Data Security on Cloud by Cryptographic Methods Using Machine Learning Techniques

Swetha Gadde[1], J. Amutharaj[2], S. Usha[3]

[1]Research Scholar, Department of Computer Science and Engineering,
Rajarajeswari college of Engineering, Affiliated to VTU, Bengaluru, Karnataka, India
[2]Department of Information Science and Engineering, Rajarajeswari college of Engineering,
Affiliated to VTU, Bengaluru, Karnataka, India
[3]Department of Computer Science and Engineering, Rajarajeswari college of Engineering,
Affiliated to VTU, Bengaluru, Karnataka, India

Abstract
On Cloud, the important data of the user that is protected on remote servers can be accessed via internet. Due to rapid shift in technology nowadays, there is a swift increase in the confidential and pivotal data. This comes up with the requirement of data security of the user's data. Data is of different type and each need discrete degree of conservation. The idea of data security data science permits building the computing procedure more applicable and bright as compared to conventional ones in the estate of data security. Our focus with this paper is to enhance the safety of data on the cloud and also to obliterate the problems associated with the data security. In our suggested plan, some basic solutions of security like cryptographic techniques and authentication are allotted in cloud computing world. This paper put your heads together about how machine learning techniques is used in data security in both offensive and defensive ventures, including analysis on cyber-attacks focused at machine learning techniques. The machine learning technique is based on the Supervised, UnSupervised, Semi-Supervised and Reinforcement Learning. Although numerous research has been done on this topic but in reference with the future scope a lot more investigation is required to be carried out in this field to determine how the data can be secured more firmly on cloud in respect with the Machine Learning Techniques and cryptographic methods.

Keywords:
Cloud Computing, Cryptographic Methods, Encryption, Machine Learning Techniques, Secrecy.

I. INTRODUCTION

Cloud Computing is an emerging effective distributed abode that avail oneself of the plan of distributing, computing capacity, connectivity, storing, and virtualization. Spreading between wide network i.e., Cloud on internet facilitate a huge pool of methods, sharing media and storage media which needs to deliver on-demand facilities. This will assist the end-users to come after the plans of distribution, safety, isolation and flexibility. Security matters are the leading strenuous issue in cloud domain and the indispensable obstacle for elevating of IT companies which give users on-demand facilities. These security matters can be envisioned at network phase, application phase, authentication phase, virtualization phase and authorization phase. The two main causes for the security cover in cloud computing are that these days, mostly everyone store their data on cloud. So, the foremost concern is on the security of user's data and the crucial data should not get meddled while transporting over the network[1], [2].

It is obligatory to guarantee the Integrity, Availability and Confidentiality of user's data. The authenticated user's data is being accessed by the unauthenticated user. To resolve these threats, cryptographic methods can be applied in cloud servers[3], [4]. Although when the user is abrogated, using a particular cryptographic method is not ample to indemnify the safety of data and to run the Access Control techniques in Cloud Computing world. These systems are tried on encryption for data safety. It could be very costly to encrypt the whole data when it comes to time and memory. Therefore, to resolve this issue it would be preferable if we initially divide our sensitive data and then try encryption techniques [5]. It would mark well founded outcomes if we restructure the facts depending upon its secrecy degree. In the area of machine learning, the data categorization is a way of differentiating the group of undivided data illustrative put with the assistance of build classifier [6].

When an instruction set of close data samples is established, a classifier is set up. Large radius of justified instructions data samples are needed to evolve an adequate classifier. This development asks a new prototype of assistance where data categorization to its different clients or users is provided by servers on cloud [7]. Precisely, the data will processed by the server spontaneously and therefore, divides the user's information samples present on distant personal servers. Although the confidential data can be accessed by the unauthorized third party servers. Furthermore, even if the servers give the categorization services to its users, any description or data set identification should not be disclosed. Therefore, a technique that makes sure the secrecy of the server's training set and user's

Manuscript received May 5, 2022
Manuscript revised May 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.5.49

information samples is needed. Thus, a decipher model is indispensable to get ahead of the disapproved client from approaching the enciphered data as well as to generate validate keys for authorized users. Fig. 1 shows the Cloud Computing (CC) service models. The cloud computing is basically an evolving practical dispersed environment which utilizes multifarious concepts of distribution, power processing, storage, connectivity, along with virtualization.

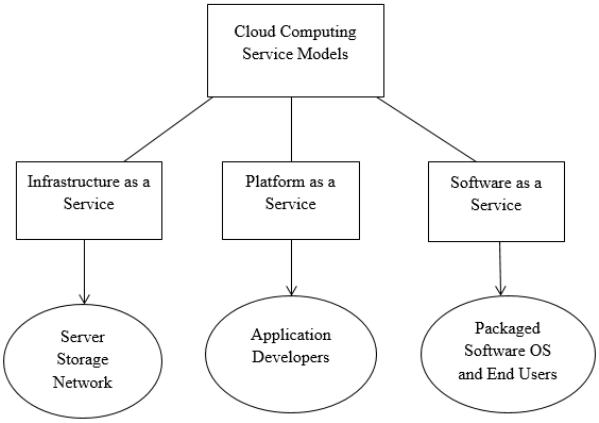

Fig. 1. Illustrates the Cloud Computing (CC) service models.

II. LITERATURE REVIEW

The earlier proposed solutions in machine learning algorithm for improving cloud safety has their own pros and cons as well. Therefore, with this paper two methods has been introduced. One based on the machine learning algorithm and another based on cryptographic method. In this part we will talk about all the survey which we did related to this topic.

P. Chinnasamy et al. carried a research on the topic of "Efficient data security using hybrid cryptography on cloud computing" in their research paper. In this paper, authors have proposed a model that classifies the data according to its security parameters. The performance of the existing KNN is improved by appending it with ensemble learning technique. The basic algorithms of ensemble learning i.e., base level-0 and Meta level-1 are modified. This will improve prediction capability and classification accuracy of existing KNN technique [8].

P. Yang et al. conducted a survey on the topic of "Data Security and Privacy Protection for Cloud Storage" in their study. The main purpose of this paper was to make sure about the safety threat of files stored on cloud using discrete methods of cryptography. In context with this paper author has informed about the symmetric and Asymmetric techniques of cryptographic methods which are popular for encrypting and decrypting of data. In this Data Encryption Standard and Advanced Encryption Standard algorithms has been described in brief. Each and every step of both the algorithms has been put forward in this paper. Another technique which is introduced here is RC-2 Encrypting Algorithm [9].

M. A. Ako et al. carried a research on the topic of "Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data" in their research. In this paper basically they focused on two issues i. e. storage and safeguard of data on cloud. To give safety to the data stored on the cloud, encryption algorithms such as Rivest–Shamir–Adleman (RSA) with Triple Data Encryption Standard (DES) approach has been used. The encrypted information is stored in the database, which is further assigned to the client depending on their priority level achieved using MBFD method. Nevertheless, to improve the secrecy of the designed model of cloud, a categorization approach is used i.e., with Artificial Neural Network (ANN) cross-validation has been done using optimization method (whale) [10].

In 2020, Umer Ahmed Butt in his paper titled, "A Review of Machine Learning Algorithms for Cloud Computing Security" had discussed about Cloud computing safety issues, threats, and solutions that requires one or more Machine Learning techniques. He reviewed distinct Machine Learning techniques that are needed to cope up with the cloud safety matters including supervised, semi-supervised, unsupervised and reinforcement learning. This is one of the most innovative methods as malfunction of data is increasing nowadays we can secure our data by this hybrid approach. In his review paper he had also mentioned about various methods of distinct researchers so that we could get better vision for cryptographic techniques [11]. The authors provides a detailed overview on cloud security.

III. METHODOLOGY

A. Design

The Cloud Computing design includes a start-to-end design that represents each layer of the Open Systems Interconnection (OSI) Model. Fig. 2 shows the cloud computing architecture. Cloud Computing is a complex design with many areas of vulnerability. The basic components of Cloud Computing are such as:

- Cloud Provider: An organization for making, or administrating, available to invested individuals.
- Cloud Consumer: An association that manages relationship, career, and utilizes administrations from the cloud givers.
- Cloud Broker: A substance that organizes the use, implementation, and transportation of cloud profits and manages links between cloud suppliers and cloud customers.
- Cloud Auditor: A gathering that can direct the self-sufficient examination of cloud organizations, information system activities, implementation, and security of cloud users.

• Cloud Carrier: A medium that provides a system to the cloud consumers with cloud administrations from cloud suppliers.

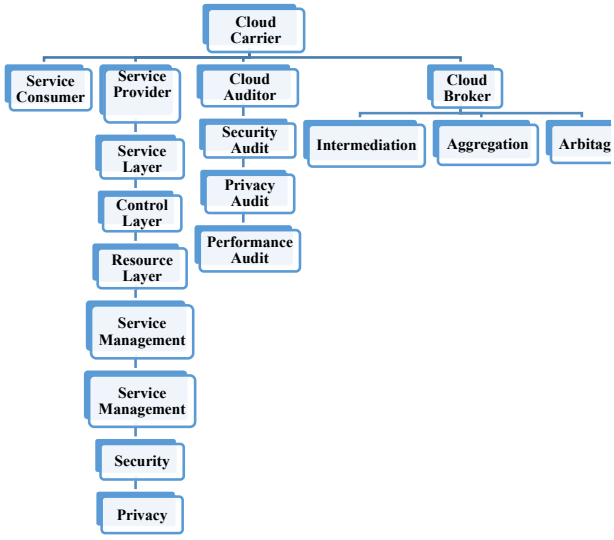

Fig. 2. Illustrates the cloud computing architecture.

B. Instrument

This test work was executed on a system that is composed of the ensuing system configurations mounted with the CloudSim together with the Edge CloudSim simulator; 16 GB of the RAM (Random Access Memory), 64-bit OS (operating system), Windows 10. For many reasons, CloudSim has further been extended by independent researchers. It is basically a framework for modelling and simulation of cloud computing infrastructures and services. It has been one of the most popular open source cloud simulators in academia and research.

C. Data Collection

Cloud Computing has four distribution prototype: private, hybrid, public, and community. Each distribution stereotype has different expenses and worth propositions. Hence, settling the distribution prototype is a strenuous and censorious decision. Fig. 3 demonstrates the cloud distribution prototype.

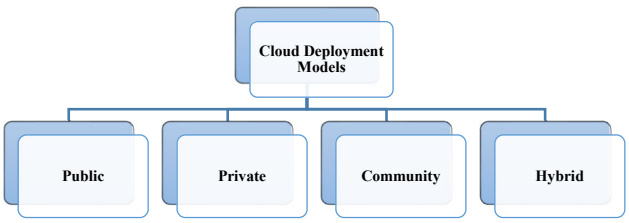

Fig. 3. Illustrates the Cloud deployment models.

D. Data Analysis

The cryptographic algorithm for data security on cloud that we have used is AES (Advanced Encryption Standard). AES is the most trusted algorithm used by the US government as well as other organizations. Though exceptionally indispensable in 128-bit format, it also uses 192-bit and 256-bit keys for demanding encryption purposes. AES-256, which has a key length of 256 bits, supports the largest bit size and is practically unbreakable by brute force based on current computing power, making it the strongest encryption standard. Table I shows the selected key size and possible combinations.

TABLE I
ILLUSTRATES THE SELECTED KEY SIZE AND POSSIBLE COMBINATIONS.

Key Size	Possible Combinations
1 bit	2
2 bits	4
4 bits	16
8 bits	256
16 bits	65536
32 bits	4.2×10^{9}
56 bits (DES)	7.2×10^{16}
64 bits	1.8×10^{19}
128 bits (AES)	3.4×10^{38}
192 bits (AES)	6.2×10^{57}
256 bits (AES)	1.1×10^{77}

Each round in the algorithm consists of four steps.

1. Substitution of the bytes

In the first step, the bytes of the block text are substituted based on the rules dictated by the predefined S-boxes (short for substitution boxes).

2. Shifting the rows

Next comes the permutation step. In this step, all rows except the first are shifted by one.

3. Mixing the columns

In the third step, the Hill cipher is used to jumble up the message more by mixing the block's columns.

4. Adding the round key

In the final step, the message is XORed with the respective round key.

Let's take a look at the order in which these operations execute. It will be as follows-

• Key expansion sets the round key list which is used on each round plus an additional (and initial, as you'll see) round.

• AddRound is the first step to obfuscate the data. Immediately, we have scrambled the data.

• Now we get into the rounds of intense data scrambling. Again, depending on the selected cipher bit, the number of rounds will differ.

• For 9, 11, or 13 rounds, depending on the cipher bit selection, the following will be performed in such order; SubBytes, ShiftRows, MixColumns, AddRound.
• At the 10, 12, or 14 rounds respectively, we perform the final set of operations which are in the order- SubBytes, ShiftRows, and AddRound.

Algorithm:
Step 1: Derive the set of round keys from the cipher key.
Step 2: Initialize the state array with the block data (plain text).
Step 3: Add the initial round key to the starting state array.
Step 4: Perform nine rounds of state manipulation.
Step 5: Perform tenth and final round of state manipulation.
Step 6: Copy the final state array out as the encrypted data (cipher text).

The machine learning technique for data security on cloud that we have used are Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning.

1. Supervised Learning

It is where you have input variables (x) and an output variable (y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y=f(X)$$

The main aim is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (y) for the data.

a) Supervised Neural Network: In this, the information is known. The predicted yield of the neural system is compared with the real yield. Given the mistake, the parameters are changed and afterward addressed the neural system once more. The administered neural system is used in a feed-forward neural system.

(b) K-Nearest Neighbor (K-NN): A basic, easy to-execute administered ML calculation that can be used to solve both characterization and regression issues. A regression issue has a genuine number (a number with a decimal point) as its yield. For instance, it uses the information in the table below to appraise somebody's weight given their height.

(c) Support Vector Machine (SVM): A regulated ML algorithm used for both gathering and relapse challenges. It is generally used in characterization issues. The SVM classifier is a frontier that separates the two classes (hyper-plane).

(d) Naïve Bayes: A regulated ML algorithm that uses Bayes' theorem, which accepts that highlights are factually free. Despite this assumption, it has demonstrated itself to be a classifier with effective outcomes.

2. Unsupervised Learning

It is where you have only input data (X) and no corresponding output variables. The main aim is to model the underlying structure or distribution in the data in order to learn more about the data.

(a) Unsupervised Neural Network: The neural system has no earlier intimation about the yield of the information. The primary occupation of the system is to classify the information based on several similarities. The neural system verifies the connection between diverse source of information and gatherings.

(b) K-Means: One of the easiest and renowned unsupervised ML algorithms. The K-means algorithm perceives k number of centroids, and a short time later generates each data to the closest gathering, while simultaneously maintaining the centroids as little as could be typical considering the present circumstance.

(c) Singular Value Decomposition (SVD): One of the most broadly used unsupervised learning algorithms, at the center of numerous proposals and dimensionality reduction frameworks that are essential to worldwide organizations, such as Google, Netflix, and others.

3. Semi-Supervised Learning

The problems where you have a large amount of data (X) and only some of the data is labelled (Y) are Semi-Supervised learning problems. It comes in between Supervised and Unsupervised Learning. Many real world machine learning problems fall into this area.

4. Reinforcement Learning

Also called as agent, it continuously learns from the environment in an iterative fashion. In this process, the agent learns from its experiences of the environment until it explores the full range of possible states.

IV. RESULTS AND DISCUSSION

In this paper, we have focused on the data security on cloud with the use of several cryptographic methods and machine learning techniques. For this purpose, we have used the Advanced Encryption Standard (AES) as a cryptographic solution and different Machine Learning Techniques have also been brought up. The outcomes of our suggested system has been denoted and talk about in this part. Due to quick shift in technology these days, there is a speedy increase in the private and important data. In contrast with enduring work, the suggested work has been shown. In agreement with the following survey, it is being noticed that the methodology presented in this research paper is providing enhanced and authentic outcomes. Fig. 4 shows the Encryption time in the existing work and the proposed work. Fig. 5 shows the Decryption time in the existing work and the proposed work. For higher standard of secrecy and safety distinct cryptographic methods are needed. The presentation of the suggested system has been assessed by following variables, Accuracy, Encryption Time, Decryption Time and Error Rates.

1) Accuracy: It is the count of accurately classified cases to the total digits of accurately and inaccurately classified cases.

2) Encryption Time: Encrypting data or any information in a manner that only authenticated user should be able to get its access.

3) Decryption Time: It is the way of decrypting the data or information back into its original form.

4) Error Rate: The quantification of the efficacy of transmission channel.

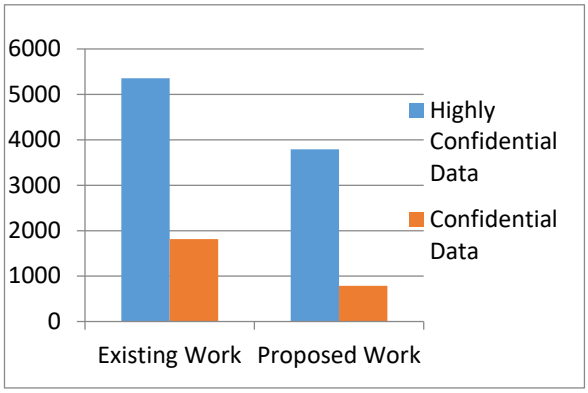

Fig. 4. Illustrates the Encryption time in the existing work and the proposed work.

Fig. 5. Illustrates the Decryption time in the existing work and the proposed work.

Though, a large survey has been done in context of this topic but keeping in mind the future scope a lot more information and research is required to be carried out in this area to detect how the information can be saved more accurately on cloud in respect with the cryptographic methods and Machine Learning Techniques. The need of data security from third party on cloud is a major concern nowadays which requires a lot of efforts needed to be done. Cloud computing and safety is an evolving topic in the field of computing and information technology. Hence as new methods are being developed to improve the safety of the cloud data, the methods to violate the security methods are also emerging. Therefore, a technique that works completely in data safety might not be as effective as provided enhanced and advanced safety threat is emerging in the area of cloud safety. Therefore, there is always a chance and scope for more research to improve the cloud safety. Therefore, more research work should be conducted in these conditions to bring enhancement in the provided algorithms or something completely new should be brought up that is more systematic, more robust than the earlier algorithms.

V. CONCLUSION

In this research, we founded that data security and safety attacks are the most challenging problems in Cloud Computing. Here, we know how machine learning and cryptographic methods can be put in terms of security from both the defensive and attacker's perspective. So it is very clear that machine learning technique and cryptographic methods are powerful weapons that can be handed-down for computerizing composite offensive and defensive cyber-attacks. Therefore, with cyber culprits also grasping machine learning algorithms, we are supposed to undergo more enlightened and big threats powered by Artificial Intelligence. It is hence of essential significance that safety experts as well as machine learning professionals stay alongside with the current development in machine learning techniques. Various types of ML Techniques e.g., ANNs, K-NN, Naïve Bayes, SVM, K-Means, SVD and various Cryptographic methods e.g., AES were introduced as solutions to mark the secrecy issues in Cloud. We assessed many suggested methods used in ML techniques for cloud safeguard. We represented a systematic assessment and survey of the suggested techniques and put a spotlight on their pros and cons. This paper can act as basis for future research that can focus on analyzing existing security solutions and the various challenges of leveraging machine learning to develop and deploy scalable cybersecurity systems in production environments.

REFERENCES

[1] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," 2010, doi: 10.1109/INFCOM.2010.5462174.

[2] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," *IEEE Trans. Parallel Distrib. Syst.*, 2013, doi: 10.1109/TPDS.2012.97.

[3] X. Dong, J. Yu, Y. Luo, Y. Chen, G. Xue, and M. Li, "Achieving an effective, scalable and privacy-preserving data sharing service in cloud computing," *Comput. Secur.*, 2014, doi: 10.1016/j.cose.2013.12.002.

[4] J. Lee, "A view of cloud computing," *Int. J. Networked Distrib. Comput.*, 2013, doi: 10.2991/ijndc.2013.1.1.2.

[5] S. Shilpashree, R. R. Patil, and C. Parvathi, "'Cloud computing an overview,'" *Int. J. Eng. Technol.*, 2018, doi: 10.14419/ijet.v7i4.10904.

[6] P. Mulinka and P. Casas, "Stream-based machine learning for network security and anomaly detection," 2018, doi: 10.1145/3229607.3229612.

[7] N. Taleb and E. A. Mohamed, "Cloud computing trends: A literature review," *Academic Journal of Interdisciplinary Studies*. 2020, doi: 10.36941/ajis-2020-0008.

[8] P. Chinnasamy, S. Padmavathi, R. Swathy, and S. Rakesh, "Efficient data security using hybrid cryptography on cloud computing," 2021, doi: 10.1007/978-981-15-7345-3_46.

[9] P. Yang, N. Xiong, and J. Ren, "Data Security and Privacy Protection for Cloud Storage: A Survey," *IEEE Access*. 2020, doi: 10.1109/ACCESS.2020.3009876.

[10] M. A. Ako, "Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data," *Cryptogr. Netw. Secur.*, 2017.

[11] U. A. Butt *et al.*, "A review of machine learning algorithms for cloud computing security," *Electronics (Switzerland)*. 2020, doi: 10.3390/electronics9091379.