Intelligent Malware Detection System Based on Behavior Analysis in Cloud Computing Environment.

Jibu K Samuel
Department of Computer Engineering
Amal Jyothi College Of Engineering
Kerala, India
jibuksamuel2023@cs.ajce.in

Mahima Thankam Jacob
Department of Computer Engineering
Amal Jyothi College Of Engineering
Kerala, India
mahimathankamjacob2023@cs.ajce.in

Melvin Roy
Department of Computer Engineering
Amal Jyothi College Of Engineering
Kerala, India
melvinroy2023@cs.ajce.in

Sayoojya P M
Department of Computer Engineering
Amal Jyothi College Of Engineering
Kerala, India
sayoojyapm2023@cs.ajce.in

Anu Rose Joy
Department of Computer Engineering
Amal Jyothi College Of Engineering
Kerala, India
anurosejoy@amaljyothi.ac.in

***Abstract*—IBMD(Intelligent Behavior-Based Malware Detection) aims to detect and mitigate malicious activities in cloud computing environments by analyzing the behavior of cloud resources, such as virtual machines, containers, and applications.The system uses different machine learning methods like deep learning and artificial neural networks, to analyze the behavior of cloud resources and detect anomalies that may indicate malicious activity. The IBMD system can also monitor and accumulate the data from various resources, such as network traffic and system logs, to provide a comprehensive view of the behavior of cloud resources.**

IBMD is designed to operate in a cloud computing environment, taking advantage of the scalability and flexibility of the cloud to detect malware and respond to security incidents. The system can also be integrated with existing security tools and services, such as firewalls and intrusion detection systems, to provide a comprehensive security solution for cloud computing environments. Overall, the Intelligent Behavior-Based Malware Detection System on a cloud computing environment provides a powerful tool for detecting and mitigating malicious activities in cloud computing environments, helping organizations to secure their critical data and systems.

***Index Terms*—Virtualization, Cloud computing, , behavioral detection, malware detection.**

I. Introduction

Malware detection in cloud computing is the process of identifying and preventing malicious software or code from infiltrating cloud-based systems, applications, and data. With the growing popularity of cloud computing, the risk of malware attacks has become a major concern for organizations that rely on the cloud to store and manage their data. Malware attacks can result in data theft, system disruption, and financial losses. To combat these threats, various malware detection techniques are employed, such as signature-based detection, behavior-based detection, and machine learning-based detection. These techniques analyze and monitor cloud-based systems for any suspicious activities, and upon detection, they take prompt action to prevent any further damage. With the increasing sophistication of malware attacks, the need for robust malware detection in cloud computing has become more important than ever. By implementing effective malware detection systems, organizations can assure the and integrity and security of their data in the cloud.

Intelligent Behavior-Based Malware Detection System on Cloud Computing Environment is a cutting-edge technology designed to detect and prevent malware attacks in cloud computing environments. The system utilizes advanced artificial intelligence and behavior-based techniques to identify and neutralize malicious activities in real-time. With the increasing adoption of cloud computing, the need for robust malware detection systems has become paramount. The intelligent behavior-based approach offers an effective solution to counter the ever-evolving nature of malware attacks. This system provides enhanced security and protection to organizations, their data, and their clients, ensuring that they remain safe and secure in the cloud.

A. Cloud Computing

Cloud computing refers to the delivery of computing services over the internet, such as servers, storage, databases, networking, software, and analytics. Users can access and use these services on-demand and only pay for what they use. Cloud computing allows for scalability, reliability, and cost-effectiveness compared to traditional on-premises computing. Cloud computing offers several benefits, including:

- Scalability: Cloud computing allows users to scale up or down their computing resources according to their needs, making it easier to handle sudden increases in demand.

- Cost-effectiveness: Cloud computing can reduce the costs associated with maintaining on-premises infrastructure, such as hardware, maintenance, and upgrades.
- Reliability: Cloud computing providers typically offer high availability and redundancy, which ensures that services remain accessible and reliable even in the event of hardware or software failures.
- Accessibility: Cloud computing allows users to access their applications and data from anywhere with an internet connection, enabling remote work and collaboration.
- Security: Cloud computing providers typically offer robust security measures, such as encryption and firewalls, to protect users' data and applications.

II. Related Work

Gupta proposed an original approach for detecting malware in cloud environments[1].Pattern-based malware detection is a technique that involves searching for specific signatures or patterns in software code that are associated with known malware. In the context of cloud architecture, this technique can be applied to scan files and data being uploaded to cloud storage or accessed from cloud-based applications, in order to detect and prevent malware infections. The process of pattern-based malware detection typically involves comparing code patterns in the data being scanned with a database of known malware signatures. If a match is found, the system can take actions such as quarantining the file, alerting the user, or deleting the file altogether.Pattern-based malware detection is just one of many techniques used in cloud security. It can be effective in detecting known malware, but may be less effective against newer or more sophisticated threats. As such, it is often used in combination with other security measures such as behavioral analysis, machine learning, and artificial intelligence-based techniques to provide a more comprehensive security solution.

The CloudEyes system, which was developed by Sun , is a cloud-based malware detection [2].The system, called Cloudeyes, aims to address some of the security challenges faced by centralized cloud security monitoring systems, such as data privacy and trust issues. The Cloudeyes system consists of multiple nodes that are distributed across the cloud infrastructure and communicate with each other via a blockchain network. Each node monitors a specific aspect of the cloud infrastructure, such as network traffic or application logs, and generates security events based on predefined rules. These security events are then recorded on the blockchain as transactions.Smart contracts are used to automate the process of validating and responding to security events. The contracts define the conditions under which a security event is considered valid and specify the appropriate response. For example, a contract might specify that if a node detects a certain type of malware, it should quarantine the affected resource and alert the system administrator.The Cloudeyes system also incorporates a reputation system that assigns a reputation score to each node based on its behavior and performance. Nodes with higher reputation scores are given more responsibility in the system, such as being responsible for validating security events generated by other nodes.The paper concludes that Cloudeyes has the potential to provide a more secure and trustworthy cloud security monitoring system than centralized systems, by leveraging blockchain and smart contract technology to address data privacy and trust issues. However, the authors acknowledge that further research is needed to evaluate the performance and scalability of the system.

Xiao et al. proposed Cloud-based malware detection game for mobile devices with offloading [3].A cloud-based malware detection game for mobile devices is a game that tests a player's ability to identify and prevent malware attacks on their mobile devices. The game uses cloud-based resources to perform malware detection and analysis, allowing for real-time feedback on the player's decisions.In addition, the game may also include offloading capabilities, which allows the mobile device to offload resource-intensive tasks to the cloud. This can help to conserve battery life and improve performance on devices with limited processing power. The offloading may involve sending data to the cloud for analysis or processing, then receiving the results back on the mobile device.Overall, a cloud-based malware detection game with offloading can provide a fun and engaging way to educate users about cybersecurity and help them protect their mobile devices from malware and other cyber threats.

The SplitScreen malware detection system was proposed by Cha that enables efficient and distributed malware detection [6].The paper proposes a novel approach to malware detection that utilizes multiple screens to partition the input data and distribute the analysis to multiple machines. The main motivation behind the research is to address the growing problem of malware detection, which is becoming increasingly complex and difficult due to the growing number of malware types and the sophistication of their attack methods. The SplitScreen system divides the input data into smaller partitions, which are then distributed to different machines for processing. Each machine runs a detection algorithm on its partition of the data and sends the results back to the central controller, which combines the results and produces a final verdict on whether the input is malware or not. The system is designed to be highly efficient, with a low processing overhead and the ability to scale to large datasets. The paper presents experimental results that demonstrate the effectiveness of the SplitScreen system. The authors compared the performance of SplitScreen with a single machine running the same malware detection algorithm and found that SplitScreen achieved significantly better results in terms of detection rate and processing time.

Win conducted research on cyber-attacks that target the virtualization infrastructure that underlies cloud computing services. The paper "Detection of malware and kernel-level rootkits in cloud computing environments" discusses the problem of detecting malicious software, specifically malware and kernel-level rootkits, in cloud computing environments. The authors note that these types of attacks are particularly dangerous in cloud computing environments, as they can spread quickly across multiple virtual machines. The paper

proposes a new method for detecting malware and rootkits in cloud computing environments, called the ”Virtual Machine Introspection” (VMI) technique. VMI involves monitoring the behavior of virtual machines from outside of the virtual machine itself, allowing for the detection of malicious activity without relying on the virtual machine’s internal monitoring tools.

Indirapriyadarshini introduced a novel method for detecting malware in the cloud environment using machine learning algorithms. The paper titled ”Malware Detection through Cloud Computing and Machine Learning”” presents a comprehensive approach to malware detection.The authors note that traditional antivirus software is often ineffective at detecting new and evolving forms of malware, and that machine learning techniques can be used to improve malware detection rates. The paper describes the implementation of a machine learning-based malware detection system, which consists of three main components: a data collection module, a feature extraction module, and a classification module. The data collection module collects malware samples from various sources, while the feature extraction module extracts relevant features from these samples. The classification module uses a machine learning algorithm to classify the samples as either malware or benign.

III. PROPOSED SYSTEM

The proposed system for detecting malware using machine learning algorithms and cloud computing consists of several key components. First, users submit suspicious files to the system via a computer network. This submitted files are then executed to get traces in different virtual machines, and the gathering of traces is accomplished with dynamic tools.These traces were obtained by a detection agent that operates on behavior-based principles .

The process involves collecting traces, which are then used to generate behaviors. These behaviors are categorized based on predetermined rules to create features, and the most significant features are chosen using suggested algorithms before being transmitted to the detection agents. The detection agents consist of both learning-based and rule-based approaches, with the former using machine learning algorithms to train the chosen features while the latter evaluates features based on pre-defined feature sets. Following analysis by the detection agents, the samples are classified as either malicious or benign and saved in the database. Finally, the user is informed of the analysis results, indicating whether the submitted file is malicious or not.

This proposed system provides a comprehensive approach to detecting malware using behavior-based detection and feature extraction, allowing for the detection of new and unknown forms of malware. By using machine learning algorithms and cloud computing, the system can improve the overall security of computer systems and networks.

A. Proposed Architecture Model

The system architecture for the proposed model is depicted in Figure 1.

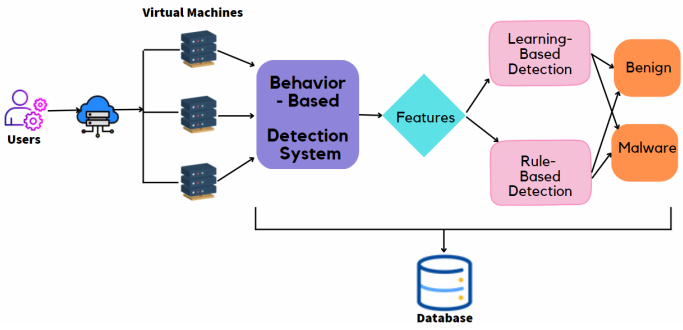

Fig. 1. Proposed malware detection in cloud environment

B. Behavior Formation, Feature Identification, and Selection

The manual analysis of malware and feature extraction is a labor-intensive process that demands a significant amount of time and effort. As a result, there is a pressing need for an automated system capable of automatically analyzing malware and extracting features. However, malware often conceals its actual behaviors by performing both related and unrelated actions, underscoring the importance of identifying the interrelated actions and extracting genuine features when generating a dataset.

To address this issue, the proposed CBCM model efficiently creates and selects features. The model is engineered to avoid generating excessive unrelated features that may increase detection time while lowering detection rates.Dynamic analysis tools such as Process Monitor, Debuggers, API Monitor, Auto-runs and Process Explorer are employed to examine the executable file in this approach. The analysis is carried out on multiple virtual machines (VMs), and the resulting execution traces are gathered and transmitted to the behavior-detection agent. The CBCM model is utilized by the behavior-detection agent to generate behaviors and features based on the collected traces.. The model intertwines behavior formation, feature identification, and selection to ensure that only the most pertinent features are selected for classification.

Overall, the proposed CBCM model offers a comprehensive strategy for malware feature extraction and selection, effectively dealing with the challenges of interrelated and unrelated actions. The application of this model can improve the effectiveness and precision of identifying malware in cloud computing systems.

C. Learning based detection

The proposed feature selection algorithms are used to identify the most discriminative features, and the resulting row vector representation for each program sample includes the frequency values of each feature property. Whenever a property is repeated, its value is the number of times it appears, and any non-repeated properties are assigned a value of 0. This process results in a feature vector representation for each program sample in the dataset.Following this, learning algorithms are employed to train the system.

The proposed malware detection agent that utilizes machine learning algorithms such as Random Forest, LMT, C4.5, SLR, SMO, and KNN for training on selected dataset features uses both cross-validation and holdout methods during the training phase to evaluate the performance of the classifiers. Decision trees, particularly C4.5, RF and LMT, were chosen for training and testing because they provide accurate and scalable results in the cloud environment and are well-suited for the dataset's feature distribution to distinguish between malware and benign software. The C4.5 algorithm selects the most appropriate feature for decision tree placement based on gain ratio and recursively selects the feature with the highest gain as the splitting criterion.

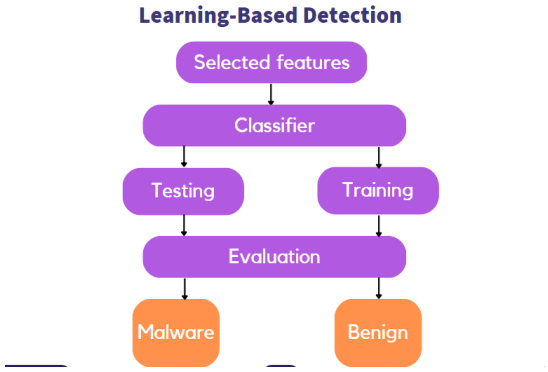

Fig. 2. Learning Based Detection System

D. Rule-based detection agent

The proposed system includes a rule-based behavior malware detection agent that operates on various cloud machines. Unlike the learning-based detection agent, which requires a training phase, the rule-based detection agent detects malware based on a pre-defined property list. This list is generated from malware behaviors and contains features that distinguish malware from benign samples by identifying malicious behavior patterns.New malware features are added to the list as they are found, keeping it constantly up-to-date.
After the feature creation and selection process is done, the feature values are classified into four groups according to their occurrence frequency: low, medium, high, and very high. If the program's characteristics match those on the list, it is considered to be malicious software. This rule-based approach eliminates the need for training data and can rapidly identify known malware based on pre-defined patterns.

IV. RESULT AND DISCUSSIONS

Various sources were used to collect malware samples, including Das Malwerk, MalwareBazaar, Malware DB, Malware Benchmark,Malshare,Tekdefense, ViruSign, VirusShare, and KernelMode.The execution comprises of three tasks.

A. Task 1:Model Building

First step is to extract,vectorize,and preprocess the data for the task.Next,a deep neural network architecture is designed in Keras.The model is trained,validated,and tested

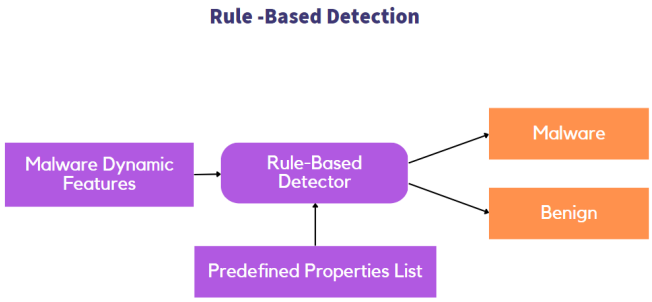

Fig. 3. Rule Based Detection System

using the extracted data while adjusting hyperparameters to improve performance.Once trained,the model and its weights are saved.Finally,a classification method is developed to differentiate between benign and malicious PE files.

B. Task 2:Deploy model in the Cloud

In this task,AWS Sagemaker is the chosen cloud service.The saved model is uploaded and loaded onto Sagemaker to enable its deployment.To deploy the model on the AWS cloud, you need to set up a notebook instance using AWS Sagemaker and execute all commands from within that notebook.Once the necessary libraries for endpoint creation are imported,the saved model and its weights are uploaded to the notebook instance.

Fig. 4. Endpoint creation

C. Task 3:Create a Client

The task involves creating Python code that can take a PE file as input and convert it into a feature vector that is compatible with the model.The feature vector is then run on the cloud API, and the results (i.e.,the nature of the file - Malware or Benign - or probabilities of each) are printed.The main objective of this task is to develop Python code that can classify the nature of PE files by performing all necessary operations within the cloud API previously created. To establish a connection with the AWS Sagemaker API,you can utilize the bot03 library and provide the required keys and token IDs from the AWS CLI.

Fig. 5. Client Execution

CONCLUSION

In the context of cloud computing, the paper suggests a two-component mechanism for detecting malware, comprising of the cloud and the client. The client submits potentially harmful file samples to the cloud, and then receives a report detailing whether the sample is malicious or not. The cloud system consists of three primary stages, namely analysis, detection based on behavior, and categorization.

During the analysis phase, various tools are utilized to scrutinize file samples, while virtual machines are used to gather execution traces, which are later forwarded to the behavior-detection agent. In the behavior-based detection phase, multiple factors such as system calls, types of system calls, system call paths, system resources, and file types are taken into account to create behaviors and features that distinguish between malicious and harmless patterns. The classification phase involves sending the chosen features to both the learning-based and rule-based agents, which categorize file samples as either malware or benign. Finally, the outcomes are transmitted back to the behavior-based detection agent, assessed, and conveyed to the client.

The proposed system is a comprehensive approach to malware detection, incorporating behavior-based detection, feature extraction, and machine learning-based classification. This system has the potential to improve computer system and network security by detecting and preventing the spread of malicious software.

REFERENCES

[1] Gupta, Shaw, S., Chakraborty, S. A pattern-based malware detection technique in cloud architecture. IEM, Kolkata.
[2] Sun, Wang, Buyya. CloudEyes: Cloud-based malware detection with reversible sketch for resource-constrained IoT devices.
[3] Huang, Du,Xiao, Li, X. Cloud-based malware detection game for mobile devices with offloading. IEEE Transactions on Mobile Computing.
[4] Krishnan,Abdelsalam, Huang, Sandhu. Malware detection in cloud infrastructures using convolutional neural networks. In Proceedings of IEEE 11th International Conference on Cloud Computing.
[5] Huang, Shen, Zhou, Yu, Fan, Cao. Multistage signaling game-based optimal detection strategies for suppressing malware diffusion in fog-cloud-based IoT networks, IEEE.
[6] Jang J, Moraru, Brumley, D., Andersen, D. G. (2011). SplitScreen: Enabling efficient, distributed malware detection. Journal of Communication and Networks.
[7] Young,. The Technical Writer's Handbook. Mill Valley, CA: University Science.