Advanced Machine Learning Techniques for Enhancing Data Security in Cloud Computing Systems

Dr.Ravinder Mogili
Professor and Head
Dept. of Computer Science and Engineering
Jyothishmathi Institute of Technology and Science
(JNTUH)
Karimnagar,Telangana,India
mogili.ravinder@jits.ac.in

Dr.G Srilatha
Associate Professor
Dept. of Computer Science and Engineering
Jyothishmathi Institute of Technology and Science
(JNTUH)
Karimnagar,Telangana,India
gksrilatha8@gmail.com

Sujana Patil
UG Student
Dept. of Computer Science and Engineering
Jyothishmathi Institute Of Technology and Science
(JNTUH)
Karimnagar,Telangana,India
sujanapatil04@gmail.com

Revelli Varshini
UG Student
Dept. of Computer Science and Technology
Jyothishmathi Institute of Technology and Science
(JNTUH)
Karimnagar,Telangana,India
revellivarshini@gmail.com

Pathem Manichandana
UG Student
Dept. of Computer Science and Technology
Jyothishmathi Institute of Technology and Science
(JNTUH)
Karimnagar,Telangana,India
pathemmanichandana@gmail.com

Kodam Rasagnan
UG Student
Dept. of Computer Science and Technology
Jyothishmathi Institute of Technology and Science
(JNTUH)
Karimnagar,Telangana,India
bunnykodam3@gmail.com

Abstract— Cloud computing is one of the most important technologies in today's digital ecosystem. Various organizations around the globe use cloud computing services to store their huge amount of data. Even though cloud computing offers many advantages, including flexibility, scalability, and cost-effectiveness, it is one of the most vulnerable technologies in terms of cyber attacks, including data breaches, insider attacks, malware, and denial-of-service attacks.

Traditional techniques, including firewall protection, cannot be considered effective in detecting new types of attacks. To mitigate these types of attacks, this research proposes a machine learning-based cloud computing security framework, including Random Forest, Deep Neural Network, and Reinforcement Learning (Q-learning).

Experimental outcomes revealed that the proposed model, including a Deep Neural Network, achieved a high accuracy of 97%, compared to Random Forest, which achieved an accuracy of 95%, and Reinforcement Learning, which achieved an accuracy of 88%.

Keywords— Cloud Security, Machine Learning, Deep Learning, Reinforcement Learning, Intrusion Detection.

I. INTRODUCTION

Cloud computing has revolutionized the way businesses handle their data. Rather than using physical servers, businesses are using distributed cloud computing for storing their data.
As cloud computing is becoming more and more popular, the chances of cyber attacks are also growing. Cloud computing is an attractive target for cyber attacks because the cloud stores sensitive information for businesses.
Traditional security systems are based on rules and signatures. Although these are effective for detecting known attacks, they are not effective for detecting unknown attacks.
Machine learning is a more intelligent approach for solving the problem. Machine learning is different from traditional approaches in the sense that machine learning is able to learn from the data. Machine learning is effective in detecting unknown attacks.
In the present research, we are going to use advanced machine learning for improving cloud computing security.

II.LITERATURE SURVEY

Several researchers have worked on improving the security features of the cloud computing environment.
Some of the research work involves the development of blockchain technology-based Public Key Infrastructure (PKI) for improving the authentication and trust management features. The system minimizes the weaknesses associated with traditional hierarchical PKIs.
Other research work involves the development of lightweight encryption schemes using blockchain technology for ensuring the safe retrieval of data and verification of integrity.
In the area of intrusion detection, machine learning algorithms such as Random Forest and Deep Learning have been found to perform well for intrusion detection compared to traditional intrusion detection systems. However, the majority of the work done so far uses only one algorithm and lacks adaptive response mechanisms.
Our work is different from the previous work in that we are using supervised learning and reinforcement learning together in the cloud computing environment.

III.PROBLEM STATEMENT

With the advent of cloud computing, organizations are increasingly exposed to various cybersecurity risks, including:

- Data breaches
- Insider attacks
- Distributed Denial-of-Service (DDoS) attacks
- Malware injections
- Zero-day attacks

Traditional security solutions are static in nature, meaning they cannot change dynamically in response to new attacks. They also tend to generate a high rate of false alarms, failing to detect sophisticated attacks.

There is an urgent need for an intelligent, adaptive, and automated security solution that can:

- Accurately detect attacks
- Minimize false alarms
- React dynamically to attacks
- Be scalable in large-scale cloud computing environments

IV.PROPOSED SYSTEM

The system will incorporate the following major machine learning algorithms:

Threat Detection using Random Forest

Random Forest is a machine learning technique that involves the combination of multiple decision trees. It is especially effective in dealing with high-dimensional data and provides reliable classification results.

The results obtained using the Random Forest technique in the experiment were:

- 95% Accuracy
- 0.92 Precision
- 0.96 Recall
- 0.94 F1 Score

Advanced Threat Analysis using Deep Neural Network

Deep Neural Networks have many hidden layers that help in the analysis of data. They are especially effective in detecting anomalies in data that other techniques might miss.

The results obtained using the DNN technique in the experiment were:

- 97% Accuracy
- 0.94 Precision
- 0.98 Recall
- 0.96 F1 Score

The high value of the recall score indicates that the DNN technique performed exceptionally well in identifying the threats in the data.

Adaptive Threat Response using Q-Learning

Q-learning is a reinforcement learning technique that involves learning the right actions based on the rewards and penalties associated with the actions. It will be used in the system to determine the right course of action when the threats have been identified.

Results:

- 88% Detection Rate
- 0.05 False Positive Rate
- 0.12 False Negative Rate

V.SYSTEM ARCHITECTURE

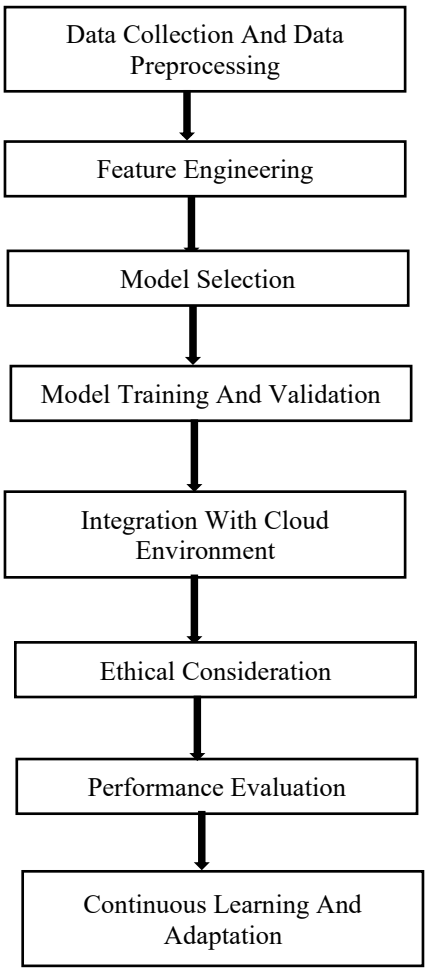

Fig 1

Fig 1 explains about the system architecture of the "Advanced Machine Learning Techniques For Enhancing Data Security In Cloud Computing.

Workflow:

1. Cloud data collection (network logs, user activity logs)
2. Data preprocessing and feature extraction
3. Model training (RF and DNN)

4. Threat classification
5. Adaptive response using Q-learning

VI.METHODOLOGY

1) Data Collection / Dataset Preparation
•Created a dataset containing Safe and Suspicious files.
•Diverse files of various types were included:
oSafe files: PDF, TXT, JPG, DOCX.
oSuspicious files: EXE, Unknown Extension, Script files.

2) Feature Extraction (File-Based)
Three main features are extracted for each uploaded file by the system:
✓File Size.
✓File Type/Extension (pdf, txt, exe, etc.).
✓Entropy (Shannon Entropy).
•Measures the randomness in the file data.
•If entropy is high, it could imply that the malware is compressed/encrypted/obfuscated.

3) Machine Learning Model Training
Supervised learning is used.
Models used:
•Logistic Regression.
•Decision Tree.
•Random Forest (Chosen).
✓Random Forest is chosen because it has:
•High accuracy.
•Less overfitting.
•Good performance with mixed feature sets.

4) Risk Prediction
Features are extracted from an uploaded file.
Model predicts:
•Safe.
•Suspicious.

VII.RESULTS

The proposed system was successfully implemented to improve the security of the data in the cloud computing systems using machine learning techniques. The dataset was successfully collected and processed using various Python libraries such as NumPy, Pandas, and Scikit-learn. Exploratory Data Analysis (EDA) was performed to analyze the data and identify the necessary features that need to be considered for developing the machine learning model. After processing the data, various machine learning algorithms were successfully implemented to train the system for normal and suspicious activity detection in the cloud computing environment.

Fig : 1

Fig : 2

Fig : 3

Fig : 4

The "Admin Accessing User" page helps the administrator view the details of all the registered users in the system. Through this page, the admin can monitor the details of the users, including the username, email ID and location details stored in the database. The administrator can monitor the activity of the users and can ensure that the system is being utilized in the right manner. This helps the admin gain better control over the application and ensures the security and correct functioning of the cloud computing systems

Fig : 5

VIII.CONCLUSION

This study aims to investigate the power of machine learning to secure cloud computing systems. Machine learning was tested in three experiments to identify its advantages and disadvantages. The results showed that the Random Forest model enhanced cloud computing security in Experiment 1. The model correctly classified security threats with 95% accuracy, 0.92 precision, 0.96 recall, and 0.94 F1 Score. The model balanced true and false positives. It is a good model for cloud computing security prevention. In Experiment 2, the researchers tested the deep learning model DNN. The model had outstanding results with 97% accuracy. It even differentiated hazards from permissible actions with 0.94 accuracy, 0.98 recall, and 0.96 F1 Score. For cloud computing data breach security threats, the DNN model's ability to identify complex patterns is a powerful tool. In Experiment 3, the researchers developed the reinforcement learning model Q-Learning for security analysis. The model successfully identified threats with 88% accuracy, balancing true and false positives. However, the model had a high 0.05 false positive rate. The 0.12 false negative rate must be addressed to improve the reinforcement learning model for detecting dangers.

IX.REFERENCES

[1] M. Talamo, F. Arcieri, A. Dimitri, and C. H. Schunck, "A blockchain based PKI validation system based on rare events management," *Futur. Internet*, vol. 12, no. 2, 2020, doi: 10.3390/fi12020040.

[2] H. Du, J. Chen, F. Lin, C. Peng, and D. He, "A Lightweight Blockchain-based Public-Key Authenticated Encryption with Multi-Keyword Search for Cloud Computing," *Secur. Commun. Networks*, vol. 2022, 2022, doi: 10.1155/2022/2309834.

[3] N. E. El-Attar, D. S. El-Morshedy, and W. A. Awad, "A New Hybrid Automated Security Framework to Cloud Storage System," *Cryptography*, vol. 5, no. 4, p. 37, 2021, doi: 10.3390/cryptography5040037.

[4] I. Sudha and R. Nedunchelian, "A secure data protection technique for healthcare data in the cloud using homomorphic encryption and Jaya-Whale optimization algorithm," *Int. J. Model. Simulation, Sci. Comput.*, vol. 10, no. 6, pp. 1–22, 2019, doi: 10.1142/S1793962319500405.

[5] N. Kaaniche and M. Laurent, "Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms," *Comput. Commun.*, vol. 111, pp. 120–141, 2017, doi: 10.1016/j.comcom.2017.07.006.

[6] H. Du, J. Chen, M. Chen, C. Peng, and D. He, "A Lightweight Authenticated Searchable Encryption without Bilinear Pairing for Cloud Computing," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/2336685.

[7] A. N. Jaber and M. F. Bin Zolkipli, "Use of cryptography in cloud computing," *Proc. - 2013 IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2013*, no. May 2016, pp. 179–184, 2013, doi: 10.1109/ICCSCE.2013.6719955.