Al-Salam Journal for Engineering and Technology
Journal Homepage: http://journal.alsalam.edu.iq/index.php/ajest
e-ISSN: 2790-4822p-ISSN: 2958-0862

A Sample Proposal Enhancing the Security of the Cloud Computing System Through Deep Learning and Data Mining

Israa Ezzat Salem[1]*, Karim Hashim Al-Saedi[1]

[1]Computer Science Department, College of Science, Mustansiriyah University, Baghdad, IRAQ.

*Corresponding Author: Israa ezzat salem

DOI: https://doi.org/10.55145/ajest.2024.03.01.001
Received June 2023; Accepted August 2023; Available online August 2023

ABSTRACT: Malware or malicious applications can cause catastrophic damage to not only computer systems but also data centers, web, and mobile applications from various industries; the Ministry of Interior, in particular, is the most important educational institution because they are more vulnerable to security breaches. Keeping stakeholder data safe from unwanted actors is a big concern that brings us to the concept of malware detection and prevention. Deep learning and data mining using artificial intelligence (AI) can be an efficient approach for developing anti-malware systems. Following suit, this study gave a thorough examination of malware detection methodologies and procedures. Initially, we attempted to provide a comprehensive description of malware, artificial intelligence, and data mining, as well as a listing of these technologies. The suggested system was described (whether this data is files, photographs, videos, or import limitations and is processed and identified by mining and deep learning data, and the system was trained on data). So far, our findings suggest that artificial intelligence and data mining can be used to construct anti-malware systems to detect and prevent malware assaults or security threats in software applications geared toward technological wonderland and its real-world application in the Ministry of Interior. To conclude, we outline dozens of possibilities for overcoming the observed restrictions and intend to expressly continue our efforts toward significant advancements in malware detection and prevention by implementing this proposal. We give a detailed look at the current ways to find malware, their flaws, and ways to make them more effective. We also explain how we're working on integrating the system. Our study shows that adopting future approaches to developing malware detection applications should provide significant advantages. Understanding this structure should help researchers do more research on malware detection and prevention using AI and data mining.

Keywords: Cloud computing, Malware worms detection, Deep learning, Data mining

1. INTRODUCTION

Cyberattacks on the global economy have substantially escalated in recent years. Steve Morgan predicts that by 2021, cyberattacks will have destroyed $6 trillion of the world economy [1]. Researchers have found that there are now approximately 1 million malware files created per day [2] and that the cost of malware, particularly for CPS [3] and important systems, is also rising. Backdoors, banking trojans, and fraudulent mobile applications have all increased significantly, according to the McAfee research [4]. One of the greatest risks to conducting a cyber assault in the area of information security is malware. Malware is any program that harms the victim's device, including viruses, worms, rootkits, backdoors, and ransomware [2]. the system's owner's permission. Malware has increased alarmingly over the last ten years, and there is currently no method that is widely used to identify all malware in the field. In order to avoid detection by security systems, this new type of malware employs sophisticated obfuscation and encapsulation methods. Because of this, it is almost hard to find complex malware using a conventional technique. By examining the executable program, malware detection is the process of finding the existence of harmful software. Malware detection methods have been proposed in a variety of approaches, including conventional and cutting-edge methods. Traditional approaches, such as signature-based, heuristic, and behavior-based detection methods, as well as model-based detection methods, have been in use for more than ten years. Different methodologies, such as machine learning, deep learning, edge computing, and cloud computing, constitute the foundation of advanced technology. While the signature-based detection technique is known to be efficient in terms of time and memory use, it is unable to identify malware that is

*Corresponding author: farg.israa@uomustansiriyah.edu.iq

not yet known to exist. Heuristic, behavioral, and model-based detection techniques can find a lot of malware, but they cannot find zero-day malware. Similar to the approaches used for signing, routing, and behavior-based methods, discovery methods based on deep learning and edge computing (mobile devices) also fail to identify sophisticated and common malware. Malware detection systems are moving away from old methods and toward more modern ones. One of the best modern detection techniques is cloud-based detection. It has two components: client and server. In cloud computing, the server does the analysis and decides whether or not the specific suspicious file is dangerous software after receiving a suspicious file from the client via the Internet. The server employs a number of detecting variables during the analytic process to enhance performance. Threads, system calls, static and dynamic features, API traces, application traces, and hybrid features are all employed during the feature extraction stage. According to recent research, a cloud-based detection technique improves the detection rate of both known and undiscovered malware [5], [6], and offers a more in-depth examination of each malware sample. The cloud-based detection strategy has a number of benefits over conventional techniques. For virus detection, the cloud environment offers greater computing capacity and significantly bigger datasets. The same infection might leave behind several execution traces [5]. Additionally, it enhances the capability of CPS, mobile devices, and personal gadgets to identify objects. The disadvantages include lost data management, client and server overhead, a lack of real-time monitoring, and restricted use of the infrastructure. The following contributions are made by this study, which gives a thorough analysis of a cloud-based malware detection strategy for deep learning and data mining:

- Gives an overview of recent academic works on deep learning and data mining methods for cloud-based malware detection.
- Explains how to leverage the cloud to safeguard against viruses and secure cyber-physical systems.
- Explains current trends in malware development and disguising methods.
- Discuss current issues and novel methods for malware detection.
- Provides a platform for cloud-based malware detection that uses citation, deep learning, behavior, and signature methodologies.

The remainder of this essay is structured as follows. The second segment discusses related work on cloud-based malware detection techniques for data mining and AI. A description of cloud-based malware detection systems is provided in the third part. Section IV provides a summary of for data mining and AI. Section Five presents a description and assessment of a cloud-based malware detection strategy. The suggested framework for our strategy is presented in Section VI. Section VII provides the conclusion and recommendations for further study.

2. RELATED WORK

By adhering to the policy of detection and prevention, malware detection serves to safeguard the system from different forms of harmful assaults. Although there are many current malware detection methods, with the development of malware technology, artificial intelligence adoption is essential for effective and reliable malware protection apps. Finding the harmful source code is the initial step in malware detection.

- To find malware source code repositories, the largest malware source code database must be identified., researchers [7] devised a method called SourceFinder. According to the study, the suggested technique has an 86 percent recall rate and an 89 percent accuracy rate when identifying malware repositories. It uses SourceFinder to identify 7504 malicious source code repositories, and then analyzes the features and qualities of those repositories. It is standard practice to employ machine learning algorithms to find malware. Probably a lot more. An assessment of malware detection methods is presented together with a thorough review of the static, dynamic, and hybrid methodologies by Niharika Sharma [8]. Additionally, the author speeds up the identification process by fusing data mining and machine learning methods. The research also assesses several machine learning- and data mining-based malware detection methods.
- Sanjay Sharma et al. [9] suggest utilizing machine learning approaches to identify malware using an approach based on opcode occurrence. Using data from the Kaggle Microsoft malware classification challenge dataset, the researchers put five classifiers to the test: LMT, REPTree, Random Forest, NBT, and J48Graft. According to a demonstration, the proposed approach can identify the infection with nearly 100% accuracy.
- Despite difficulties in using machine learning for intrusion detection, such as novel computing paradigms and novel evasion strategies, Three significant issues are presented by Sherif Saad et al. [10] that hinder the effectiveness of machine learning-based malware detection. The researchers then suggest potential strategies to get over the limitations before talking about the key behavioral analysis that will rule the future generation of antimalware systems.
- • In addition to machine learning, malware detection uses cloud computing, network-based detection systems, virtual machines, and a mix of different methods and technologies. Deep learning and artificial intelligence are being used more and more to find malware. By combining binary visualization with self-organizing incremental neural networks, Irina Baptista et al. [11] have come up with a new way to find malware. In a demonstration of recognizing malicious payloads in many file formats, including Portal Document File.pdf and Microsoft Document File.doc files, the results of testing show that ransomware can be found with an accuracy of 91.7% and 94.1%,

respectively. The authors say that the proposed method worked well and had a good rate of incremental detection, which made it possible to find new malware in real time.

- A virtual analyst was created using artificial intelligence in a separate research by Syam and Vankata [12] to guard against attacks and collect the necessary measures. The researchers divide their data into supervised and unsupervised categories, turn the latter into the former using analyst input, and then automatically update the program. Over time, it uses the Active Learning Mechanism to improve the algorithm, making it stronger and more effective.
- An innovative Bayesian optimization-based approach is proposed by a team of researchers from Kennesaw State University [13] for automatic hyperparameter optimization that produces the best DNN architecture. The demonstration results demonstrate the usefulness of the framework, and the paper investigates the NSL-KDD, a benchmark dataset for network intrusion detection. Accuracy, precision, recall, and f1-score all show that the DNN Architecture successfully detects much more invasion. The random search optimization-based method is inferior to the BO-GP-based methodology.; for the KDDTest+ and KDDTest-21 datasets, respectively, BO-GP attained the greatest accuracy of 82.95 percent and 54.99 percent. Yuan et al [14]. Give a deep learning method for connecting the aspects of the static and dynamic analyses of Android applications. Additionally, they turned on DroidDetector, an Android malware detection engine based on deep learning that can determine if a file exhibits malicious behavior. They evaluated DroidDetector and did a thorough analysis of the components that deep learning largely explores to film malware totally using a huge number of Android applications. With access to more setup details, deep learning seems to be especially appealing for classifying Android malware. DroidDetector can recognize objects with an accuracy of 96.76 percent, which is higher than that of conventional machine learning techniques. Ding and others [15] To find malware, propose an affiliation mining technique based on API invites. Unique methodologies are introduced to increase the speed of OOA mining identification. To improve governing quality, criteria are proposed to identify the API and to eliminate APIs that cannot visit objects distinctively. Experiments demonstrate that the proposed systems can significantly increase OOA running speed. In our studies, the time spent extracting information is lowered by 32%, and the time spent arranging is decreased by 50%. Eskandari and others [16] The HDM-Analyzer is a novel hybrid strategy that incorporates areas of interest into dynamic and reliable high-frequency investigative procedures while maintaining accuracy at a respectable level. By using real-world data gathered by researching components, HDM-Analyzer can forecast the majority of core leadership; thus, they have no overhead performance. This paper's primary responsibility is to adopt the desired viewpoint exactly while researching and standardizing the component in a consistent examination, keeping in mind the final goal of increasing the correctness of the established inquiry. The overheads of the execution were, in fact, incurred during the learning phase; as a result, they have no impact on the light extraction stage of the examination process. The preliminary findings demonstrate that HDM-Analyzer outperforms static inquiry methodologies and components in terms of overall accuracy and multifunctional time quality.
- Miao et al [16]. Give a bi-layer behavior reflection approach in light of the semantic assessment of dynamic API sequences. Processes on delicate frame assets and complex procedures are isolated in understandable semantic levels. By examining the reliance on the data, raw API calls are coupled at the lower layer to extract low-layer practices. In the upper class, comprehensive interpretation of low-class traditions is combined to create higher-class practices that are more intricate. A high-dimensional vector space is then presented using several low-layer techniques. Now, distinct methods may be directly used using a number of well-known machine learning accounts. Furthermore, it is advised to create OC-SVM-Neg, a single-class boosting vector machine (OC-SVM) that gains from unfavorable and accessible instances, in order to address the problem of not sufficiently studying thoughtful projects or drastically unbalanced malware and friendly projects. The experiment shows that the recommended extraction approach employing OC-SVM-Neg outperforms dual works on the incorrect rate of caution and speculative ability. Based on links between system call pools, Nikolopoulos and Polenakis [18] created a graph-based model that determines if an unknown software sample is harmful or benign and classifies a malicious program into one of a number of well-known malware family groups. Customers used system call dependency charts, often known as ScD graphs, which were created from effects gathered during dynamic pollution study and were more accurate. When we apply our recognition and arrangement systems to a weighted coordinated graph, also known as a Gr-graph, which is a specific graph of a group's relationship that results from scd-graph after collecting separate subsets of its peaks, drastic changes take place. The authors built their model to withstand these changes. For the classification strategy, the authors provided saMe-similitude and NP-similarity metrics that integrate saMe-NP convergence, as well as a measure of comparability in Delta for the discovery method. Finally, they evaluated their malware detection and classification model, determining the efficiency of the model against malware based on detection rates and classification precision.

3. BACKGROUND THEORY

This section reviews several key subjects related to data security and intrusion detection using deep learning and data mining:

3.1 ARTIFICIAL INTELLIGENCE AND MALWARE

The goal of this research was to examine malware characteristics, such as polymorphisms, as opposed to existing malware kinds like worms, viruses, etc. Malware does not, however, have any universally agreed-upon standard characteristics, according to the definition given in the malware read. This essay will instead concentrate on well-known malware that targets the Ministry of the Interior's data. In this section, the findings on malware are organized into the following categories.

• malware that utilizes machine learning and deep learning methods from artificial intelligence.
• Malware with AI technologies Incorporated
• Malware behaving like biological equivalents
• Malware behaving like biological equivalents

1) Malware with AI technologies Incorporated: There is virtually little material that specifically states that malware uses AI. One virus specifically mentioned is called Zellome, and it includes genetic algorithms (GA) as a kind of brute-force method to produce decryptor routines to support its polymorphic behavior [19]. Research on this malware by Symantec came to the conclusion that artificial intelligence technologies were not used well enough [20]. However, some people were interested in how it used AI.

2) Malware exhibits intelligent-like behaviours: Researchers and anti-virus developers have noticed that certain software exhibits intelligence [21], such as non-predictive actions [22], as a result of studies into the behavior of malware. There are malware programs like Storm that demonstrate certain artificial intelligence qualities, such as automatically altering their defensive strategies to thwart any attempts to halt their spread [23].

3) Malware behaving like biological equivalents: Some malware exhibits traits of artificial life or mimics the symptoms of diseases or their biological analogues. According to studies [12], there are observably striking parallels between biological viruses that infect live things and their computer-based analogues. For instance, Kienzle and Elder's research [24] indicated that the majority of computer worms are descended from worms that may be found in nature. The capacity to proliferate at the expense of the host and the capacity to infect their host through an opening share similarities. Both are able to spread on their own without human intervention. Both have the ability to delay taking action for a while. Similar creatures combine their abilities to produce more dangerous behaviors. Malware includes the Nimda worm, a concoction of two distinct worms that surfaced after the September 11 terrorist attacks on the United States. Also, it has been demonstrated that malware acts similarly to genuine parasites.It's interesting to note that, according to Fumell and Ward [25], malware has been on the rise, with the main factor driving this surge being a profit-oriented motive. The propagation of malware infection has been tried to be modeled using biological epidemic models by researchers. A homogeneous epidemic model proved effective in simulating the spread of random-scanning worms, according to Chen and Ji [26]. Some academics have even argued that viruses and malware may be examples of artificial life. Artificial life has traits like the capacity for development and evolution, as well as the capacity for self-reproduction, information storage, and self-representation. According to Spafford [26], computer viruses display characteristics that have been identified as characteristics of artificial life, such as information on how they portray themselves. However, there are a number of serious flaws discovered, such as the computer virus' reliance on its host machine, that prevent him from acknowledging computer viruses as artificial life.

3.2 DATA MINING IN MALWARE DETECTION

Signature-based detection is the most popular detection approach, and it is at the heart of every commercial antivirus application [27][28]. Malicious code authors have devised a variety of stealth strategies to circumvent detection by standard signature-based algorithms. Because standard signature-based detection approaches are unable to detect this new sort of malware, virus researchers are focusing on developing more broad, scalable characteristics that can recognize harmful activity as a process rather than a single static signature. Analysis is classified into static and dynamic analytics. Static analysis scans the computer code without actually executing it, while dynamic analysis executes the program in a real or virtual environment. While static analysis is free of the execution cost, it has limits when there is a dynamic decision point in the program's control flow. Dynamic analysis examines software execution to detect whether behavior is potentially dangerous. These two techniques are also used in tandem [29]. Dynamic analysis, on the other hand, is only used during decision points in the program's control flow. We offer a static analysis approach that uses data mining methods to automatically extract important behavior from malware and clean programs in this study. As the major categorization characteristic, we propose employing instruction sequences retrieved from a clean and malicious software decompiler. The behavior represented by these sequences of instructions is crucial if they are to distinguish between malicious and clean programs, and the following is the distinction between our technique and other static analysis approaches discussed in the relevant study area. First, the proposed method treats data mining as a full process, from data preparation through model construction. Despite the fact that data preparation is a critical phase in the data mining process, practically all current static analysis approaches discussed in the relevant study section do not go into depth about it. Second, instead of employing a fixed length byte like n-gram, all of the characteristics were a learning sequence derived via deconstruction.

3.3 CLOUD COMPUTING IN MALWARE DETECTION

Many claims that some kind of cloud computing is now commonplace due to the Internet's pervasiveness in everyday life. A precise definition of cloud computing is necessary since this study has a strong emphasis on cloud computing technologies. Cloud computing is difficult to describe. There are several meanings, all of which have the Internet in common. Cloud computing enables a single device or region to regularly use the Internet by utilizing all the features that are pre-installed on computers. Another part of it is using shared computer resources while having local servers control the apps. Users of cloud computing do not have to worry about where or how their data is stored. They simply start using the services whenever they want to. The main advocates of this technology are virtual appliances and virtualization (Hypervisor). Cloud computing allows customers to select the service model that best meets the needs of their environment by offering a number of service models. Three different cloud service models are software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) [4] [5]:

- Software-as-a-service (SaaS): The customer makes advantage of the provider's cloud-hosted apps. Salesforce.com CRM Application, as an example.
- Platform-as-a-service (PaaS): Own apps are deployed by users into the cloud infrastructure. The supplier must support the programming languages and application development tools utilized. Google Apps is an illustration.
- Infrastructure-as-a-service (IaaS): Clients can contribute storage, networks, processing power, and other resources, as well as install and run any software, from apps to operating systems. The kind of data uploaded using cloud computing is depicted in Figure 1.

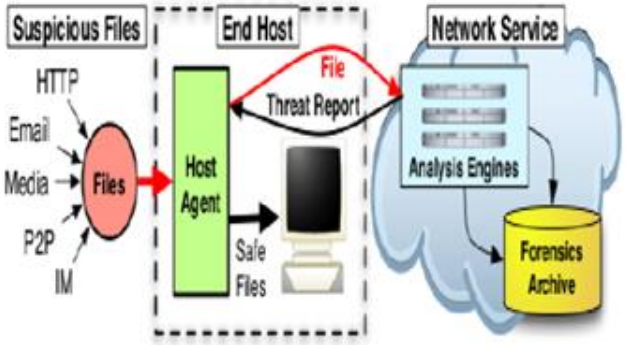


FIGURE 1. - The way cloud computing systems work

4. METHODS OF RESEARCH

The primary goal of the methodological proposal is to identify, research, and explore the relevant existing methods in my study in order to conduct the study on malware detection approaches utilizing artificial intelligence for deep learning and data mining in cloud computing. We first performed a search to identify potential research papers from scientific databases using a predefined search, we had to select these search strings to avoid results from irrelevant research papers, these keywords are based on malware and artificial intelligence for deep learning and data mining because of the supposed problems Addressed in our research the constant threat to data by attackers, especially when using the cloud and the difficulty of detecting malicious worms as a result of the rapid spread and the emergence of new worms, the goal of the proposal was to create a new method for securing data, especially if it was observed that most sites of some institutions do not adopt a security system. By deep learning the behavior of worms and the effective features of data mining for worm patterns. We will train the system on existing data, where the dataset for training used IDS dataset, but it is applied to the Ministry of Interior that has important data to preserve in the testing, especially when the cloud is a source for that data, especially for ministries with sensitive information.

In this section, we talk about malware detection systems and show the results and possible limits of data analysis by extracting critical data using data mining and deep learning and making a detection decision. This is especially important for data uploaded to the cloud, which is more likely to have its privacy violated. As shown in the figure below, the way it works is like this:

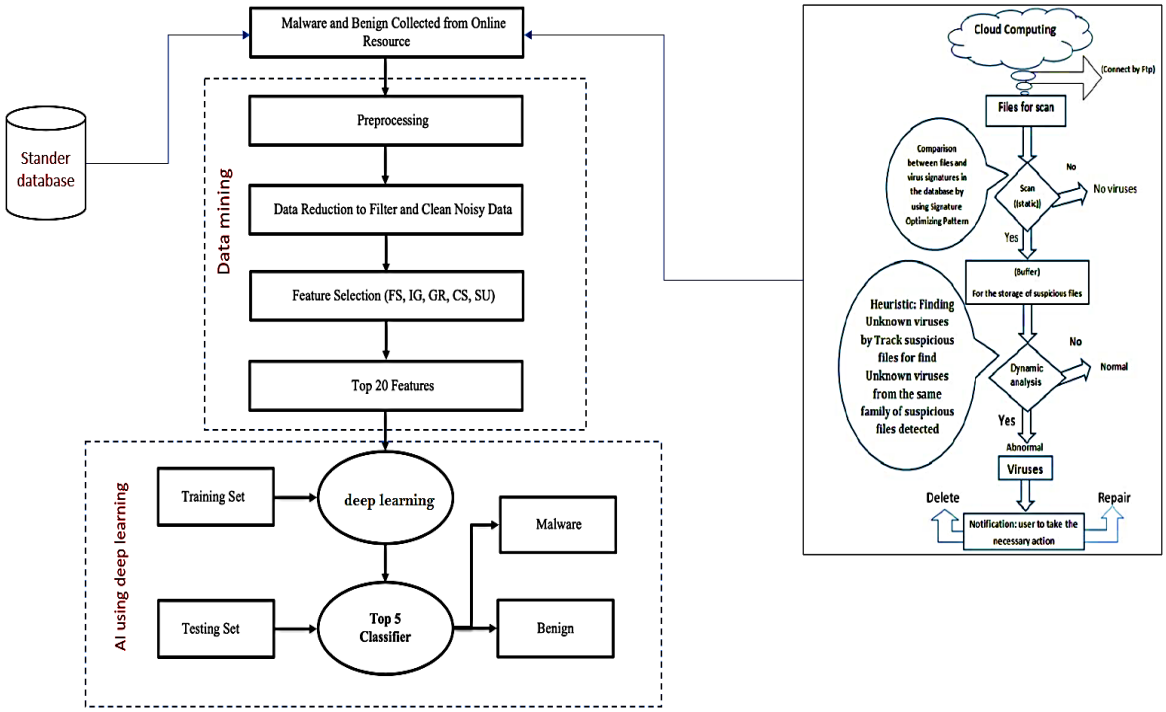


FIGURE 2. - Proposal for research methodology

Figure (2) shows how data processing, feature selection, classifier training, and malware detection are all steps in the flow of malware detection. First, data sets of both bad and good web apps are collected through the IDS Kaggle website. Malware detection systems will be built in a way that uses AI and data mining to process malware datasets and analyze malware to figure out what makes it what it is Data mining uses the following metrics to select 20 features: Fisher Score (FS), Chi-Square (CS), Information Gain (IG), Gain Ratio (GR), and Uncertainty Symmetric (US) (US). The system must compare various classifiers on FS, CS, IG, GR, and US to train the classifier for deep learning. This allows the system to find unknown malware.

4.1 STANDER DATASET

In this section, we will use a host-based IDS dataset with the malware execution traces needed to train and develop a successful IDS system for the cloud. However, the latest datasets are outdated, lack traces of a malware attack, and cannot be adapted into a hypervisor for IDS development in the cloud. Therefore, developing a dataset of traces of malware attacks captured from a hypervisor or a VMM from the cloud will help create IDS for the cloud to protect against such attacks in the future through artificial intelligence and data mining methods. Here, we will not rely only on this data, we will rely on the data that can be uploaded by the cloud, whether it is files, videos, photos, and others. Then we will test the system on a dataset from the Ministry of Interior or one of the ministries that need more protection. Now we are in touch with the Ministry of Interior and the Intelligence Department if they can help implement the system in their ministry. We will analyze the data with malware and benign software to analyze those data in order. Malware detection techniques can be put into three groups: those that look for signatures, those that look for unusual behavior, and those that look for malware. In this section, we talk about malware detection systems and give some factors that could affect the results.In our research, we will rely on all types of malware, as shown in Figure (3).

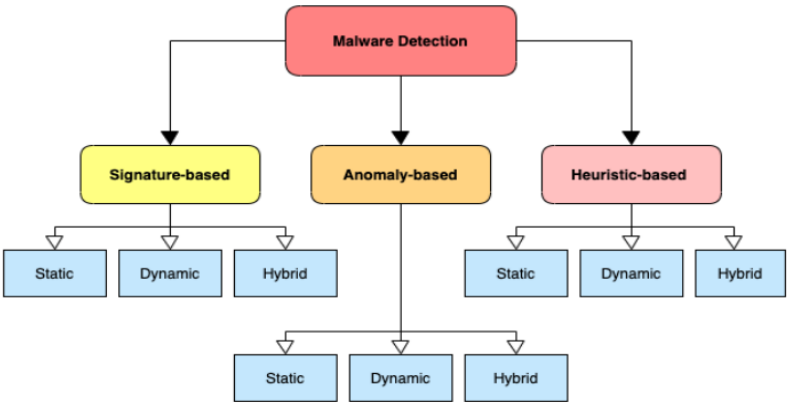
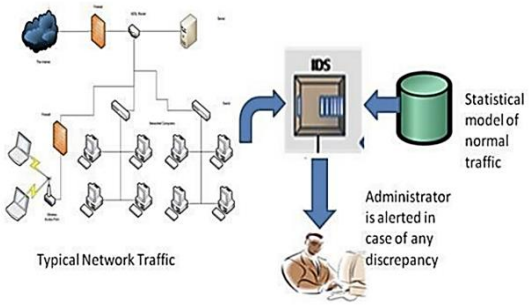


FIGURE 3. - Malware Detection Techniques classification

-signature-based method, the developers use a database containing virus signatures, scan the file, and evaluate the information using this database to detect malware in the database. If the information matches the database data, then the file contains viruses. The main advantage of this method is effective for known malware, however, it has limitations in detecting unknown malware. Figure (4) shows that IDS maintains a statistical model of traffic that can also be referred to as a database, IDS accepts traffic from different sources and matches it with statistical traffic to see if it is malicious or not, and then saves it as an administrator.

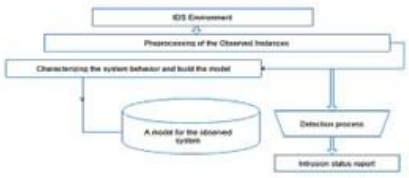


Signature based Intrusion Detection System (IDS)

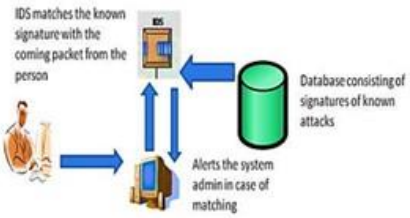
The methodology used in Signature based IDS

FIGURE 4. - Signature malware for IDS dataset.

-Here we will introduce all kinds of malicious data to test where fault-based network intrusion detection plays a vital role in addressing security issues and protecting networks from malicious activities, especially in the data types used (IDS). Defect-based methods circumvent the drawbacks of signature-based methods by applying classification techniques to the operations of a malware detection system in order to identify any known or undiscovered malware. This shift from pattern-based detection to a classification-based method of determining what is normal and what is not helps in the identification of malware activity.



Common anomaly-based network IDS



Anomaly Based IDS

FIGURE 5. - Anomaly-based on IDS dataset

-Heuristics-Based Detection Technology The application of AI to signature and anomaly-based detection systems increases the effectiveness of malware detection. In order to adapt to environmental change and enhance predictability, a deep learning algorithm for neural networks has been implemented to the malware detection system in order to improve the categorization approach. The method employs properties such as inheritance, selection, and combination to gain access to optimal solutions from various sources without prior system knowledge. Combining

statistical and mathematical methodologies enhances the heuristic approach of earlier methods. Figure 6 represents the features of the exploratory methods.

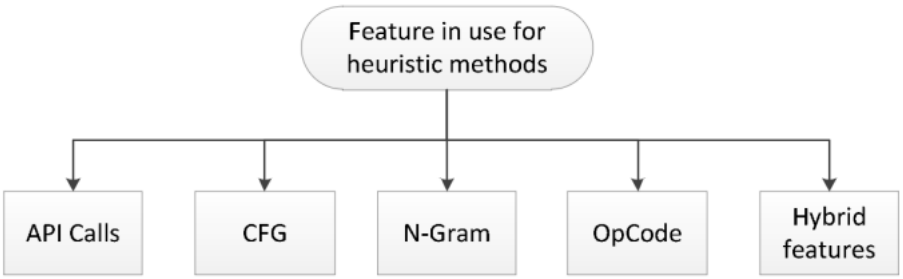


FIGURE 6. - Feature for heuristic methods

4.2 DATA MINING

Once we've found the data that has the malware, we'll use data mining, data filtering, and any other mining techniques to process it. In deep learning, it's important to choose the right features. In the proposed method, there are 1808 features, and many of them don't give the right amount or even take from them. In our issue, it's important to maintain accuracy by reducing the number of details. So, we first used Fisher Score (FS) to choose features, and then we looked at four other ways to choose features. The filtering approach is followed by the five-feature selection method used in this method. In this method, the relationship between each feature and the category (malicious or safe software) is found, and each feature's impact on the rating is calculated. This method works with any classification algorithm that is not encapsulated, and it lets you compare how well different classifiers work. In this method, Fisher score (FS), information gain (IG), gain ratio (GR), chi-square (CS), and uncertainty symmetry (US) are used. Using these criteria for selecting features We chose the 20 best features.

4.3 AI USING DEEP LEARNING

In this part, we'll talk about malware detection technologies that use AI for deep learning, as well as the results and possible limits of these technologies. This keeps the host-based intrusion detection system (IDS) from being hidden, but it also moves the IDS away from the host to make it harder to attack. Evaluation is the ability to control how the host and the main application talk to each other with the help of a virtual machine monitor. However, the suggested solution is limited in terms of how quickly it can be modified and how prone it is to error.

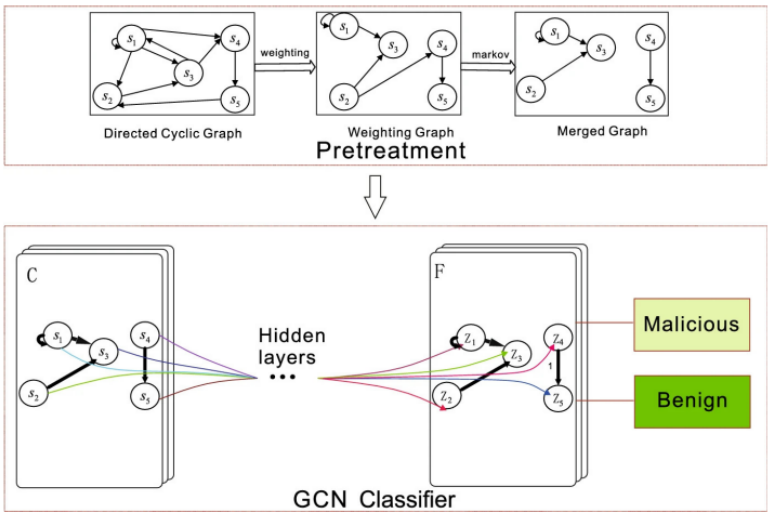


FIGURE 7. - GCN-based Malware Detection System Framework.

It proposes a method for classifying malware using a graphical convolutional network that can adapt to various infection characteristics. The method begins by extracting the API call sequence from the malicious code and constructing a vector cycle graph. The graph's convolutional network is then used to create a classifier using the Markov chain and principal component analysis techniques. The method also looks at and compares its own results. Figure 7 shows a framework for a malware detection system that is built on GCN.

5. CONCLUSION

Malware or malicious applications can do a lot of damage to computer systems, data centers, and web and mobile apps in many different industries. The Ministry of the Interior is the most important educational institution because security breaches are more likely to happen there. Making sure that stakeholder data is safe from bad actors is a big challenge, which brings us to the idea of detecting and stopping malware. Using deep learning and data mining as part of artificial intelligence (AI) can be a good way to make anti-malware systems. In line with this trend, this study gave a detailed look at the techniques and methods used to find malware. In the beginning, we tried to give a clear overview of malware, AI, data mining, and their listing. We talked about the proposed system that we came up with to stop hackers from getting into files, images, videos, or import restrictions that were uploaded to the cloud. This data is processed and found through data mining and deep learning, and the system was trained on data. So far, our research indicates that artificial intelligence and data mining can be an useful starting point for developing anti-malware systems to detect and prevent malware assaults or security issues in software programs, both in the technical marvel and the Department of the Interior. To reach a conclusion, we discuss numerous strategies to circumvent the challenges we've identified. We would also like to make it clear that we will continue to work diligently to make significant improvements in malware detection and prevention using this idea.

FUNDING

No funding received for this work

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their efforts.

CONFLICTS OF INTEREST

The authors declare no conflict of interest

REFERENCES

[1] S. Levy and J. Crandall, “The program with a personality: Analysis of elk cloner, the first personal computer virus,” Jul. 2020.
[2] A. P. Namanya, A. Cullen, I. Awan, and J. Pagna Diss, “The world of malware: An overview,” Sep. 2018.
[3] R. Samet and Ö. Ö. Tanrıöver, “Using a Subtractive Center Behavioral Model to Detect Malware,” Security and Communication Networks, 2020.
[4] D. Oudyal, D. Dasgupta, Z. Akhtar, and K. D. Gupta, “Malware analytics: Review of data mining, machine learning and big data perspectives,” Dec. 2019.
[5] O. Kayode, D. Gupta, and A. S. Tosun, “Towards a distributed estimator in smart home environment,” in 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). IEEE, 2020, pp. 1–6.
[6] A. Singh and A. Jain, “Study of cyber-attacks on cyber-physical systems,” in Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), 2018, pp. 26–27.
[7] M. O. F. Rokon, R. Islam, A. Darki, E. Papalexakis, and M. Faloutsos, “Sourcefinder: Finding malware source-code from publicly available repositories,” in RAID, 2020.
[8] N. Sharma and B. Arora, “Data mining and machine learning techniques for malware detection,” in Rising Threats in Expert Applications and Solutions, V. S. Rathore, N. Dey, V. Piuri, R. Babo, Z. Polkowski, and J. M. R. S. Tavares, Eds. Singapore: Springer Singapore, 2021, pp. 557–567.
[9] S. Sharma, R. Challa, and S. Sahay, Detection of Advanced Malware by Machine Learning Techniques: Proceedings of SoCTA 2017, Jan. 2019, pp. 333–342.
[10] S. Saad, W. Briguglio, and H. Elmiligi, “The curious case of machine learning in malware detection,” 2019.
[11] I. Baptista, S. Shiaeles, and N. Kolokotronis, “A novel malware detection system based on machine learning and binary visualization,” May 2019, pp. 1–6.
[12] M. Zekri, S. El Kafhali, N. Aboutabit, and Y. Saadi, “DDoS attack detection using machine learning techniques in cloud computing environments,” in 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), 2017, pp. 1–7.
[13] S. A. Repalle and V. R. Kolluru, “Intrusion detection system using AI and machine learning algorithms,” Dec. 2017.

[14] J. Sun, R. Wyss, A. Steinecker, and P. Glocker, "Automated fault detection using deep belief networks for the quality inspection of electromotors," Technisches Messen, vol. 81, no. 5, pp. 255–263, 2014.
[15] S. Tsimenidis, T. Lagkas, and K. Rantos, "Deep learning in IoT intrusion detection," Journal of Network and Systems Management, vol. 30, no. 1, pp. 1–40, 2022.
[16] M. Elsisi, et al., "Effective IoT-based Deep Learning Platform for Online Fault Diagnosis of Power Transformers Against Cyberattack and Data Uncertainties," Measurement, 2022, Art no. 110686.
[17] S. Malik, A. K. Tyagi, and S. Mahajan, "Architecture, Generative Model, and Deep Reinforcement Learning for IoT Applications: Deep Learning Perspective," in Artificial Intelligence-based Internet of Things Systems. Springer, Cham, 2022, pp. 243-265.
[18] "IoT deep learning," MIT News, Nov. 13, 2020. [Online]. Available: https://news.mit.edu/2020/iot-deep-learning-1113. Accessed: Feb. 5, 2024.
[19] D. Kajaree and R. Behera, "A Survey on Healthcare Monitoring System Using Body Sensor Network," Int. J. Innov. Res. Comput. Commun. Eng., vol. 5, no. 2, pp. 1302–1309, 2017.
[20] F. C. C. Garcia, C. M. C. Creayla, and E. Q. B. Macabebe, "Development of an Intelligent System for Smart Home Energy Disaggregation Using Stacked Denoising Autoencoders," in International Symposium on Robotics and Intelligent Sensors, IRIS 2016, IEEE, Japan, 2016.
[21] T. J. Saleem and M. A. Chishti, "Deep Learning for Internet of Things Data Analytics," Procedia Computer Science, vol. 163, pp. 381–390, 2019.
[22] S. M. Tabish, M. Z. Shafiq, and M. Farooq, "Malware detection using statistical analysis of byte-level file content," in Proceedings of the ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics, 2009, pp. 23-31.
[23] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," arXiv preprint arXiv:1901.03407, 2019.
[24] Z. Cui, F. Xue, X. Cai, Y. Cao, G. G. Wang, and J. Chen, "Detection of malicious code variants based on deep learning," IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3187-3196, 2018.
[25] A. Azmoodeh, A. Dehghantanha, and K. K. R. Choo, "Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning," IEEE Transactions on Sustainable Computing, vol. 4, no. 1, pp. 88-95, 2018.
[26] D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis," Computers & Security, vol. 81, pp. 123-147, 2019.
[27] M. M. Mijwil, M. Aljanabi, and ChatGPT, "Towards Artificial Intelligence-Based Cybersecurity: The Practices and ChatGPT Generated Ways to Combat Cybercrime," Iraqi Journal for Computer Science and Mathematics, vol. 4, no. 1, pp. 65-70, Jan. 2023. https://doi.org/10.52866/ijcsm.2023.01.01.0019.
[28] M. M. Mijwil, H. K. Hiran, R. Doshi, M. Dadhich, A. H. Al-Mistarehi, and I. Bala, "ChatGPT and the Future of Academic Integrity in the Artificial Intelligence Era: A New Frontier," Al-Salam Journal for Engineering and Technology, vol. 2, no. 2, pp. 116-127, Apr. 2023. https://doi.org/10.55145/ajest.2023.02.02.015.
[29] P. Szor, The Art of Computer Virus Research and Defense. Addison Wesley for Symantec Press, New Jersey, 2005.