International Journal of Safety and Security Engineering
Vol. 14, No. 1, February, 2024, pp. 225-230
Journal homepage: http://iieta.org/journals/ijsse

Advances in Malware Analysis and Detection in Cloud Computing Environments: A Review

S. Madhusudhana Rao*, Arpit Jain

Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, India

Corresponding Author Email: madhusudhanarao.s@nic.in

https://doi.org/10.18280/ijsse.140122

Received: 15 June 2023
Revised: 21 December 2023
Accepted: 9 January 2024
Available online: 29 February 2024

Keywords:
malware detection, cloud computing, machine learning (ML), data security, network security, cyber-physical systems (CPS), intrusion detection

ABSTRACT

Cloud computing, integral for data storage and online services, presents significant advantages over traditional data storage and distribution methods, including enhanced convenience, on-demand storage, scalability, and cost efficiency. Its growing adoption in securing Internet of Things (IoT) and cyber-physical systems (CPS) against various cyber threats offers numerous opportunities. Despite the continuous evolution of malware and the lack of a universally effective detection method, cloud environments provide a promising approach for malware detection. Cloud computing, recognized for its efficiency, scalability, flexibility, and reliability on elastic resources, is widely utilized in the IT industry to support IT infrastructure and services. However, one of the foremost security challenges faced is malware attacks. Certain antivirus scanners struggle to detect metamorphic or encrypted malware in cloud environments due to complexity and scale, allowing such threats to evade detection. High detection rates with precision in reducing false positives are essential. Machine learning (ML) classifiers, a vital component in Artificial Intelligence (AI) systems, require training on extensive data volumes to develop credible models with high detection rates. Traditional detection methods face challenges in identifying complex malware, as modern malware employs contemporary packaging and obfuscation techniques to circumvent security measures. This paper provides a detailed discussion on detecting malware in cloud environments and the advantages of cloud computing in safeguarding IoT and CPS from cyber attacks. It presents a survey on malware analysis and detection models, aiding researchers in identifying limitations of traditional malware detection models in cloud environments and inspiring the design of innovative models with enhanced quality of service levels.

1. INTRODUCTION

In recent years, cyberattacks on global economies have grown more frequent. Steve Morgan estimates that by 2021, the world economy will have lost over $6 trillion due to cyberattacks. Malware generation is a major concern for data security, with over one million harmful files being developed daily [1]. Particularly for cyber-physical and mission-critical systems, the cost of malware is rising rapidly. According to McAfee, there has been an uptick in malicious mobile apps, banking Trojans, and backdoors. Cloud computing is a model of distributing computing resources across a network of remote machines [2]. In a cloud computing setup, data and applications can be accessed and managed by anyone with the appropriate credentials [3]. One or more computers in a network may perform these operations simultaneously [4]. These machines may be real or purely logical. Multiple virtual servers can be hosted on a single physical server and shown to the user as a single device. Since these virtual servers are not tied to the real server in any way, they are free to relocate and increase or decrease their resources as needed without harming the user experience [5].

The present method for detecting malware is not without its flaws. Start with a plethora of highly accurate classification methods that extract network functionalities through file disassembly [6]. The two most common ways of decompiling files using these methods are either delivering suspicious files to the server or installing professional applications on the tenant's virtual machine [7]. Most cloud tenants are hesitant to provide security service providers access to sensitive files due to privacy concerns, making network function extraction impractical [8]. Additionally, the majority of signatures are produced through static analysis of binary file code. However, malicious programmers frequently find ways to evade this method by implementing obfuscation-specific code modifications. Finally, the device's antivirus software's dependability is questionable [9]. Even with the use of virtual environment modeling and file behavioral checks for dynamic analysis, advanced malware continues to be recognized and elude detection [10].

The use of the cloud for detection purposes is among the most cutting-edge options accessible today [11]. This is achieved by utilizing a client and a server [12]. Servers in the cloud can detect malicious software in files sent to them by clients in the cloud [13]. Recent studies have demonstrated that malware detection rates can be improved and that each malware sample can be thoroughly analyzed using cloud-based detection [14]. Malicious software was created with the

intention of infiltrating and possibly destroying a computer system without the owner's knowledge. There are primarily two types of malicious software: file infectors and independent malware. Worms, backdoors, malware, rootkits, adware, and so on are all examples of malware that can manifest in different ways and cause different types of damage [15]. Because most modern malware uses many polymorphic layers to avoid detection or uses side processes to automatically upgrade to a newer version at relatively short intervals, traditional signature-based malware detection methods are becoming more and more useless [16].

Opting for cloud-based detection instead of more traditional methods has many advantages. With the advent of cloud computing, which makes use of processing power and enormous databases, malware detection has the potential to become more rapid and accurate [17]. Computers, mobile devices, and cyber-physical systems can all benefit from the cloud-based method's enhanced object identification capabilities. But this approach has many flaws, such as wasteful use of resources, loss of control over data, and no real-time monitoring [18]. This overview paper contributes significantly to the field by offering a comprehensive evaluation of the cloud-based malware detection technique [19]. This article provides a concise overview of recent scholarly work on the subject, discusses the latest trends in malware generation and evasion tactics, and, following an examination of the challenges encountered thus far, proposes novel approaches to malware detection. It also introduces a malware detection system that runs in the cloud and can spot dangerous software using a combination of heuristics, deep learning, signatures, and behavior analysis [20]. Providing a comprehensive review of the method and summarizing current research on cloud-based malware detection is the goal of this work [21]. Current challenges in malware detection are analyzed, and some possible solutions are proposed after a discussion of recent advances in malware creation and evasion techniques.

Malware is used by hackers to take advantage of security holes in the systems they compromise. These gaps can be anything from a simple buffer overflow to a fatal weakness in a network's underlying protocol [22]. The classification of malware has gotten more difficult as many strains share characteristics with numerous families. Viruses were the first type of malware discovered, followed by spyware created for more mundane purposes like stealing from friends or making money. However, as time has progressed, more sophisticated forms of malware have arisen, posing a serious threat to organizations of all sizes and to governments worldwide [23]. Traditional malware and next-generation malware are the two main groups of malicious software [24]. Malware that has been around for a while is easy to spot and delete from computers. However, the malware of the next generation is more pervasive and tougher to eliminate [25]. Malicious software, or malware, is software designed to do harm to its target system. Malware comes in a wide variety of forms, from viruses and worms to backdoors, rootkits, and even ransomware. Malware categories, defining features, and known families are outlined in Table 1.

Storage, computation, messaging, streaming media, developer tools, and security are just some of the many services that have benefited from the rise of cloud computing as a mainstream paradigm. The various cloud deployment models and services are depicted in Figure 1. With cloud computing, information may be accessed from any computer at any time. However, cloud malware offers a serious security risk due to its ability to steal personal information and login passwords, hijack digital devices, destroy systems, and even steal identities. Malware in cloud environments can hide on virtual machines, making detection more difficult. Several methods, such as those based on signatures, behaviors, and machine learning, have been proposed for detecting cloud-based malware. Amazon Web Services (AWS), Microsoft Azure, and Google Cloud all provide security services designed to identify malicious activity in the cloud. For instance, the Amazon Sentinel Service is a service for detecting hazards that uses anomalous activity to spot malware, and Google has released a detector for contemporary dangers.

Table 1. Types of malware and primary characteristics

Malware Types	Main Characteristics
Virus	Malware that is widespread and well-known
Worm	use networks to spread Permit illegal access to the CPS systems
Trojan Horse	The program that sends confidential information to third parties seems to be standard software
Backdoor	evades security measures and makes systems accessible from a distance
Rootkits	grant special access. Cover up their shady code from the host system
Ransom ware	the data on the compromised machine is encrypted
Obfuscated malware	It uses deceptive methods to blend in with the systems

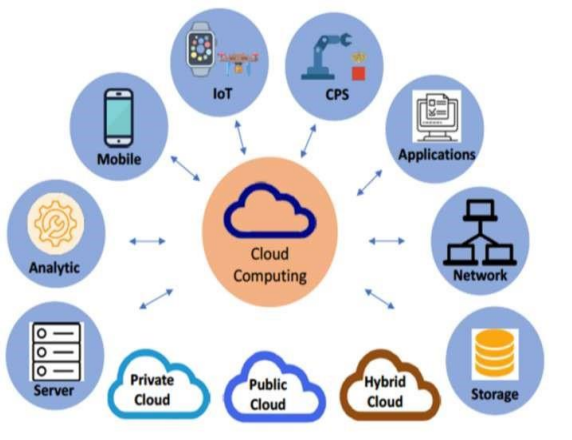

Figure 1. Cloud deployment models

The Canadian Institute for Cybersecurity supplied all of the data used in this investigation. A wide variety of malware logs are among the many data files included in the collection. A wide range of models can be trained using these recovered log features. The samples contained around 51 different families of malware. With 279 columns and 17,394 rows, the dataset had more than 17,394 data points from various locations.

Recently, virus detection has been one of the many areas where cloud computing has been put to use. Since cloud-based detection methods have significant advantages over traditional detection methods, many different cloud-based detection strategies and approaches have been developed [26]. Fundamental principles, feature extraction strategies, and algorithm varieties are often used in assessing these techniques. The cloud-based malware detection system is represented in Figure 2.

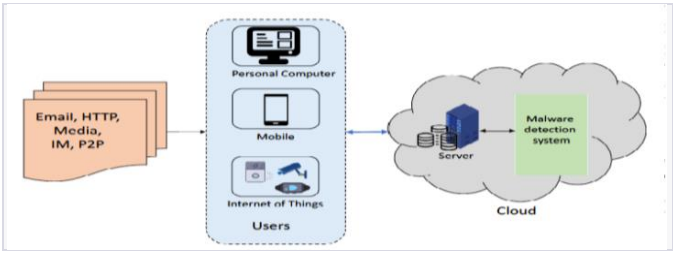

Figure 2. Cloud-based malware detection system

Malware detection researchers face challenges when dealing with datasets that include redundant and/or superfluous features, since learning approaches struggle to handle data that contains these types of features. In fact, most learners' accuracy drops and operational costs rise due to unnecessary or redundant behaviors for a given class. Hence, improving the total performance of the produced models requires continuously improving the work of producing datasets, which includes feature extraction and selection phases.

2. LITERATURE SURVEY

Encryption is essential for storing sensitive data in the cloud. Attribute-Based Encryption (ABE)-based access control is a highly effective way of safeguarding the privacy and authenticity of data kept in the cloud. You might be worried about its scalability and performance with the ABE Cipher-Text Policy because it doesn't let you add or remove computational nodes during runtime. Existing approaches also have the problem of single-point-of-failure (SPoF). In order to ensure data sharing on public cloud storage with high levels of availability, Qaisar et al. [1] presented a scalable multi-agent architecture for systems based on CP-ABE. This architecture will be used in the proposed study. A cloud host was suggested by the author as a way for authorized agents to communicate with users without jeopardizing system security and privacy. Using the efficacious power of the cutting-edge Gemini method to combat malware, the author introduced a novel approach to cloud security. Gemini is a helpful approach for discovering similarities between graph embeddings using only binary codes. The proposed study solves problems with efficiency and scalability while offering a method for cloud-based malware detection. Scalability, efficiency with multiple agents, and malware detection are the three main topics it tackles.

When it comes to online data storage and service delivery, one of the most fascinating new developments is cloud computing. In order to safeguard computer-based systems against cyber-related dangers, it is advantageous to employ this rapidly developing technology rather than more traditional protection measures. Computer-based systems necessitate protection, including critical infrastructure, CPS, traditional PCs, mobile devices, and the Internet of Things (IoT). Any piece of software that aims to undermine the security, privacy, or availability of a computer system is considered malicious software (malware). In order to identify the ever-increasing malware attack surface, Aslan et al. [2] suggested a cloud-based, intelligent behavior-based detection technique. The proposed method starts with gathering malware samples from different virtual machines, which makes it easy to extract unique properties. The next step is for the author to supply certain attributes to the learning-based and rule-based detection agents so that they can differentiate between safe and malicious files. By studying 10,000 code samples, we were able to determine how effective the proposed method was. The proposed approach accurately and efficiently detects both known and unknown malware. In comparison to state-of-the-art methods, the proposed strategy has also produced superior results.

Cloud computing has emerged as a new standard in industry, thanks to the fast expansion of the industrial Internet, and it has the potential to revolutionize the way businesses operate. In recent years, cloud-based service models have become increasingly attractive to many enterprises. A major concern, however, is the possibility of security flaws, such as undiscovered malware assaults on virtual domains. To protect virtual domains within a cloud-based service platform, Mishra et al. [3] proposed an introspection-based security solution called VMShield to fight malware in cloud infrastructure. By seeing processes' actions in real time through the hypervisor's virtual memory, VMShield stops malware from evading the security technology. Because it uses reflection to find covert attacks, the suggested solution outperforms static and dynamic state-of-the-art procedures. In order to select useful characteristics, VMShield uses the meta-heuristic methodology known as binary particle swarm optimization, which it obtains from the Bag of N-gram method applied to system calls. Using a Random Forest (RF) classifier, the monitored programs are categorized as either good or bad processes. This approach has the potential to detect different types of malware more efficiently than the traditional signature-matching method.

Cloud computing is becoming more popular among organizations as a means to support their diverse range of software programs. With these services, businesses can cut down on expenses related to hardware management, scalability, and maintenance. Popular cloud service providers (CSPs) that offer infrastructure as a service (IaaS) to companies like these include Amazon Web Services, Microsoft Azure, and Google Cloud Platform. The security of cloud services has thus become an important concern for CSPs, as they are increasingly being targeted by cybercriminals due to the increasing popularity of cloud computing. It is commonly known that malware poses a significant and ubiquitous threat to cloud-based IaaS. Research by Kimmel et al. [4] examined the efficacy of RNN-based deep learning techniques for malware detection in cloud virtual machines. By monitoring system metrics such as CPU, memory, and disk usage during malicious attacks, these models learn malware behavior over time.

The IoT paves the way for a hyper-connected world where every object is linked to the internet and can exchange data instantly. Smart city and smart industrial applications combine the IoT with 5G and AI technology. Potential security breaches are increasing in proportion to the proliferation of interconnected devices, networks, and applications. Detecting IoT malware has been the subject of several studies as a means of protecting against the risks offered by malicious code. The fast development of new and variant IoT malware makes it difficult to predict when existing methodologies will no longer be able to accurately identify harmful IoT code detected by static analysis. To mitigate the effects of malware on IoT devices, Jeon et al. [5] proposed DAIMD, a system that can detect both classic and newly evolved IoT malware. In an embedded cloud environment, the DAIMD technique uses a

convolutional neural network (CNN) model to dynamically analyze IoT malware. Through dynamic analysis in a nested cloud environment, DAIMD is able to extract memory, network, virtualized systems of files, processes, and system call behaviors from malware that is specific to the Internet of Things.

Due to the severity of the threats posed by Android malware, the need to identify it has become critical. Privacy issues and communication bottlenecks are commonplace in cloud Android malware detection. In actual use, the need for offline updates highlights the need for on-device training. However, on-device training is challenging to accomplish, particularly for those high-complexity malware detectors, due to the restricted resources of mobile devices. Using the recently described wide learning method, Yuan et al. [6] developed a lightweight on-device Android malware detector to address this issue. In order to train its models, the detector mostly used one-time computing. As a result, it may be trained in its entirety or in stages, right on the go, with a mobile device. The detection accuracy of the detector is comparable to that of the models based on deep learning multilayer perceptron (MLP) and CNN, while it surpasses the shallow learning-based models supported by support vector machines (SVM) and AdaBoost. In addition, compared to current detectors, the proposed model can better withstand hostile cases, and this is only going to become better with on-device model retraining.

There are many security gaps in 5G technology. Due to their continued use of the 4G network infrastructure, many current networks have security holes that make them vulnerable to 5G attacks. Without universally accepted security standards, cyber attacks and malware on 5G IoT devices could spread unchecked. Machine learning, improved edge computing, post-quantum cryptography, and other forms of AI-driven communication are all expected to play a role in the deployment of the 6G network of the future. Edge computing enables the dispersed incorporation of the processing capacity of supercomputing servers onto the devices at the edge of a network. With this enhancement, users in even the most remote areas will enjoy the same high level of service, catalyzing the rapid expansion of related apps. It is becoming more difficult to detect malware infections in this intricate environment. Examining the theoretical and experimental works on data-driven malware detection in the large-scale data-intensive domain, Uysal et al. [7] focus on continuous learning, new ideas in multi-domain to multi-target learning, challenges with unknown data, imbalanced data, and data scarcity, and new ideas in explainability via visualization with a multi-labelling approach that can identify malware based on its recipes.

Recent developments in computing have led to an uptick in the amount of time people spend in virtual environments. Because of the COVID-19 pandemic, this process has been accelerated. Since moving their operations online, cybercriminals have shifted their attention away from the real world. The reason behind this is that committing a crime online is much easier than in the real world. Cybercriminals frequently use harmful software, often called malware, to initiate cyberattacks. The latest malware incorporates sophisticated methods of obfuscation and packing. These techniques of malware hiding make detection and classification much more difficult. New approaches, distinct from the norm, are required to successfully combat new forms of malware. Modern malware is so complex that traditional ML methods, which are a branch of artificial intelligence, cannot identify it all. The deep learning (DL) approach provides optimism for the future of malware detection, which is a challenging problem that calls for innovative approaches. An innovative hybrid model for malware categorization based on deep learning was introduced by Aslan and Yilmaz [8]. An innovative hybrid architecture for the optimal combination of two diverse pre-trained network models is the principal outcome of this research. Collecting data, designing the deep neural network's architecture, training the network, and finally evaluating its performance are the four main processes in this architecture.

The security of large amounts of data kept in the cloud is jeopardized by the inability to detect and prevent viruses. More research is needed in this field, even though privacy and security are important worries in the big data world. Malware must be detected and removed promptly to prevent data integrity and, eventually, system standing from being compromised. One important approach to malware detection in recent years has been ensembles, which combine numerous classifiers. Due to their scale, memory, and processing demands, as well as the high cost of data transfer during training and operation, large ensemble classifiers are not well-suited for big data in the cloud. In order to address this problem, Abawajy et al. [9] introduced Hybrid Consensus Pruning (HCP), the initial pruning method that combined various classifier types using a fast consensus mechanism. Experiments were used to see how well the HCP method worked at finding malware by comparing it to other ways of pruning very large ensemble classifiers, such as EPIC, DHCEP, and K-Means Pruning.

Secure cloud computing is an essential component of the future generation of computers. If malicious software is advanced enough to identify a security tool on a TVM, it can conceal its actions. This means that TVM layer security solutions are not reliable. Using Virtual Machine Introspection (VMI), Mishra et al. [10] laid forth a security architecture design for identifying common and atypical attacks through fine-grained monitoring of virtual machines.

Bhardwaj et al. [11] offered a unique architecture for classifying network traffic as either benign or DDoS assault traffic by combining a well-conceived stacking sparse Auto Encoder (AE) for learning features with a Deep Neural Network (DNN). The settings of AE and DNN are tuned with methods specifically developed for DDoS attack detection.

Containerization's popularity in the cloud has been on the rise recently. Kubernetes has made it easier to control software running in isolated containers. Kubernetes makes it possible to automate processes related to application administration, such as self-healing, scaling, rolling back, and updating. Attacks on pods to carry out harmful acts are another example of how security threats have grown. Crypto mining malware, which steals computing power to mine bitcoin, is among the most dangerous new strains of malware. A crypto mining process, initiated by a concealed malware executable, can operate in the background during application deployment and operation in the pod; thus, a way to detect malicious crypto mining software operating within Kubernetes pods is required. One approach that could work is to use ML to determine if a pod is hosting a crypto mining operation and then label it accordingly. In addition to detection, the network administrator will require an explanation of the ML's categorization outcome and its justification. A pod's removal or restart with a new image are both disruptive administrative actions, and the explanation will justify and support these actions. Karn et al. [12] detailed

the planning and execution of a ML-based system for detecting anomalous pods in a Kubernetes cluster by keeping an eye on the system calls made by the Linux kernel.

In order to categorize malware attacks on the IIoT, Ahmed et al. [13] proposed a 5G-enabled system consisting of a deep learning-based architecture. In this approach, convolutional neural networks are used to recognize and distinguish between distinct forms of malware based on an image representation of the virus. Ribeiro et al. [14] created a unique host-based IDPS for Android (HIDROID), which is capable of running in its entirety on a mobile device while imposing only a light computational load. It gathers information in real time, sampling characteristics that represent the consumption of limited mobile device resources. In order to create a data-driven model for harmless behavior, the detection engine makes use of statistical and machine learning technologies. When an observation is made that does not fit this model, an alert is sent, and the preventative agent takes the necessary precautions. HIDROID is convenient for everyday use because it doesn't need harmful data for training or tuning.

Two virus assaults on a group of digital nodes were studied by Varma et al. [15]. Malicious software infects the nodes and then uses their processing power for cloud computing, bitcoin mining, and other similar activities. This article was written with the assumption that the SIS compartmental model describes the transmission of viruses. Malware developers have the ability to control how much resource their host nodes utilize, which the author assumed. In the short term, more money will be available, and in the long run, malware may be detected and eliminated more rapidly if resources are used more efficiently. You can fully secure the afflicted node with anti-virus software at a reduced cost if the threat is recognized. The two players in this hypothetical race try to outdo each other by squeezing as much money as possible out of the infected nodes. Table 2 represents the traditional model analysis that indicates the working of the models and their limitations are also included.

Table 2. Traditional model analysis

Author Name	Year of Publication	Proposed Model	Remarks
Qaisar et al. [1]	2021	Data sharing on public cloud storage using high levels of availability can be guaranteed with the author's new scalable multi-agent design for systems based on CP-ABE.	The proposed models analyze only limited and known intrusion patterns in which novel attacks are not detected. The time complexity levels of the models are high that need to be reduced. The nodes that are normal in nature turned into malicious nodes that is not concentrated in traditional models.
Aslan et al. [2]	2021	As the number of malware attack vectors continues to grow, the author suggests a cloud-based intelligent behavior-based detection method. The first step of the proposed method is to gather malware samples from different virtual computers. From these samples, distinct traits may be readily extracted.	
Mishra et al. [3]	2021	The author proposed VMShield, a security approach based on introspection, to safeguard virtual domains on a platform for cloud-based services. By monitoring the hypervisor's virtual memory to record processes' actions in real time, VMShield stops malware from circumventing the security tool.	
Kimmel et al. [4]	2021	The author investigated how well cloud VMs can identify malware using deep learning approaches based on Recurrent Neural Networks (RNNs). Malware behaviour is learned by these models over time by observing system parameters like CPU, memory, and disc utilisation at runtime during malicious attacks.	
Jeon et al. [5]	2020	The author suggested a dynamic analysis for IoT malware detection (DAIMD) that can identify both well-known and newly-evolved IoT malware. The DAIMD method dynamically analyses IoT malware in a nested cloud environment, learning it with a convolution neural network (CNN) model.	
Yuan et al. [6]	2021	The author developed a lightweight on-device Android malware detector to address this issue. In order to train its models, the detector mostly used one-time computing. As a result, it may be trained in its entirety or in stages, right on the go with a mobile device.	
Uysal et al. [7]	2023	In this review, the author surveys the research on data-driven malware detection in the large-scale data-intensive domain, looking specifically at the literature on continuous learning, new ideas in multi-domain to multi-target learning, challenges with unseen/unknown data, imbalanced data, and data scarcity, and new ideas in explainability via visualisation using a multi-labelling approach.	The considered models complexity levels are high and the number of features in the traditional models are also more than need to be reduced. Feature dimensionality reduction models can be applied on the traditional models that helps reduce the training time and enhance the accuracy levels in intrusion detection.
Aslan and Yilmaz [8]	2021	Malware categorization using a hybrid model was introduced by the author as a new architecture based on deep learning. This study's main contribution is an innovative hybrid design that optimally merges two different types of pre-trained network models.	
Abawajy et al. [9]	2020	The author presented Hybrid Consensus Pruning (HCP), the first pruning technique that use a quick consensus function to merge different types of classifiers into a single one.	
Mishra et al. [10]	2020	The author presented a security architectural design for fine-grained monitoring of virtual machines using Virtual Machine Introspection (VMI) to identify both common and uncommon attacks.	

An era of rapid expansion and extensive use has dawned on the Internet of Things (IoT). Because they are unable to employ complex security measures, these low-power devices are particularly susceptible to infestation. An anomaly detection solution for IoT networks that utilizes edge computing to uncover hidden threats is proposed by Lakshman Narayana et al. [16] and called ADRIoT. A traffic preprocessor, a collection of anomaly detectors tailored to individual devices, and a traffic capturer are all components of an edge's detection module. Each detector is constructed using an LSTM autoencoder in an unsupervised way, which safeguards against future zero-day vulnerabilities without requiring labeled attack data. Once a client device connects to the edge, it will retrieve the suitable detector from the cloud and process it locally. A single-edge device's limited resources, such as a home router, also make it difficult to implement such a detection module. To get over this problem, the author came up with a multiedge collaborative strategy that uses a neighborhood network to manage additional traffic by combining the capabilities of multiple nodes.

Because of its widespread use, Android is a prime target for cybercriminals. This highlights the critical nature of finding effective countermeasures to these attacks. Recent advances in machine learning have provided a practical answer to the problem of malware detection based on the identification of distinct traits. Malware detectors powered by machine learning have a lot of capabilities, but attackers can still exploit those characteristics to their advantage. Consequently, creating new skills that can reliably detect malicious behavior is one of the main tasks of the Android security department. Ullah et al. [17] presented a novel feature representation approach for malware identification that combines API-Call Graphs (ACGs) with byte-level picture representation.

Every single person who uses the internet these days is dealing with malware. A new kind of harmful software called polymorphic malware can change its behavior in response to different threats and situations. Because it is constantly evolving, polymorphic malware evades traditional signature-based malware detection techniques. Malware and other harmful threats were detected using many machine learning algorithms by Akhtar and Feng [18]. The implementation of the most accurate method was determined by its high detection ratio. For a more complete picture of the system's precision, the confusion matrix included measurements for both false positives and false negatives.

3. DISCUSSIONS

It is challenging for a cloud malware analysis and detection system to identify the latest strains of malware, such as polymorphic and metamorphic malware. Malware detection software relies on behavioral detection approaches that are intrinsically linked to the system/environment the malware is currently operating in. Creating an environment like that on the cloud is challenging. A cloud service can employ virtual computers to simulate a client's on-premises infrastructure. Complete system replication necessitates repeatable cloud-based playback. Replaying a program with concurrency and inter process communication might lead to non determinism. Deterministic replay requires serializing thread access, but this can cause some exceptions to go unnoticed if they were thrown in the original application when two or more threads accessed the same object simultaneously. This exception in the program could be the entrance to malware. In this scenario, the replay won't be able to spot the infection. In addition, it is impractical to create an exact copy of every system because doing so could use up excessive amounts of resources.

Malware has emerged as one of the biggest problems affecting internet users today. Malicious software has evolved into a new breed called polymorphic malware, which is far more versatile than older infections. Typical malware detection approaches rely on signatures, however polymorphic malware is always changing its signature traits to evade detection. A number of machine learning algorithms are analyzed to detect dangers and infections. The most accurate algorithm was chosen and enhanced version can be improved for system utilization when the detection ratio was high. The confusion matrix's measurement of false positive and false negative counts was an advantage since it gave more insight into the system's performance. Using the results of malware analysis and detection with machine learning algorithms to calculate the difference in correlation symmetry integrals, it was showcased that harmful traffic on computer systems could be detected, leading to improved network security using machine learning techniqie.

Finding malware that shares similar behaviours within a family is possible using both static and dynamic learning approaches [10]. Dynamic analysis accounts for malicious files' activity by monitoring data flows, recording function calls, and adding monitoring code to dynamic binaries [11], in contrast to static analysis that only looks at the contents of files without actually running them. By analyzing both static and behavioral artifacts, machine learning algorithms can better understand the dynamic nature of modern malware and detect more sophisticated attacks that would otherwise go undetected by signature-based methods. Machine learning-based solutions outperform newly disclosed malware since they are not signature-dependent. Accurately obtaining and representing features is possible with the help of deep learning algorithms that can execute feature engineering independently [12].

Due to the immaturity of cloud computing and the lack of standardized application programming interfaces used by cloud suppliers, most cloud users are forced to re-write their programs when migrating between cloud providers. If an antimalware engine has to be relocated to a different cloud, its communication component will typically need to be rewritten. Cloud customers need just demand standardization and interoperability to put an end to this issue once and for all. Based on the limitations identified in the survey, there is strong requirement to build a framework for malware detection in cloud environments and to detect VM level and Services Level attacks. It is necessary to build a model using machine learning methods for classification of attacks using existing data sets. Fine tune and testing the model in appropriate environment after preprocessing the data collected with the attacks detected at VM level and services level need to be performed for achieving enhanced quality of service levels.

4. CONCLUSION

Cloud computing offers several advantages, such as cost-effectiveness, scalability, and high availability on elastic resources. The platform is also widely used in IT for enabling technological infrastructure and services. However, malware assaults are a major security concern due to the complex

ecosystem and large size of services. Malware that changes or encrypts itself can evade some of the most prevalent antivirus scanners' detection capabilities. A high identification rate and precise precision detection are critical for avoiding a high false-positive rate. An example of a service that could be offered through the cloud is malware detection. This study provides a concise overview of the systems that are currently in use. In addition to providing a compilation of the most recent literature and data for easy comparison and analysis, this study can be utilized as incentive to improve existing models and develop new methodologies for cloud malware investigation and detection. The model's features can reduce the quantity of data sent between users and the cloud, protect the user's device from infection, and shorten the time it takes to detect malware. Through its harmful behaviors towards the integrity, availability, and confidentiality of system resources and services, malware has an impact on computing systems, including cloud systems. Because the implementation of cloud computing environments is becoming more widespread, users are implicitly dependent on them for many services because they provide users with cost-saving services. Therefore, it is crucial to maintain cloud security. In this study, a survey on malware categories that could abuse cloud computing infrastructure is presented. This survey helps numerous researchers to design innovative and intellectual solutions for accurate detection of malware in cloud environment. In future, the suggested models can be implemented for the accurate detection of intrusions in the network with feature dimensionality reduction models and also with dynamic malware detection with self repair strategies.

REFERENCES

[1] Qaisar, Z.H., Almotiri, S.H., Al Ghamdi, M.A., Nagra, A.A., Ali, G. (2021). A scalable and efficient multi-agent architecture for malware protection in data sharing over mobile cloud. IEEE Access, 9: 76248-76259. https://doi.org/10.1109/ACCESS.2021.3067284

[2] Aslan, Ö., Ozkan-Okay, M., Gupta, D. (2021). Intelligent behavior-based malware detection system on cloud computing environment. IEEE Access, 9: 83252-83271. https://doi.org/10.1109/ACCESS.2021.3087316

[3] Mishra, P., Aggarwal, P., Vidyarthi, A., Singh, P., Khan, B., Alhelou, H.H., Siano, P. (2021). VMShield: Memory introspection-based malware detection to secure cloud-based services against stealthy attacks. IEEE Transactions on Industrial Informatics, 17(10): 6754-6764. https://doi.org/10.1109/TII.2020.3048791

[4] Kimmel, J.C., Mcdole, A.D., Abdelsalam, M., Gupta, M., Sandhu, R. (2021). Recurrent neural networks based online behavioural malware detection techniques for cloud infrastructure. IEEE Access, 9: 68066-68080. https://doi.org/10.1109/ACCESS.2021.3077498

[5] Jeon, J., Park, J.H., Jeong, Y.S. (2020). Dynamic analysis for IoT malware detection with convolution neural network model. IEEE Access, 8: 96899-96911. https://doi.org/10.1109/ACCESS.2020.2995887

[6] Yuan, W., Jiang, Y., Li, H., Cai, M. (2019). A lightweight on-device detection method for android malware. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51(9): 5600-5611. https://doi.org/10.1109/TSMC.2019.2958382

[7] Uysal, D.T., Yoo, P.D., Taha, K. (2022). Data-driven malware detection for 6G networks: A survey from the perspective of continuous learning and explainability via visualisation. IEEE Open Journal of Vehicular Technology, 4: 61-71. https://doi.org/10.1109/OJVT.2022.3219898

[8] Aslan, Ö., Yilmaz, A.A. (2021). A new malware classification framework based on deep learning algorithms. IEEE Access, 9: 87936-87951. https://doi.org/10.1109/ACCESS.2021.3089586

[9] Abawajy, J.H., Chowdhury, M., Kelarev, A. (2015). Hybrid consensus pruning of ensemble classifiers for big data malware detection. IEEE Transactions on Cloud Computing, 8(2): 398-407. https://doi.org/10.1109/TCC.2015.2481378

[10] Mishra, P., Varadharajan, V., Pilli, E.S., Tupakula, U. (2018). Vmguard: A VMI-based security architecture for intrusion detection in cloud environment. IEEE Transactions on Cloud Computing, 8(3): 957-971. https://doi.org/10.1109/TCC.2018.2829202

[11] Bhardwaj, A., Mangat, V., Vig, R. (2020). Hyperband tuned deep neural network with well posed stacked sparse autoencoder for detection of DDoS attacks in cloud. IEEE Access, 8: 181916-181929. https://doi.org/10.1109/ACCESS.2020.3028690

[12] Karn, R.R., Kudva, P., Huang, H., Suneja, S., Elfadel, I.M. (2020). Cryptomining detection in container clouds using system calls and explainable machine learning. IEEE Transactions on Parallel and Distributed Systems, 32(3): 674-691. https://doi.org/10.1109/TPDS.2020.3029088

[13] Ahmed, I., Anisetti, M., Ahmad, A., Jeon, G. (2022). A multilayer deep learning approach for malware classification in 5G-enabled IIoT. IEEE Transactions on Industrial Informatics, 19(2): 1495-1503. https://doi.org/10.1109/TII.2022.3205366

[14] Ribeiro, J., Saghezchi, F.B., Mantas, G., Rodriguez, J., Abd-Alhameed, R.A. (2020). Hidroid: Prototyping a behavioral host-based intrusion detection and prevention system for android. IEEE Access, 8: 23154-23168. https://doi.org/10.1109/ACCESS.2020.2969626

[15] Varma, V.S., Hayel, Y., Morărescu, I.C. (2022). A non-cooperative resource utilization game between two competing malware. IEEE Control Systems Letters, 7: 67-72. https://doi.org/10.1109/LCSYS.2022.3186620

[16] Lakshman Narayana, V., Lakshmi Patibandla, R.S.M., Pavani, V., Radhika, P. (2023). Optimized nature-inspired computing algorithms for lung disorder detection. In: Raza, K. (eds) Nature-Inspired Intelligent Computing Techniques in Bioinformatics. Studies in Computational Intelligence, vol 1066. Springer, Singapore. https://doi.org/10.1007/978-981-19-6379-7_6

[17] Ullah, F., Srivastava, G., Ullah, S. (2022). A malware detection system using a hybrid approach of multi-heads attention-based control flow traces and image visualization. Journal of Cloud Computing, 11(1): 1-21. https://doi.org/10.1186/s13677-022-00349-8

[18] Akhtar, M.S., Feng, T. (2022). Malware analysis and detection using machine learning algorithms. Symmetry, 14(11): 2304. https://doi.org/10.3390/sym14112304

[19] Shaukat, K., Luo, S., Varadharajan, V. (2023). A novel deep learning-based approach for malware detection. Engineering Applications of Artificial Intelligence, 122: 106030. https://doi.org/10.1016/j.engappai.2023.106030

[20] Zheng, L., Zhang, J. (2022). A new malware detection method based on VMCADR in cloud environments. Security and Communication Networks, 2022: 4208066. https://doi.org/10.1155/2022/4208066
[21] Narayana, V.L., Sujatha, V., Sri, K.S., Pavani, V., Prasanna, T.V.N., Ranganarayana, K. (2023). Computer tomography image based interconnected antecedence clustering model using deep convolution neural network for prediction of COVID-19. Traitement du Signal, 40(4): 1689-1696. https://doi.org/10.18280/ts.400437
[22] Narayana, V.L., Sujatha, V., Sri, K.S., Pavani, V., Prasanna, T.V.N., Ranganarayana, K. (2023). Computer tomography image based interconnected antecedence clustering model using deep convolution neural network for prediction of COVID-19. Traitement du Signal, 40(4): 1689-1696. https://doi.org/10.18280/ts.400437
[23] Nahmias, D., Cohen, A., Nissim, N., Elovici, Y. (2020). Deep feature transfer learning for trusted and automated malware signature generation in private cloud environments. Neural Networks, 124: 243-257. https://doi.org/10.1016/j.neunet.2020.01.003
[24] Wang, X., Zhang, J., Zhang, A., Ren, J. (2019). TKRD: Trusted kernel rootkit detection for cybersecurity of VMs based on machine learning and memory forensic analysis. Mathematical Biosciences and Engineering, 16(4): 2650-2667. https://doi.org/10.3934/mbe.2019132
[25] Jian, Y., Kuang, H., Ren, C., Ma, Z., Wang, H. (2021). A novel framework for image-based malware detection with a deep neural network. Computers & Security, 109: 102400. https://doi.org/10.1016/j.cose.2021.102400
[26] Pinhero, A., Anupama, M.L., Vinod, P., Visaggio, C.A., Aneesh, N., Abhijith, S., AnanthaKrishnan, S. (2021). Malware detection employed by visualization and deep neural network. Computers & Security, 105: 102247. https://doi.org/10.1016/j.cose.2021.102247