AI-Driven Cloud Malware Detection and Classification System

[1] DASI JYOSHNA (22K91A0564), [2] CHANAGANI UDAY KIRAN (22K91A0549)

[3] ARRAVOLU PRASHANTH (22K95A0522), [4] DASARI RAKHI (22K91A0563)

GUIDE: Mrs.M. Amani

Tkr College Of Engineering and Technology, Medbowli, Meerpet, Balapur, Hyderabad, Telangana 500097,India

ABSTRACT

The growing number of cyber threats, especially in cloud systems over the last few years has increased the demand for speedy and reliable incident response schemes. This project has implemented an AI-based incident response system to better detect and respond to security incidents in cloud environments. You use machine learning methods with cloud platforms such as the Google Cloud and Microsoft Azure with better efficiency and scalability to your work. Flask is used to develop an automated pipeline, composed of modules for network traffic classification, web intrusion detection and incident-based malware analysis. The system was evaluated by using publicly available datasets such as NSL-KDD, UNSW NB15, and CIC-IDS-2017. As a result, the Random Forest algorithm yielded accuracies of 90%, 75% and 90% for these datasets and moreover also had a precision rate as high as 96% in malware analysis. And also a neural network model then we got an accuracy of 90%. To manage the high processing requirements, cloud-based GPUs and TPUs are utilized, ensuring faster computation. Containerization techniques are also applied to make the system flexible, scalable, and easy to deploy across different cloud platforms. Overall, the proposed system helps in reducing the time required to respond to security incidents, lowers operational risks, and provides a cost-effective solution. This project demonstrates how combining artificial intelligence with cloud infrastructure can improve modern cyber security practices.

Keywords: Cyber incident, digital forensics, artificial intelligence, machine learning, cloud Enviroments,CIC-IDS-2017,NSL-KDD.

INTRODUCTION:

In recent years, cyber attacks have been increasing across different industries, which clearly shows the need for better and faster incident response systems. Many reports from government agencies and cyber security organizations point out that although cyber breaches are common, only a small number of organizations are properly prepared to handle them. These reports also highlight that companies using Artificial Intelligence (AI) and Machine Learning (ML) for threat detection and response are able to reduce the overall cost

and impact of data breaches. This shows how important it is to adopt modern, AI-based security solutions. The system supports multi-cloud environments using tools like Docker and Kubernetes, which makes it flexible and easy to deploy. The network traffic classifier analyzes real-time data to identify suspicious activities. It is trained and tested using well-known datasets such as NSL-KDD, UNSW-NB15, and CIC-IDS-2017, which helps in improving detection accuracy. The Web Intrusion Detection System (WIDS) focuses on identifying unusual behavior in web traffic to prevent unauthorized access. It works by analyzing important features from HTTP server logs.

LITERATURE REVIEW:

In the field of cybersecurity, many research studies have pointed out that traditional security methods are not enough to handle modern cloud-based threats. Earlier systems mostly depended on signature-based or rule-based techniques, which work well for known attacks but fail to detect new or unknown threats like zero-day attacks and complex multi stage intrusions. Because of these limitations, researchers started using machine learning techniques such as Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes for network traffic classification. These methods have shown better performance in terms of accuracy and reduced false alarms, especially when tested on standard datasets like NSL-KDD, UNSW-NB15, and CIC-IDS-2017. In recent years, deep learning methods like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Autoencoders, and LSTMs have also been used to improve detection. These models are capable of learning complex patterns in network traffic, which helps in identifying unusual or suspicious behavior more effectively. Similarly, in Web Intrusion Detection, algorithms like Isolation Forest, One-Class SVM, and autoencoders have been found useful in detecting abnormal web requests and log patterns. To address these challenges, technologies like Docker and Kubernetes are widely used for deploying systems that can scale automatically based on workload. Many researchers also suggest that using AI models along with cloud platforms like Google Cloud and Microsoft Azure improves both performance and efficiency, especially when supported by GPUs and TPUs. However, most of the existing systems still focus on only one aspect, such as network detection or malware analysis, and do not provide a complete solution. Some popular tools also have issues like high false positives, limited

scalability, or dependency on manual work.

PROBLEM DEFINITION:

The rapid adoption of cloud-based storage systems has significantly increased the risk of malware infiltration through user-uploaded files. Most existing cloud platforms lack integrated, real-time malware detection mechanisms, allowing potentially harmful files to be stored and shared without proper inspection. This exposes systems to serious threats such as data breaches, unauthorized access, and system compromise. Furthermore, users are often not provided with sufficient insights into the safety of their uploaded files, and traditional antivirus solutions are not seamlessly integrated into cloud environments, resulting in delayed or ineffective threat detection.

To address these challenges, there is a need for an intelligent cloud security system that can automatically analyze and classify uploaded files using machine learning techniques. The proposed system focuses on real-time malware detection by extracting key file features such as entropy, size, and suspicious patterns, and applying models like Random Forest for accurate classification. Additionally, it provides detailed analytics and instant alerts to users when malicious activity is detected, thereby enhancing transparency, improving security, and ensuring safe cloud storage operations.

PROPOSED SYSTEM:

The proposed system, titled **"AI Driven Cloud Malware Detection and Classification System,"** focuses on addressing the increasing security threats associated with cloud-based platforms. As organizations and individuals rely heavily on cloud storage and services, the likelihood of cyber incidents such as malware injection, unauthorized access, and data breaches has significantly increased. Traditional security mechanisms often fail to provide real-time detection and response, leading to delayed mitigation and potential system damage. Therefore, there is a critical need for an intelligent system that can proactively monitor, detect, and respond to cyber threats within cloud environments.

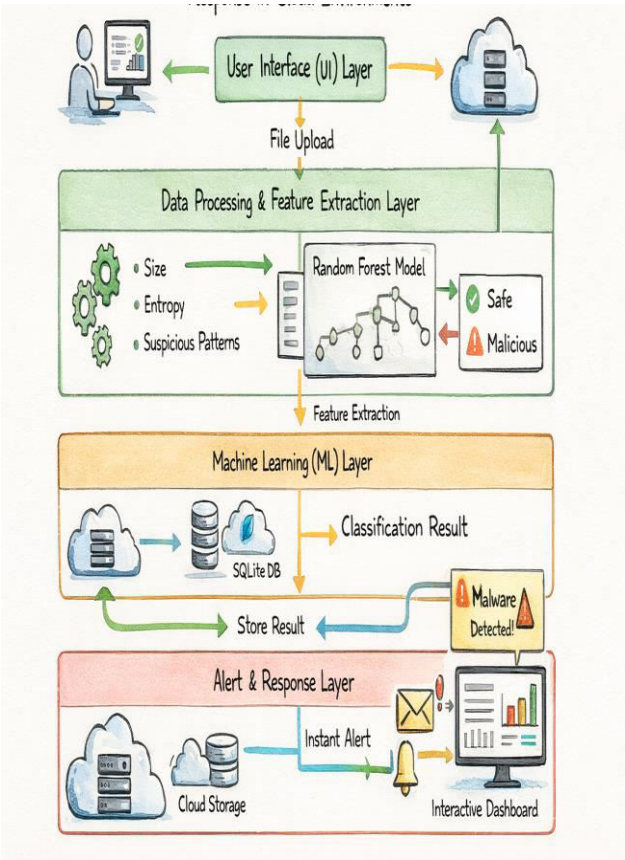
The system leverages artificial intelligence and machine learning techniques to enhance the accuracy and speed of cyber incident detection. By analyzing uploaded files and system activities using features such as entropy, file size, and behavioral patterns, the model—implemented using algorithms like Random Forest—can effectively classify data as safe or malicious. The integration of real-time

analytics and visualization dashboards allows users to monitor file activities and risk levels efficiently. This not only improves threat detection but also provides transparency and better understanding of potential vulnerabilities within the system.

In addition to detection, the proposed system emphasizes rapid response to identified threats. It incorporates automated alert mechanisms that notify users instantly when suspicious or malicious activity is detected, enabling quick preventive actions. The system also maintains structured storage and logging of all activities, supporting traceability and further analysis. Overall, this AI-powered approach ensures a more secure, responsive, and efficient cloud environment by combining intelligent detection, real-time monitoring, and proactive incident response mechanisms.

SYSTEM ARCHITECTURE:

The system architecture of the proposed **AI Driven Cloud Malware Detection and Classification System** is designed as a multi-layered framework that ensures secure file handling and real-time threat analysis. It begins with the **user interface layer**, where users upload files or interact with the system through a dashboard. The uploaded data is passed to the **data processing and feature extraction layer**, which analyzes key attributes such as file size, entropy, and behavioral patterns. These features are then fed into the **machine learning layer**, where a trained Random Forest model classifies the file as safe or malicious. The results are stored in the **database layer** (e.g., SQLite or cloud storage), while the **alert and response module** generates instant notifications for detected threats. Finally, the **visualization layer** presents analytics and reports through an interactive dashboard, enabling users to monitor system activity and respond effectively to cyber incidents in real time.



IMPLEMENTATION:

The implementation of the proposed **AI Driven Cloud Malware Detection and Classification System** is developed as a web-based application that integrates frontend, backend, and machine learning components. The user interacts with the system through a dashboard interface where files can be uploaded for analysis. Once a file is submitted, the backend—implemented using frameworks such as Flask—handles the upload, stores file metadata in a database (e.g., SQLite), and initiates preprocessing. During this stage, important features such as file size, entropy, and suspicious patterns are extracted to form a structured dataset suitable for machine learning evaluation.

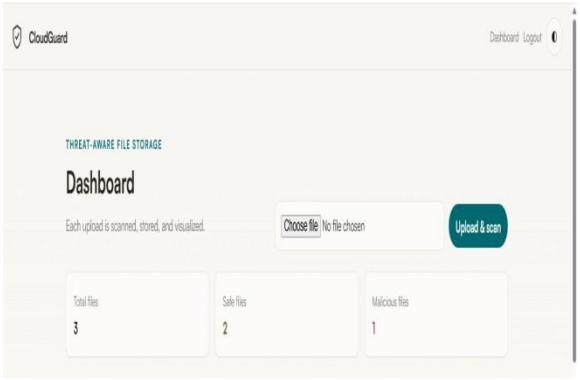
These extracted features are then passed to a trained Random Forest model, which classifies the file as either safe or malicious based on learned patterns. The prediction results, along with calculated metrics like confidence score and risk indicators, are stored and displayed on the dashboard for user interpretation. If a malicious file is detected, the system triggers an automated alert mechanism, such as sending an email notification to the user, ensuring immediate awareness and response. Additionally, the platform provides visual analytics for each uploaded file, enabling users to monitor trends and make informed decisions, thereby enhancing the overall security and efficiency of cloud-based operations.

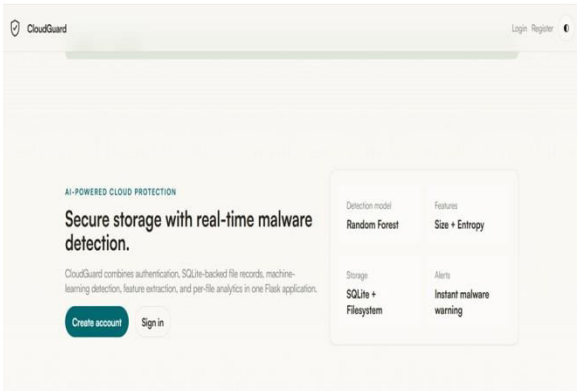
RESULTS AND DISCUSSION

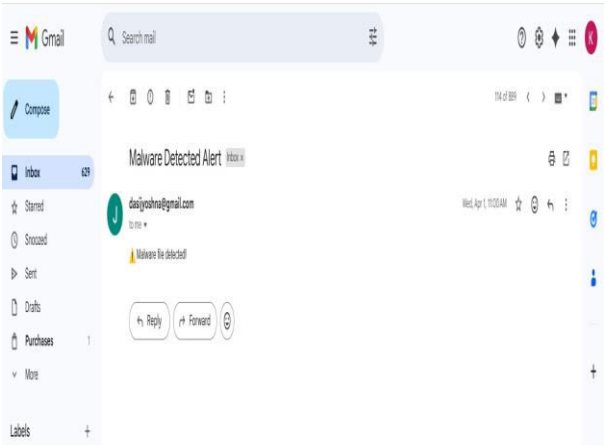
The results of the proposed system demonstrate its ability to effectively classify uploaded files into safe and malicious categories with clear visualization and measurable metrics. From the dashboard, it is evident that the system successfully processed multiple files, identifying the total number of uploads along with a breakdown of safe and malicious files. Each file is analyzed using extracted features such as entropy, external risk, and suspicious ratio, and the results are displayed with corresponding confidence levels. For instance, files labeled as safe show low entropy variation and minimal risk scores, while the malicious file exhibits comparatively higher entropy and suspicious indicators, validating the model's capability to distinguish abnormal patterns.

In addition to classification, the system provides real-time feedback through visual graphs and alert mechanisms. The feature

comparison charts help users understand how different parameters contribute to the final prediction, enhancing transparency in decision-making. Furthermore, when a malicious file is detected, the system triggers an immediate email alert, ensuring that users are promptly informed about potential threats. This combination of accurate prediction, detailed analytics, and instant notification highlights the effectiveness of the system in improving cloud security and enabling quick response to cyber incidents.







CONCLUSION:

In this project, we explored how Artificial Intelligence (AI) can be used in cybersecurity, especially for handling incidents in cloud environments. We designed and developed a complete system that includes a network traffic classifier, a malware analysis module, and a web intrusion detection system. Through this, we were able to understand how AI and machine learning can help in detecting cyber threats more effectively. During the implementation, we deployed our system on cloud platforms like Google Cloud and Microsoft Azure, which helped us

understand how scalable and flexible AI-based security solutions can be in real-world scenarios. We tested our models using standard datasets such as NSL-KDD, CIC-IDS 2017, and UNSW-NB15, along with malware samples. The Random Forest model gave good accuracy in most cases, while deep learning models improved precision, although they required more computational resources. We also used containerization techniques, which made the system easy to deploy and scale. Tools like T-Pot helped in collecting real-time attack data, and the ELK Stack was useful for log analysis and visualization. These tools, along with AI techniques, made the system more effective in monitoring and responding to threats. Overall, this project helped us understand the importance of AI in modern cybersecurity. It shows that using AI and machine learning can improve threat detection and response speed compared to traditional methods. As cyberattacks are becoming more advanced, developing intelligent and automated systems like this will be very important in the future to protect data and systems.

REFERENCE

[1] GOV.U.K., Official Statistics Security Breaches Survey. Accessed: Mar. 4, 2024. Available:https://www.gov.uk/government/statistics/cyber-security-breaches-survey 2024/cyber-security-breaches-survey-2024.

[2] IBM. (2024). Cost of a Data Breach Report. Accessed: Oct. 17, 2024. [Online]. Available: https://www.ibm.com/reports/data-breach.

[3] J. N. Angelis, R. S. Murthy, T. Beaulieu, and J. C. Miller, "Better angry than afraid: The case of post data breach emotions on customer engagement," IEEE Trans. Eng. Manag., vol. 71, pp. 2593–2605, 2022, doi: 10.1109/TEM.2022.3189599.

[4] D. Cotroneo, A. Paudice, and A. Pecchia, "Empirical analysis and validation of security alerts filtering techniques," IEEE Trans. Depend. Secure Comput., vol. 16, no. 5, pp. 856 870, Sep. 2019, doi: 10.1109/TDSC.2017.2714164.

[5] S. Rizvi, M. Scanlon, J. Mcgibney, and J. Sheppard, "Application of artificial intelligence to network forensics: Survey, challenges and future directions," IEEE Access, vol. 10, pp. 110362–110384, 2022,doi:10.1109/ACCESS.2022.3214506

[6] H. Xu et al., "Smart mobility in the cloud: Enabling real-time situational awareness and cyber-physical control through a digital twin for traffic," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 3, pp. 3145–3156, Mar. 2023.

[7] B. Żurkowski and K. Zieliński, "Root cause analysis for cloud-native applications," IEEE Trans. Cloud Comput., vol. 12, no. 1, pp. 232–250, Jan. 2024, doi: 10.1109/tcc.2024.3358823.

[8] A. Pichan, M. Lazarescu, and S. T. Soh, "Cloud forensics: Technical challenges, solutions and comparative analysis," Digit. Invest., vol. 13, pp. 38–57, Jun. 2015, doi: 10.1016/j.diin.2015.03.002.

[9] M. Sonia, Chaganti B. N. Lakshmi, Shaik Jakeer Hussain, M. Lakshmi Swarupa, N. Rajeswaran,

"Android Malware Detection Using Genetic Algorithm Based Optimized Feature Selection and Machine Learning," International-Conference-Computational Intelligence in Machine Learning, Springer, 2022.