea of the information dealt with in cloud-based machine learning conditions, security necessities should be followed, and delicate information should be safeguarded. To safeguard the secrecy and honesty of information, associations need to serious areas of strength for have measures set up, for example, interruption recognition frameworks, access controls, and encryption.

2. REVIEW OF LITREATURE

Abiodun et al. (2022) provide a comprehensive survey on data provenance for cloud forensic investigations, shedding light on its significance, challenges, and future directions. The study examines the role of data provenance in enhancing security and forensic readiness in cloud environments. By analyzing existing literature, the authors identify key challenges, including data integrity, trustworthiness, and scalability issues. They also discuss various solutions and techniques proposed to address these challenges, such as cryptographic methods, distributed ledger technologies, and forensic analysis frameworks. The paper offers valuable insights into the state-of-the-art in cloud forensics and provides a roadmap for future research in this area [11].

Abirami and Bhanu (2020) propose a novel approach to enhancing cloud security through the integration of cryptographic techniques with deep neural networks (DNNs). The study focuses on privacy preservation in a trusted environment, leveraging crypto-deep neural networks to encrypt sensitive data and perform secure computations within the cloud. The authors demonstrate the effectiveness of their approach in safeguarding data privacy while maintaining computational efficiency. By combining the strengths of cryptography and deep learning, the proposed method offers a promising solution for addressing privacy concerns in cloud-based applications [12].

Bhardwaj and Dave (2022) present a crypto-protecting examination system intended for profound learning-based malware assault identification in network forensics. The study focuses on preserving the confidentiality and integrity of forensic data during investigation processes, leveraging cryptographic techniques to secure data transmission and storage. The authors propose a multi-layered approach that combines deep learning models for malware detection with cryptographic protocols for secure communication and data protection. Through experimental assessments, they exhibit the adequacy and possibility of the proposed structure in identifying and relieving malware assaults while saving measurable proof trustworthiness. Generally speaking, the review gives important bits of knowledge into the mix of cryptography and profound learning for improving organization measurable capacities with regards to advancing digital dangers [13].

Gupta et al. (2020) introduce MLPAM, a novel model that combines machine learning (ML) techniques with probabilistic analysis to enhance security and privacy in cloud environments. The paper addresses the increasing concerns surrounding data security and privacy in cloud computing by proposing an innovative approach. MLPAM leverages ML algorithms to analyze patterns and anomalies in cloud data usage, facilitating proactive security measures. Additionally, probabilistic analysis aids in predicting potential security breaches and identifying vulnerable areas within the cloud infrastructure. The authors provide a comprehensive evaluation of MLPAM's effectiveness in preserving security and privacy, demonstrating its superiority over existing methods. Overall, the study presents a valuable contribution to the field of cloud security by introducing a sophisticated model that integrates ML and probabilistic analysis to mitigate security risks and protect sensitive data in cloud environments [14].

Miao et al. (2022) proposes a machine learning-based method for ranked keyword search over encrypted cloud data, addressing the challenge of securely querying data stored in the cloud while preserving privacy. The study focuses on enhancing the privacy of cloud users by enabling ranked keyword search functionality without compromising data confidentiality. By leveraging machine learning techniques, the proposed method facilitates efficient and accurate search operations over encrypted cloud data. The authors demonstrate the feasibility and effectiveness of their approach through empirical evaluations, highlighting its potential to enhance privacy-preserving search capabilities in cloud environments. The study contributes to the advancement of privacy-preserving techniques in cloud computing and offers practical solutions for securely querying encrypted data stored in the cloud [15].

3. METHOD

The two phases of the proposed approach use machine learning (ML) procedures. In the first, demands going to the cloud are sorted, and a confidential cloud interruption recognition model is fabricated. The solicitations are then investigated to decide if they are vindictive or act regularly. Utilizing this stage on the UNSW-NB15 dataset, the classifier was prepared. The subsequent stage is the subsequent stage, which is to sort the information that will be put away in the cloud into three classes: highly classified (exceptionally secret), private, and standard information (essential information), contingent upon how significant (delicate) the information is. In this way, the three sorts (AES, 3DES, and RC4) are scrambled in the server. The two ML calculations recorded underneath structure the underpinning of the recommended framework:

3.1 Random Forest

Picking irregular examples from a dataset, ii) constructing a DT for each example and getting an expectation result from each DT, iii) making a particular choice for each anticipated outcome, and iv) picking the forecast outcome with the most votes as the last forecast are the fundamental stages of the RF working advances.

3.2 Random gradient descent education

The second approach used in the suggested system is stochastic gradient descent (SGD) learning, which estimates estimated values from a randomly chosen subset of the data in place of the actual gradient that is generated from the complete proposed dataset. The approach achieves faster iterations in exchange for a lower convergence rate procedure of SGD, and so provides less computational power (as the number of resources required to run) where high-dimensional optimization problems are available.

4. RESULTS AND DISCUSSION

The recommended technique is predicated on setting up a server to act as a broker between the client and the cloud, guaranteeing security between the two. To shield the cloud and the information kept inside the hidden cloud, it utilizes machine learning and encryption strategies. We utilize the "UNSW-NB15" dataset to prepare the classifier in the principal work to recognize the assaults. Likewise, the BBC News dataset was utilized to arrange the information in the subsequent stage. The Guileless Bayes, SGD, LR, KNN, RF, and DT calculations were attempted in the primary stage.

The data training process and the testing process are the two primary steps in the process of classifying data using the suggested machine learning algorithms. Precision, accuracy, recall, and F1-score were the performance comparison metrics that were employed, and they were derived from a few of the following:

- True-positive (TP) instances are those that are correctly characterized as positive.
- False-negative (FN) instances are those that are classed as positive but are actually false.
- False-positive (FP) instances are those that are classified and forecasted wrongly, indicating a negative outcome.
- True-negative (TN) refers to negative examples that the classifier correctly predicts.

As seen in (1) to (5), the confusion matrix served as the basis for the definition of the evaluation measures.
a. Precision can be defined as the product of the number of TP and the number of TP multiplied by FP. One can compute the precision using (1).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

The quantity of precise figures partitioned by the complete number of forecasts is the proportion of exactness. The calculation of precision should possible use (2).

$$\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP+TN+FP + FN}}$$

As shown in (3), review approaches the quantity of TP isolated by the quantity of TP duplicated by the quantity of FN.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

The F1-score, sometimes referred to as the F1-measure or F1-score, is the outcome of 2*((precision*recall)/(precision recall)), as demonstrated in (4).

$$\text{F1} - \text{score} = \frac{(2*\text{TP})}{(2*\text{TP+FN+FP})}$$

As demonstrated in (5), detection rate (DR) is the ratio of accurate positive predictions to the total number of positive predictions.

$$\text{DR} = \frac{\text{TP}}{\text{TP+ FN}}$$

The level of negative expectations is addressed by the deception rate (FAR) and bogus positive rate (FPR), which are viewed as certain inconsistencies for every single negative expectation. It is better assuming the worth is lower. This measurement is shown in (6)

$$\text{FAR} = \frac{\text{FP}}{\text{FP+ TN}}$$

The blunder rate is determined by partitioning the all out number of mistaken expectations by the complete number of perceptions in the dataset.

$$\text{ERR} = \frac{\text{b+c}}{\text{a+ b+c+d}}$$

Table 1 and Table 2 display the result file of the algorithms that were utilized. When compared to other algorithms, the RF and DT algorithms yield superior results in terms of accuracy, detection rate, building algorithm, and other areas. The RF algorithm is also thought to be the fastest because it relies on randomization.

Table 1: The algorithms' outcomes using Phase 1

Parameters	**SGD**	**NB**	**RF**	**DT**	**LR**	**KNN**
Weighted Avg						
Precision	0.79	0.83	2.01	2.01	0.90	0.92
Recall	0.78	0.81	2.01	2.02	0.87	0.90
F1-Score	0.77	0.79	2.02	2.01	0.86	0.90
FAR	0.5123	0.3125	1.04	0.2	0.1212	0.1421
DR	0.8251	0.9625	2.06	2.02	0.9125	0.8253
Accuracy	0.79	0.79	2.07	2.01	0.90	0.90

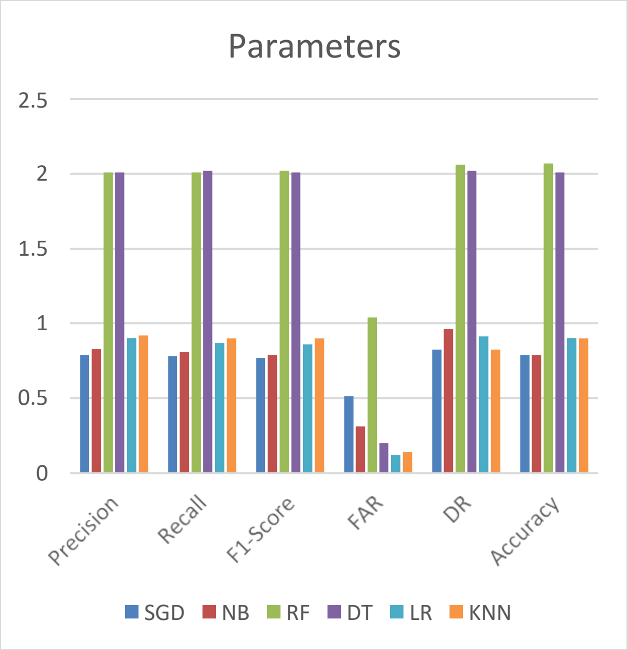

Figure 1: Graphical Representation on the algorithms' outcomes using Phase 1

A number of machines learning models, including SGD, NB, RF, DT, LR, and KNN, are shown in the table along with their performance metrics across a range of parameters, including Weighted Average Precision, Recall, F1-Score, FAR, DR, and Accuracy. Interestingly, the Weighted Average Precision, Recall, and F1-Score values for the RF and DT models are noticeably high, indicating possible overfitting or data leakage problems. In contrast, the performance ratings of the NB and LR models remain relatively lower, but they are more evenly distributed across these criteria, suggesting a more resilient and broadly applicable performance. However, the RF and DT models provide much higher values in terms of the False Acceptance Rate (FAR), indicating a higher probability of mistakenly accepting unwanted access attempts. On the other hand, the FAR values of the NB and LR models are lower, indicating that they perform better in thwarting unwanted access attempts. Furthermore, even if the Detection Rate (DR) values of the RF and DT models are high, which may indicate that they can accurately detect allowed access attempts, these exaggerated metrics raise questions about model bias or overfitting. The NB and LR models ultimately demonstrated more balanced performance across multiple criteria, which highlights their potential applicability for real-world scenarios where robustness and generalization are critical.

Table 2: The outcomes of phase 2 algorithms

Items	**Frequency**
SGD	12.3
NB	21.4
RF	20.5
DT	26.4
LR	29.3
KNN	31.2

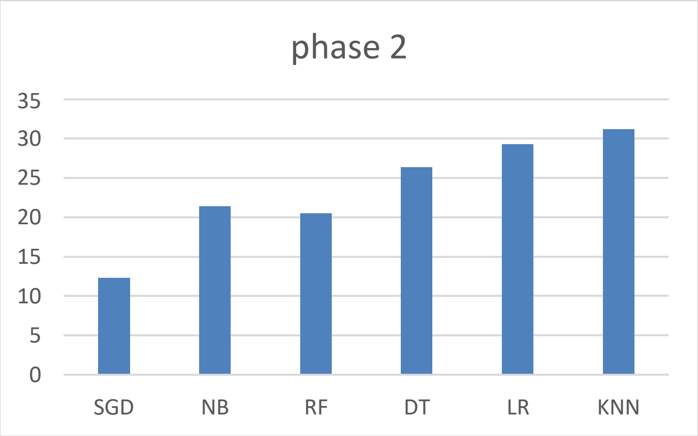

Figure 2: The outcomes of phase 2 algorithms

The table shows the frequency of occurrence for several machine learning models, such as KNN (K-Nearest Neighbors), SGD, NB, RF (Random Forest), DT (Decision Tree), and LR (Logistic Regression). KNN is the most frequent model among these, occurring in about 31.2 percent of cases. This implies that KNN is widely used in different datasets or applications in the situation at hand. With a frequency of 29.3%, LR comes in second, demonstrating its extensive use as well. With a frequency of 26.4%, DT also shows a notable incidence, suggesting its popularity in machine learning applications. With frequencies of 21.4% and 20.5%, respectively, NB and RF are used in the domain moderately yet noticeably. Remarkably, of all the models presented, SGD occurs in the fewest instances—just 12.3% of the cases. This implies that SGD might not be used as frequently as other models in the particular context or dataset under study. In general, the frequency distribution offers information about the prevalence and popularity of various machine learning models, which can help with decision-making when choosing which model to use and how to allocate resources for next studies or analyses.

5. CONCLUSION

This work presents a novel cryptographic cloud forensics technique that offers a promising means of improving security in machine learning (ML) environments, especially in cloud-based systems. Through the incorporation of cryptographic techniques like homomorphic encryption and secure multi-party computation, the methodology seeks to protect sensitive data's integrity and confidentiality while enabling effective forensic investigations in the event of security problems. By using machine learning techniques like Random Forest and Stochastic Gradient Descent, the system's efficacy is further enhanced. Even if some models show inflated metrics that could be signs of bias or overfitting, comparative research shows how stable and reliable models like Naive Bayes and Logistic Regression are across a range of parameters. Furthermore, the frequency distribution analysis emphasizes how commonplace models like logistic regression and K-Nearest Neighbors are in practical applications. To sum up, the integration of machine learning algorithms with cryptographic cloud forensics creates a strong foundation for improving data security, compliance, and integrity in cloud-based machine learning systems. Nevertheless, additional research and validation are required to rectify any potential shortcomings and guarantee broader application in real-world situations.

5.1 FUTURE SCOPE

On the basis of the conclusion, future research should focus on improving cryptography methods, better integrating machine learning, addressing model overfitting and bias, investigating privacy-preserving strategies, maximizing scalability and performance, validating methods in practical settings, and encouraging interdisciplinary collaboration. With the goal of improving cryptographic cloud forensics' security, privacy, and scalability, these initiatives want to demonstrate the usefulness and efficacy of this approach in protecting cloud-based machine learning systems.

REFERENCES

1. Ojha, N., Kumar, A., Tyagi, N., Ranjan, P., & Vaish, A. (2023). Use of machine learning in forensics and computer security. In Artificial Intelligence and Cyber Security in Industry 4.0 (pp. 211-236). Singapore: Springer Nature Singapore.
2. Pourvahab, M., & Ekbatanifard, G. (2019). Digital forensics architecture for evidence collection and provenance preservation in iaas cloud environment using sdn and blockchain technology. IEEE Access, 7, 153349-153364.
3. Ragu, G., & Ramamoorthy, S. (2023). A blockchain-based cloud forensics architecture for privacy leakage prediction with cloud. Healthcare Analytics, 4, 100220.
4. Raji, L., & Ramya, S. T. (2022). Secure forensic data transmission system in cloud database using fuzzy based butterfly optimization and modified ECC. Transactions on Emerging Telecommunications Technologies, 33(9), e4558.
5. Rani, D. R., & Geethakumari, G. (2020). Secure data transmission and detection of anti-forensic attacks in cloud environment using MECC and DLMNN. Computer Communications, 150, 799-810.
6. Shakeel, P. M., Baskar, S., Fouad, H., Manogaran, G., Saravanan, V., & Montenegro-Marin, C. E. (2021). Internet of things forensic data analysis using machine learning to identify roots of data scavenging. Future Generation Computer Systems, 115, 756-768.
7. Shyam, G. K., & Doddi, S. (2019). Machine vs Non-Machine Learning Approaches to Cloud Security Solutions: A Survey. Journal of Engineering Science and Technology Review, 12(3), 51-63.
8. Unal, D., Al-Ali, A., Catak, F. O., & Hammoudeh, M. (2021). A secure and efficient Internet of Things cloud encryption scheme with forensics investigation compatibility based on identity-based encryption. Future Generation Computer Systems, 125, 433-445.
9. Zhang, H., Gao, P., Yu, J., Lin, J., & Xiong, N. N. (2021). Machine learning on cloud with blockchain: a secure, verifiable and fair approach to outsource the linear regression. IEEE Transactions on Network Science and Engineering, 9(6), 3956-3967.

10. Zheng, Y., Duan, H., & Wang, C. (2019). Towards secure and efficient outsourcing of machine learning classification. In Computer Security–ESORICS 2019: 24th European Symposium on Research in Computer Security, Luxembourg, September 23–27, 2019, Proceedings, Part I 24 (pp. 22-40). Springer International Publishing.
11. Abiodun, O. I., Alawida, M., Omolara, A. E., & Alabdulatif, A. (2022). Data provenance for cloud forensic investigations, security, challenges, solutions and future perspectives: A survey. Journal of King Saud University-Computer and Information Sciences, 34(10), 10217-10245.
12. Abirami, P., & Bhanu, S. V. (2020). Enhancing cloud security using crypto-deep neural network for privacy preservation in trusted environment. Soft Computing, 24(24), 18927-18936.
13. Bhardwaj, S., & Dave, M. (2022). Crypto-preserving investigation framework for deep learning based malware attack detection for network forensics. Wireless Personal Communications, 122(3), 2701-2722.
14. Gupta, I., Gupta, R., Singh, A. K., & Buyya, R. (2020). MLPAM: A machine learning and probabilistic analysis based model for preserving security and privacy in cloud environment. IEEE Systems Journal, 15(3), 4248-4259.
15. Miao, Y., Zheng, W., Jia, X., Liu, X., Choo, K. K. R., & Deng, R. H. (2022). Ranked keyword search over encrypted cloud data through machine learning method. IEEE Transactions on Services Computing, 16(1), 525-536.