Machine Learning Based Malicious Detection Approaches: A Survey

Mr. Mahesh T. Dhande[1], Dr. Sanjaykumar Tiwari[2]

[1]Research Scholar, Department of Computer Science & Engineering, Monad University U.P., (India),

[2]Associate Professor, Monad University U.P., (India)

Abstract: **-** Malware is harmful software designed to breach user privacy, endanger computer systems, or obtain unauthorized access to networks. As a result of the growing number of uses for computers and the reliance on electronically stored private information, malware attacks on confidential data are becoming a serious problem for people and businesses worldwide. Thus, malware protection is essential to maintaining the security of our personal computing systems and data. Recent research articles have alternately concentrated on single-attacking techniques or malware detection systems. As far as we are aware, no survey article presents malware patterns of attack and protection techniques in tandem. This work attempts to tackle this problem by combining machine learning (ML) oriented models of detection for complex and contemporary malware with a variety of malicious attack methods. This allows us to concentrate on a taxonomy of malware attacks according to four basic dimensions: the attack's main objective, its mode of attack, its targeted exposure as well as execution process, and the kinds of malware that carry out the attack. Thorough details on methods for analysing malware are also looked into. Furthermore, a thorough discussion is held regarding current malware detection methods that use ML algorithms and feature extraction.

Keywords: *Machine Learning, Malicious, Challenges and Issues in Machine learning, Security Measure in ML.*

1. Introduction

The newest type of web-based computing network is machine learning, which provides users with flexible and easy-to-use resources to access various cloud apps. The development of many technologies has been largely driven by the internet. Among them, machine learning has received attention recently. It is an enumerated standard [1] that provides flexible implementation, facts, figures, and information storage architecture by connecting a sizable pool of public or private systems. Machine learning has the potential to transform a data center from a large-exhaustive milieu into a pricing variable setting since it is a practical means of achieving obvious cost advantages [2]. In recent years, machine learning has emerged as a new technology and is anticipated to become a major concern in the years to come. Because it's a modern technology, it needs to be protected against the newest security flaws and has other problems [3].

Since it provides both customers and suppliers of cloud computing huge cost savings as well as new business options, the approach to cloud computing has advanced dramatically and exponentially, emerging as a revolutionary trend in the field of information technology. Customers that use cloud services as needed, pool resources as a resource that may grow high or low as needed rapidly and elastically, pay only for what is consumed, and access services with a networked infrastructure are the ones who define machine learning. The way that IT is traditionally delivered as services is evolving due to machine learning. Cost savings, scalability, efficiency, asset utilization, increased efficiencies, and mobility are among the company and IT outcomes. Although machine learning service providers boasted about the performance and security of their offerings, the real security and dependability of cloud computing delivery is not up to par [4].

By connecting the cloud application to the internet, machine learning provides a mechanism for information in the cloud to be remotely retrieved and stored. By choosing cloud services, users safeguard their Metadata on the cloud data server. Cloud service providers can access or manage the data kept in the data center in the cloud. Thus, data collection for analysis in a data center in the cloud needs to be done as professionally as possible [5].

Data in the cloud needs to be protected on its own, especially when it comes to data separation within the cloud service. Virtualization, encryption, and authentication are three ways to achieve differing degrees of data separation. This improves data security against illegal access. When it comes to collaborative occupation cloud settings, which can accommodate numerous clients or clients who do not see or exchange data with one another but may utilize resources or applications in an implementation context while possibly not belonging to the same organization, it is crucial. These days, agencies are trying to prevent having to focus on the IT framework. They must focus on running their firm if they want to increase production. Compared with the traditional IT approach, machine learning offers a multitude of potential advantages. But from the perspective of the user, concerns over machine learning security represent a significant barrier to its adoption. Without explicit user control, machine learning refers to the availability of computer network services, primarily for data storage and computing power [6].

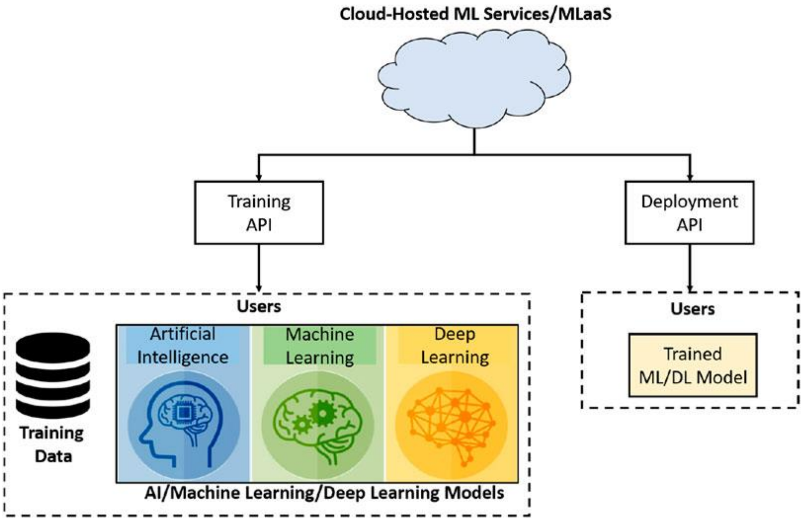

Figure 1. Architecture of Machine learning Security system [7]

For consumers of the cloud, privacy is a serious concern. This technology needs sufficient safeguards and controls to lessen consumers' worries. The majority of cloud service users are wary of having their personal data accessible or shared with other suppliers of cloud services. The user information consists of four components that require protection: (i) Usage information; obtained computer device details (ii) Private information, such as details on bank accounts, fitness, etc. (iii) Information that enables the classification of a person's personal data (iv) Outstanding computer uniqueness; only observable details, such as IP addresses, hardware identities, etc. [8].

Using the services of cloud service providers, data stored in the cloud are processed as well as accessed on a web server. As a result, machine learning is becoming more valuable, expanding as a market, and drawing a lot of interest from the corporate and educational communities. Unfortunately, there were a lot of problems with the cloud storage solution, including security concerns and limited access. Because cloud storage services are centered on two-way data exchange between service providers and customers, machine learning security considerations include integrity, transparency, availability, identification, permission, and confidentiality. As a result, there is a growing likelihood of data compromise, which can be broadly categorized into two categories: critical data and historical data. An important piece of information is one that an individual needs to know at all times, therefore any interruption or disappearance would irritate him. Furthermore, the data found in archives are frequently quite rare collectively and at non-critical times. As such, the gap in it cannot be seen as the primary issue [9].

This work is organized as follows: Section 1 provides an introduction to machine learning, while Section 2 covers its background. After that, Section 3 provides the relevant work, and Section 4 discusses issue statements. Additionally, Section 5 discusses machine learning techniques. The answer to data security concerns is covered in Section 6. Section 7 discusses Pros and Cons. After that, Section 8 offers the discussion, and Section 9 concludes with a final analysis.

2. Background Of Machine Learning

For middleware among app development platforms, commercial applications, and enterprise systems, machine learning provides a resource-sharing environment. Some Cloud operating models include subscription infrastructure services, free infrastructure services with value other platform services, and customers' profitable free marketing services. In certain respects, academics, businesspeople, researchers, and IT companies define the phrase "machine learning." Clouds are a large, easily accessible collection of virtualized tools. The ability to dynamically adjust these services to a changeable charge (scale) allows for optimal resource utilization [10].

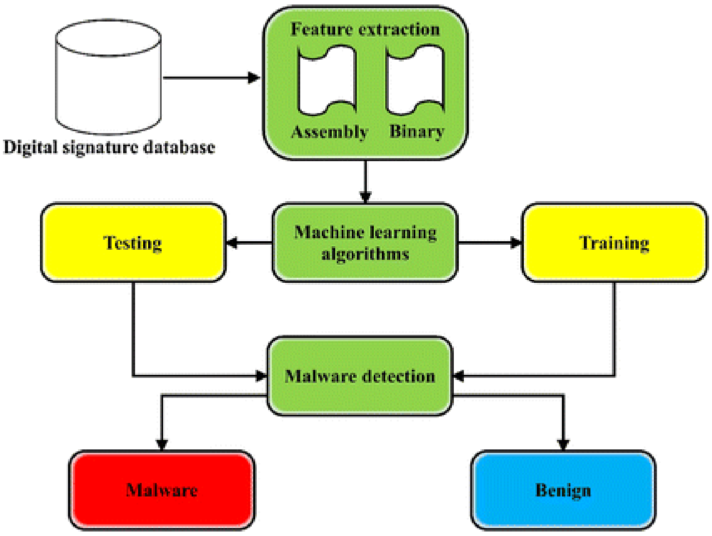

Figure 2. Machine learning based Malware detection [11]

Without a question, the most well-liked topic in the IT sector is machine learning. In addition to alternative web services companies, IBM, Microsoft, and several other IT service providers, as well as a multitude of streaming service providers, are extremely appealing in the machine learning space. The industry places emphasis on the lower cost of a machine learning platform [12]. Google, Amazon, as well as Yahoo have all alluded to a cloud computing strategy.

This accomplishment was made possible by its flexible computation online service, which launched in 2006 and provides a paid service that lets users rent devices and run their own applications. In that same year, Google made Google Docs functioning. Spreadsheets from Google and Written, two distinct components, served as the foundation for Google Docs at first. Writely, which lets renters transfer, edit, and copy papers via blogging platforms, was purchased by Google. Acquired in 2005 from two web innovations, Google Spreadsheets is an online tool that lets users create, update, edit, and share spreadsheets. Applications built with Ajax and compatible with Microsoft Excel are utilized. Spreadsheets should be saved in HTML format [13]. AT&T also entered the cloud computing space in 2006 when US internetworking became necessary. Apps for the software platform USI were available in over thirty states. In order to serve as a worldwide portal within its cloud, AT&T introduced Synaptic in 2008, combining the five US internet databases located in the USA, Europe, and Asia. With the aid of a speedier cloud, IBM introduced the IBM intelligent cloud platform in 2011. Subsequently, Apple unveiled ICloud, designed to house additional personally identifiable data (pictures, movies, audio, etc.). Microsoft started advertising the cloud in television this year, informing the public about its ability to store photos and videos for easy access. In 2012, Oracle introduced the Oracle cloud, which offers IaaS, PaaS, and SaaS, the three essential business components. The ultimate example of cloud technology today is Web 2.0, Google, Yahoo, Microsoft, and other internet providers that offer software-based business technology applications. Since machine learning has become a practical and easily available tool, a variety of people from different backgrounds—including hackers, students, and financial organizations—use virtual computers to complete their everyday tasks. This climate requires an implicit level of awareness to ensure efficiency [14].

3. Related Work

Data security in machine learning is a major concern, and various approaches are being put forth. Data security threat evaluates in machine learning are also being improved, and worries about privacy-related data storage issues are growing, as no private information can be recovered, unlike in the case of compromised emails. The studies listed below discuss several researchers' approaches to addressing cloud computing's data security issues and potential solutions [15].

Safety considerations, requirements, and questions that cloud service providers must solve in cloud engineering are covered in [16]. It covered defense-oriented countermeasures designed to lessen recognized hazards. The focus is primarily on outlier-based solutions, which work well with existing security systems rather than intrusion sensors. There are improvements in the pattern. Asserts that ML technology can reduce service costs while increasing company performance in [17]. To further this and encourage its use of the IT consumer sector, security concerns must also be addressed. They added that because internet services have data from several businesses and individuals saved in their systems, these platforms offer an entity target for cyber threats and illicit conduct. The analyst uses a machine learning inquiry to find security flaws and vulnerabilities, as well as to compile a list of five emerging attack types. 1. Service interruption 2: Malicious invader assaults 3. Network assault on the workstation side 4. Phishing attempts 5. Attacks using shared memory targets. They proposed an automatic threat classification system and verified its effectiveness by displaying threats within a designated cloud environment [18].

Eleven publications' worth of material about susceptible security threats are compiled in a paper found in [19]. The topic studied and the approach applied within the document to address the problem were computed by the researchers. It covers a few of the main machine-learning security issues and risk-reduction techniques. According to the researcher, the usage of machine learning will increase in the upcoming years as more businesses communicate data with different servers, potentially creating large-scale groups of attackers. He frequently asserts that by utilizing open source software from the outset of machine learning implementation, the likelihood of interoperability and data lock-in issues in the future can be reduced.

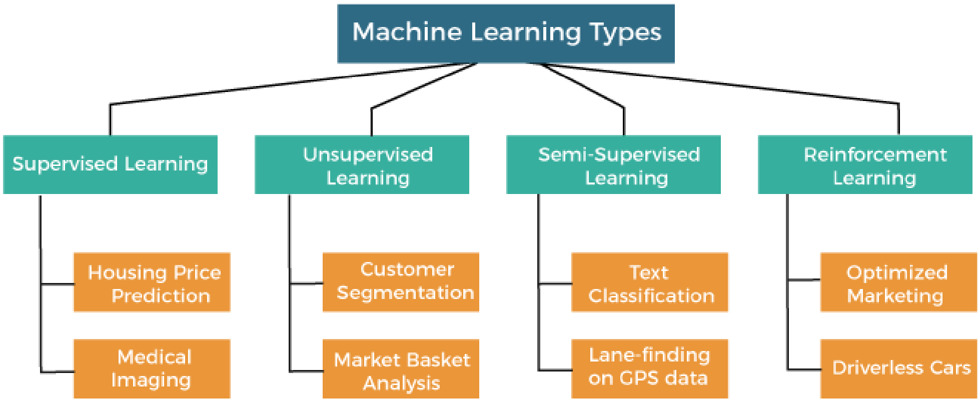

Figure 3. Types of Machine learning [20]

Listed among the benefits of CC in [21] are reduced maintenance costs, scalability, and market divergence. They contend that machine learning also gives users faster and easier service implementation while reducing complexity. Virtualization is the tactic utilized to engage with service quality. The advantages of adopting CC from a specific agenda are also covered in this. They added that CC has to implement a few restrictions. It examined security issues in machine learning as well as its methodology according to machine learning principles and deployed a few cloud computing platforms. The cloud architecture report served as the basis for the machine learning privacy framework. Additionally, we developed tools to enhance the data protection model's machine learning task.

4. Problem Statements

The organization uses machine learning to provide services, and the data is stored at the physical locations of vendors who are not within the organization's control. Therefore, security is crucial, or this cloud computing capability raised a number of security concerns and inquiries, such as data confidentiality, integrity, privacy, etc. Every organization required, or demanded, a reliable machine for instruction where the system as well as service providers protect data confidentiality. In order to foster that kind of confidence in machine learning, a system that is capable of performing authentication, verification, and encryption data transmission is needed; thus, the data's confidentiality must be maintained [22 & 23]. Confidentiality, Integrity, and Availability are the three fundamental concerns that need to be addressed in relation to the cryptographic terminology of CC security [24].

Because of its significant impact on the operations of cloud-based enterprises, timely availability of personally identifiable information is one of the most important security concerns in cloud computing. Let's take an example where a business organization or end user loses all access to cloud services and is unable to continue operating as a result of the CSP or cloud server's inability to deliver data to them in a timely manner.

The assurance that information remains unaltered while in transit and that only those who are genuinely permitted to access or alter it can do so is known as data integrity. In order to protect it further, it must not be distorted during transit, and measures must be taken to ensure that the information cannot be altered by an unauthorized person. To prevent unconstitutional access to data, confidentiality refers to ensuring that only authorized users may access the information [25].

5. Techniques Of Malware Analysis And Detection

The taxonomy of methods for malware analysis and detection is covered in this section. Malware detection begins with a deep taxonomy: each known detection approach is described in subcategories, and the relationship between each introduced detection category and the utilized data uses is determined. Malware analysis, on the other hand, is taxonomized and linked to the data types that are employed with each analysis approach. The malware detection and analysis taxonomy are displayed in Figure 3, along with the frequently used data types for each analysis strategy. The analysis approaches are provided as static, dynamic, and hybrid. Sub-detection techniques that delve deeper than the popular techniques—signal-based, behavioral-based, and heuristic-based—have been described in relation to malware detection techniques. As deep categories of the main detection approaches, static and dynamic signatures, ongoing, sequential, and typical behavioral patterns, as well as automated and manual rules, are presented. Each sub-detection strategy is linked to the most often used data types. The detection process, which defines the file's classification and, consequently, the overall precision of the detection models, is significantly and more impacted by malware analysis and data types. To identify and recognize the primary goal and function of the files under examination and then categorize them as malicious or benign files, a variety of data have been retrieved using static, dynamic, as well as hybrid analysis, including byte code, opcode, file data, registry data, and so forth. Determine the significance and trend of the various data kinds, we therefore go over the analysis techniques used in recent studies below along with the categories of data that were taken out of those studies.

5.1. Static

Examination With this method, software is examined without being run. One method for breaking down a process in a static evaluation is reverse engineering. Static evaluation matches the software to a large database of signatures by extracting the strings, entropy as, and image headers [26]. This technique has trouble with sophisticated and novel malware variants, despite its ability to swiftly identify and evaluate known malware.

Complex malware can be analyzed by sophisticated static evaluation techniques, but these procedures are frequently time-consuming and need a deep understanding of computing systems and disassembly.

5.2. Dynamic Analysis

This kind of study focuses on the characteristics of tactics that are carried out when the program is operating. Dynamic evaluation can be done in several ways, such as following information flow, tracking instructions, analyzing function parameters, monitoring function calls, and looking at AutoStart Extensibility Points. For dynamic malware analysis, modifications to files, registry entries, network activity, and access to RAM as well as hard drives can all be carefully examined. Since malware's behavior and attributes don't vary even when their structure does, dynamic evaluation is usually used for detecting new malware. However, contemporary malware has advanced in sophistication and can stop running when it recognizes behavioral analysis. Machine learning works better in these situations than both static and dynamic evaluation techniques [27].

5.3. Signature-Based Detection Method:

Similar to a fingerprint, a malware executable's signature is a unique characteristic. This method has been widely applied in antivirus software. It is not as good in detecting new viruses as other approaches, despite being faster.

5.4. Behaviour-Based Detection Method:

This technique assesses the behavior of a software to determine if it is malicious or not. Rather than examining the code or sequences, it examines the behaviors of the program. This kind of identification is made possible since the behavior remains constant or similar even when the code is altered. One significant drawback of behavior-based detection is that not all malware will display the same behaviors in a secure setting (such as a virtual machine or sandbox), which could lead to false negative results [28].

5.5. Typically, this method comprises three stages:

Removing behaviors, identifying characteristics by similar behavior, and classifying samples as benign or harmful using machine learning or data mining approaches.

5.6. Heuristic-Based Detection Method:

Heuristic approaches rely on machine learning techniques and experience to identify distinctive properties of pertinent programs. Heuristics can identify malware with similar characteristics by analyzing these features. The heuristic approach uses multiple samples to train the system, which then uses the learned system to detect malware in fresh software samples. Heuristic approaches can identify malware that signature- or behavior-based approaches might overlook, and they can even partially identify unknown malware. They might still have trouble identifying new or complex spyware, though [29].

5.7. Opcode Sequence Analysis:

Opcode sequences taken out of DEX files are utilized in studyas features for the malware detection program TinyDroid. Analysis of the operation code (opcode) sequence that makes up the executable's minimal machine instructions is known as opCode Sequence Analysis. Researchers can improve malware detection by studying these sequences to find trends and characteristics that point to malicious behavior.

5.8. API Call Analysis:

The research presented in papers [30] focuses on using API call analysis to analyze the local maliciousness of harmful programs. An operating system or library's API calls are the means by which an application can communicate with the system or other software elements. Through the analysis of API calls generated by an executable, researchers can discern trends and characteristics that could potentially point to malevolent actions.

5.9. System Call Analysis:

System calls made by an executable are frequently analyzed as part of heuristic-based detection techniques. Applications use system calls to ask the operating system for functions like memory management, file access, and process control. Through the analysis of system call patterns, experts are able to discern characteristics that point to malevolent activity.

5.10. Control Flow Graph Analysis:

Control flow graph (CFG) analysis is another feature extraction method utilized in heuristic-based malware detection. CFGs show the several paths that can be taken during program execution and describe the flow of control within a program. Researchers can spot trends and characteristics that point to malevolent activity by examining the CFG structure of a program.

5.11. Hybrid Feature Extraction:

To increase the efficacy of malware detection, researchers occasionally decide to combine several feature extraction methods. This method, called hybrid feature extraction, can use more than one feature extraction technique at once or combine static and dynamic analytic methods [31].

6. Solution To Data Security Challenges

Cryptography is suggested as a more secure method of protecting documents. Data can be stored on a cloud server more easily until files need to be encrypted. A certain community member should be granted permission by the data owner so that they can easily access the details. Heterogeneous info-centric authentication must be used to incorporate data access control. A data protection blueprint needs to include provisions for user safety, data recovery, data encryption, data integrity, authentication, and enhanced cloud data protection. The usage of data encryption as a service is recommended to guarantee data confidentiality and privacy. Use encryption to render data completely useless and to prevent other users from accessing it. Standard encryption can complicate accessibility. If the data is saved on backup disks and the keywords in the files don't change, users are urged to check before uploading the data to the cloud. By computing the file's hash, you may make sure that its information is unaltered until it is transferred to cloud servers. Although it is exceedingly difficult to keep, this hash computation can be used to check the integrity of records. SaaS ensures that all restrictions must be clear from the outset in order to separate data from different users at the device and physical levels. A distributed access management architecture might be [32].

It is utilized for access control in cloud computing. Stronger regulations apply when using passwords with attribution to identify unauthorized users. You can use permissions as a service to alert the client that they have access to that particular portion of the data. The owner can delegate the majority of computer-intensive tasks to the cloud with the aid of the fine-grained access management scheme, all while keeping the data material and servers hidden. A data-driven framework can be created and shared with cloud users for reliable data collection. A network-based prevention architecture is used to track threats in real-time [33].

7. Prons And Cons Of Machine Learning

The newest type of web-based computing network is machine learning, which provides users with flexible and easy-to-use resources to access various cloud apps. By connecting the cloud application to the internet, machine learning provides a means of remotely storing and accessing cloud data [34]. By choosing cloud services, users can archive their Metadata on a cloud-based server [35]. Cloud service providers can access or manage the data kept in a data center in the cloud. Thus, data collection for information processing within a cloud data center needs to be done as professionally as possible. Machine learning is very motivating due to its versatility, efficiency, usefulness, and cost-saving features. Although this is the newest and most promising technology, there are a lot of risks involved. Compared with the traditional IT approach, machine learning offers a multitude of potential advantages. But from the perspective of the user, concerns over machine learning security represent a significant barrier to its adoption. Machine learning is the capacity to access computer network resources, mostly for data storage and processing power, without the need for explicit user intervention. Using the services of cloud service providers, information from the cloud is processed as well as accessed on a web server. As a result, machine learning is becoming more valuable, expanding as a market, and drawing a lot of interest from the corporate and educational communities [36].

However, there are other challenges that arise when using machine learning. In this part, we have identified a number of challenges pertaining to data security in machine learning. It is imperative that we prioritize privacy and data security when utilizing the internet-cloud platform. An organization's credibility and reputation might be negatively impacted by data exposure or information loss. Data leak prevention is thought to be the the most important issue, accounting for 88% of the significant issues. Similarly, privacy and data remoteness have a 92% impact on security issues. The most important security issues with machine learning are: secrecy and authentication; availability; honesty; dependability; lack of assets and knowledge, etc. Six phases make up the data life cycle: Build, Use, Transfer, Archive, and Delete [37].

8. Discussion

One of the newest trends that is starting to gain traction in the search field right now is machine learning. With this technology, users may manage their resources anywhere, at any time. Cloud is regarded as a conscientious and noteworthy company that entered the information technology industry. As a result, the information technology industry is pushing for a shift to machine learning (ML), which calls for taking several important issues including security into account. Additionally, in order to save expenses and boost productivity, creatives must use ML. Here, we have discussed the present issues. Accurate reevaluation of solutions' appropriateness for clouds is necessary. We have effectively summarized our challenges along with their limitations and recommended models in the table. We can use encryption to safeguard our data from nefarious users. This also covers several types of encryption. We can protect our private data by using encryption. Next, in our article, we address several cloud security concerns and offer remedies. There must be several forms of security, such as software and information security. This study demonstrates how ML offers the ability to employ resources from resource pools, which contributes to the decrease of e-waste. In our paper, we list a few advantages of cloud computing that can be highly beneficial. Numerous ML-related issues, such as assurance, privacy, achievement, possession, and other non-technical concerns, need to be investigated in the upcoming research. As a result, academics face several challenges and must find solutions for both technical and non-technical issues.

9. Conclusion

Run-time malware attacks have the potential to threaten data privacy when sensitive data is transferred across hyper-connected networks. This study includes a thorough literature analysis of attack pattern taxonomy and machine learning-based malware detection and classification techniques in order to safeguard data against malware threats. Here, we've covered a few key data security concerns in machine learning, their solutions, and the advantages and disadvantages of the technology. This study demonstrates how ML offers the ability to employ resources from resource pools, which contributes to the decrease of e-waste. Future research must assess a wide range of cloud-related concerns, including those about profitability, property, efficiency, protection, privacy, and other non-technical issues. As a result, research teams encounter numerous challenges and must look into both technology and non-technical problem solutions. The security issues have to be thoroughly investigated.

References:

[1] M. Alaeiyan, S. Parsa, and M. Conti, “Analysis and classification of context-based malware behavior,” Comput. Commun., vol. 136, pp. 76–90, Feb. 2019, doi: 10.1016/j.comcom.2019.01.003.

[2] H. El Merabet and A. Hajraoui, “A survey of malware detection techniques based on machine learning,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 1, pp. 366–373, 2019, doi: 10.14569/IJACSA.2019.0100148.

[3] F. Mira, A. Brown, and W. Huang, “Novel malware detection methods by using LCS and LCSS,” 2016 22nd Int. Conf. Autom. Comput. ICAC 2016 Tackling New Challenges Autom. Comput., pp. 554–559, Oct. 2016, doi: 10.1109/ICONAC.2016.7604978.

[4] Z. Markel and M. Bilzor, “Building a machine learning classifier for malware detection,” WATeR 2014 - Proc. 2014 2nd Work. AntiMalware Test. Res., Jan. 2015, doi: 10.1109/WATER.2014.7015757.

[5] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, “A Survey on Machine Learning Techniques for Cyber Security in the Last Decade,” IEEE Access, vol. 8, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[6] A. Damodaran, F. Di Troia, C. A. Visaggio, T. H. Austin, and M. Stamp, “A comparison of static, dynamic, and hybrid analysis for malware detection,” J. Comput. Virol. Hacking Tech. 2015 131, vol. 13, no. 1, pp. 1–12, Dec. 2015, doi: 10.1007/S11416-015-0261-Z.

[7] S. Euh, H. Lee, D. Kim, and D. Hwang, “Comparative analysis of low-dimensional features and tree-based ensembles for malware detection systems,” IEEE Access, vol. 8, pp. 76796–76808, 2020, doi: 10.1109/ACCESS.2020.2986014.

[8] E. Raff et al., “An investigation of byte n-gram features for malware classification,” J. Comput. Virol. Hacking Tech., vol. 14, no. 1, pp. 1–20, Feb. 2018, doi: 10.1007/S11416-016-0283-1/TABLES/14.

[9] Y. Nagano and R. Uda, “Static analysis with paragraph vector for malware detection,” Proc. 11th Int. Conf. Ubiquitous Inf. Manag. Commun. IMCOM 2017, Jan. 2017, doi: 10.1145/3022227.3022306.

[10] A. Pektaş and T. Acarman, “Classification of malware families based on runtime behaviors,” J. Inf. Secur. Appl., vol. 37, pp. 91–100, Dec. 2017, doi: 10.1016/J.JISA.2017.10.005.

[11] A. Namavar Jahromi et al., “An improved two-hidden-layer extreme learning machine for malware hunting,” Comput. Secur., vol. 89, 2020, doi: 10.1016/j.cose.2019.101655.

[12] R. Damaševičius, A. Venčkauskas, J. Toldinas, and Š. Grigaliūnas, “Ensemble-based classification using neural networks and machine learning models for windows pe malware detection,” Electron., vol. 10, no. 4, pp. 1–26, 2021, doi: 10.3390/electronics10040485.

[13] E. Amer and I. Zelinka, “A dynamic Windows malware detection and prediction method based on contextual understanding of API call sequence,” Comput. Secur., vol. 92, p. 101760, May 2020, doi: 10.1016/J.COSE.2020.101760.

[14] P. V. Shijo and A. Salim, “Integrated Static and Dynamic Analysis for Malware Detection,” Procedia Comput. Sci., vol. 46, pp. 804– 811, Jan. 2015, doi: 10.1016/J.PROCS.2015.02.149.

[15] R. J. Mangialardo and J. C. Duarte, “Integrating Static and Dynamic Malware Analysis Using Machine Learning,” IEEE Lat. Am. Trans., vol. 13, no. 9, pp. 3080–3087, Sep. 2015, doi: 10.1109/TLA.2015.7350062.

[16] N. Kumar, S. Mukhopadhyay, M. Gupta, A. Handa, and S. K. Shukla, “Malware classification using early stage behavioral analysis,” Proc. - 2019 14th Asia Jt. Conf. Inf. Secur. AsiaJCIS 2019, pp. 16–23, Aug. 2019, doi: 10.1109/ASIAJCIS.2019.00-10.

[17] D. Gupta and R. Rani, “Improving malware detection using big data and ensemble learning,” Comput. Electr. Eng., vol. 86, p. 106729, Sep. 2020, doi: 10.1016/J.COMPELECENG.2020.106729.

[18] A. Damodaran, F. Di Troia, C. A. Visaggio, T. H. Austin, and M. Stamp, “A comparison of static, dynamic, and hybrid analysis for malware detection,” J. Comput. Virol. Hacking Tech. 2015 131, vol. 13, no. 1, pp. 1–12, Dec. 2015, doi: 10.1007/S11416-015-0261-Z.

[19] W. Han, J. Xue, Y. Wang, Z. Liu, and Z. Kong, “MalInsight: A systematic profiling based malware detection framework,” J. Netw. Comput. Appl., vol. 125, pp. 236–250, Jan. 2019, doi: 10.1016/j.jnca.2018.10.022.

[20] D. Baysa, R. M. Low, and M. Stamp, “Structural entropy and metamorphic malware,” J. Comput. Virol. Hacking Tech. 2013 94, vol. 9, no. 4, pp. 179–192, Apr. 2013, doi: 10.1007/S11416-013-0185-4.

[21] M. Wojnowicz, G. Chisholm, M. Wolff, and X. Zhao, “Wavelet decomposition of software entropy reveals symptoms of malicious code,” J. Innov. Digit. Ecosyst., vol. 3, no. 2, pp. 130–140, Dec. 2016, doi: 10.1016/j.jides.2016.10.009.

[22] K. Rieck, P. Trinius, C. Willems, and T. Holz, “Automatic analysis of malware behavior using machine learning,” J. Comput. Secur., vol. 19, no. 4, pp. 639–668, Jan. 2011, doi: 10.3233/JCS-2010-0410.

[23] D. Uppal, V. Mehra, and V. Verma, “Basic survey on Malware Analysis, Tools and Techniques,” Int. J. Comput. Sci. Appl., vol. 4, no. 1, pp. 103–112, 2014, doi: 10.5121/ijcsa.2014.4110.

[24] H. S. Galal, Y. B. Mahdy, and M. A. Atiea, “Behavior-based features model for malware detection,” J. Comput. Virol. Hacking Tech., vol. 12, no. 2, pp. 59–67, Jun. 2016, doi: 10.1007/s11416-015-0244-0.

[25] Y. Ki, E. Kim, and H. K. Kim, “A novel approach to detect malware based on API call sequence analysis,” Int. J. Distrib. Sens. Networks, vol. 2015, Jun. 2015, doi: 10.1155/2015/659101.

[26] G. Liang, J. Pang, and C. Dai, “A Behavior-Based Malware Variant Classification Technique,” Int. J. Inf. Educ. Technol., vol. 6, no. 4, pp. 291–295, 2016, doi: 10.7763/IJIET.2016.V6.702.

[27] L. Xiaofeng, J. Fangshuo, Z. Xiao, Y. Shengwei, S. Jing, and P. Lio, “ASSCA: API sequence and statistics features combined architecture for malware detection,” Comput. Networks, vol. 157, pp. 99–111, Jul. 2019, doi: 10.1016/j.comnet.2019.04.007.

[28] B. Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, “Deep learning for classification of malware system call sequences,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9992 LNAI, pp. 137–149, Dec. 2016, doi: 10.1007/978-3-319-50127-7_11.

[29] A. A. E. Elhadi, M. A. Maarof, and B. I. A. Barry, “Improving the detection of malware behaviour using simplified data dependent API call graph,” Int. J. Secur. its Appl., vol. 7, no. 5, pp. 29–42, 2013, doi: 10.14257/ijsia.2013.7.5.03.

[30] W. Mao, Z. Cai, D. Towsley, and X. Guan, “Probabilistic Inference on Integrity for Access Behavior Based Malware Detection,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9404, pp. 155–176, 2015, doi: 10.1007/978-3-319-26362-5_8.

[31] M. Kruczkowski and E. Niewiadomska-Szynkiewicz, “Support vector machine for malware analysis and classification,” Proc. - 2014 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. - Work. WI-IAT 2014, vol. 2, pp. 415–420, Oct. 2014, doi: 10.1109/WIIAT.2014.127.

[32] A. Mohaisen, O. Alrawi, and M. Mohaisen, “AMAL: High-fidelity, behavior-based automated malware analysis and classification,” Comput. Secur., vol. 52, pp. 251–266, Jul. 2015, doi: 10.1016/j.cose.2015.04.001.

[33] P. Vadrevu and R. Perdisci, “MAXS: Scaling malware execution with sequential multi-hypothesis testing,” ASIA CCS 2016 - Proc. 11th ACM Asia Conf. Comput. Commun. Secur., pp. 771–782, May 2016, doi: 10.1145/2897845.2897873.

[34] D. Bekerman, B. Shapira, L. Rokach, and A. Bar, “Unknown malware detection using network traffic classification,” 2015 IEEE Conf. Commun. NetworkSecurity, CNS 2015, pp. 134–142, Dec. 2015, doi: 10.1109/CNS.2015.7346821.

[35] D. Arivudainambi, V. K. Varun, S. C. S., and P. Visu, “Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance,” Comput. Commun., vol. 147, pp. 50–57, Nov. 2019, doi: 10.1016/j.comcom.2019.08.003.

[36] M. Rabbani, Y. L. Wang, R. Khoshkangini, H. Jelodar, R. Zhao, and P. Hu, “A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing,” J. Netw. Comput. Appl., vol. 151, p. 102507, Feb. 2020, doi: 10.1016/j.jnca.2019.102507.

[37] M. Ghiasi, A. Sami, and Z. Salehi, “Dynamic VSA: a framework for malware detection based on register contents,” Eng. Appl. Artif. Intell., vol. 44, pp. 111–122, Sep. 2015, doi: 10.1016/j.engappai.2015.05.008.