IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 10, Issue 3, March 2023

DOI: 10.17148/IARJSET.2023.10309

Machine Learning Algorithms for Cloud Computing Security: A Systematic Review

Sagar Mal Nitharwal[1]

Computer Science, Malaviya National Institute of Technology, Jaipur, India[1]

Computer Science, BTKIT Dwarahat, Almora, India[1]

Abstract: Cloud computing, also known as CC, refers to the on-demand availability of network resources, most notably data storage and processing power, which does not require any additional or direct administration on the part of the users. CC has recently surfaced as a collection of public and private data centres that provide customers with a unified online platform regardless of location. The term "edge computing" refers to a new computing paradigm that moves computation and information storage closer to the end users of a network to speed up reaction times and increase available transmission capacity. Mobile processing Cloud (MCC) is an approach to mobile application delivery that uses distributed processing. The rapid acceptance of computing models is slowed down by the fact that cloud computing (CC) and edge computing (edge computing) have security issues. These security issues include a vulnerability for clients and association acknowledgment. The study of computer techniques that can learn and become more efficient on their own through exposure to new data is known as machine learning (ML). In this review article, we present an analysis of CC security threats, issues, and solutions that utilized one or more ML algorithms. These solutions were implemented to combat these threats. We look at the various machine learning techniques, such as supervised, unsupervised, semi-supervised, and reinforcement learning, that are utilized to solve the problems associated with cloud security. After that, we evaluate the effectiveness of each method by contrasting its characteristics, as well as its positive and negative aspects. In addition to this, we outline potential future research directions to protect CC models.

Keywords: cloud security; security threats; cybersecurity; machine learning; network-based attacks; storage-based attacks.

I. INTRODUCTION

Cloud computing (CC) is a relatively new paradigm that has recently emerged as a new framework for facilitating and delivering extended this methodology through virtualization innovation to streamline sending and operating a more extensive scope of use on edge servers. CC was developed in response to a growing need to facilitate and deliver these services. The distributed concept of this paradigm presents a change in the distributed computing security strategies currently in use. In addition, information should be encrypted using one-of-a-kind encryption systems because it may pass through several distributed hubs affiliated with the web before ultimately being stored in the cloud. Edge nodes could also be asset-required devices, limiting the available options for security strategies. It is possible to shift the accountability for information away from the service providers and onto the end users if the information is maintained at the network's edges. Activities provided through the use of the Internet [1].

Typical financial constraints and rising computational costs have necessitated increased data storage, analysis, and presentation, significantly changing today's cloud paradigm [2]. CC refers to the accessibility of end-users resources on demand, particularly information storage and processing capacity, and does not require a direct or special organization from the customer. The term "distributed computing" is widely used but can have a variety of connotations depending on whom you talk to. Distributed computing provides clients access to public and private data on a singular platform accessible via the internet [3]. However, CC is plagued by many security issues that prevent it from being rapidly adopted as a computing paradigm. One of these issues is a vulnerability for clients and associations. Edge computing is a CC variant used to process time-sensitive data. It provides application developers and service providers with distributed processing capability at the system's periphery. The most recent edge processing ML techniques address security concerns and ensure more effective data management [4]. ML refers to the application of artificial intelligence, which permits systems to naturally take in new information and improve themselves without needing to be explicitly customized. ML is primarily concerned with developing computer programs that can determine an appropriate learning rate and use it to educate themselves [5].

The strategy for learning begins with perceptions or data, such as models, direct understanding, or heading, to the channel for structures in information and select better decisions later on the subject of the provided models. This paper aims to

DOI: 10.17148/IARJSET.2023.10309

analyze the legal problems and security threats posed by distributed computing using ML algorithms as the primary research methodology. The ever-increasing demand for distributed computing services is a key factor contributing to the concept's emergence as a candidate for consideration as a business strategy that could reduce the cost of infrastructure and operations. As a result, effectively managing the safety and privacy risks present in a distributed environment is essential, as is correctly addressing the associated tool fault issues [6]. The analysis and discussion of the safety problems and issues associated with distributed computing using ML algorithms and the valid steps to explain such issues are presented here.

II. CLOUD THREATS

Cloud Security threats

The primary concerns regarding CC's security can be broken down into three categories: confidentiality, integrity, and availability. These topics are touched on briefly in this section.

1. An insider threat to client information, the risk of an attack from the outside, and data problems are all included in the category of confidentiality threats. The first significant security challenge is an insider risk to customer data, which refers to unauthorized or unlawful access to customer data by an employee working for a cloud service provider. This type of access poses a threat to customer data. Second, the possibility of an assault from the outside is becoming more significant for cloud applications hosted in an unsecured area. This danger pertains to cloud clients and applications subjected to remote software or hardware attacks. Third, information leakage is an unlimited risk to cloud bargain data due to human error, a lack of instruments, and secured access failures, after which anything is conceivable. This risk is caused by human error, lack of instruments, and failures.

2. The risks to information integrity include the risks of information separation, poor customer access control, and the risk of information quality being compromised. First, there is the possibility of information isolation, which incorrectly combines the definitions of security parameters, unwise design of virtual machines (VMs), and off-base client-side hypervisors. This is a complicated problem within the cloud, which provides the assets connecting the clients; if the assets change, it could impact the trustworthiness of the information. Next up is poor client access control, which, due to inefficient access and character control, has various problems and threats that permit attackers to damage information assets.

3. The impact of progress on the board, the inaccessibility of the organization, the interruption of the physical availability of assets, and ineffective recovery strategies are all examples of availability threats. The first effect is the effect of progress on the board, which incorporates the impact of the testing client entrance for various clients and the effect of changes to the foundation. Alterations to hardware or software running in a cloud environment can harm an organization's ability to provide access to its resources. Next up is the inaccessibility of services, which includes the inaccessibility of domain name system (DNS) organizations registering software and assets. Additionally, the system's data transfer capability may not be available.

Attacks on the Cloud

1. Network-based attacks: Attacks based on network Port scanning, attacks using botnets, and spoofing attacks are the three kinds of system attacks discussed here. A port scan is beneficial to hackers and of considerable interest to them because it allows them to evaluate an attacker and collect information necessary to initiate an attack successfully. Regarding port scanning, attackers typically conceal their identities, while network defenders don't bother to do so. This distinction is based on whether or not the network’s defense regularly searches ports. A botnet is a collection of web-connected devices that have been infected with software and that are vulnerable to attack by cybercriminals. When a hacker or piece of malevolent software impersonates data to carry out operations on behalf of another user (or system), this is known as a spoofing attack. It occurs when an intruder purports to be someone else (or another machine, such as a phone) on a network to manipulate other machines, devices, or people into engaging in real activities or handing over sensitive information.

2. Attacks based on virtual machines (VMs) have numerous security flaws caused when numerous VMs are hosted on the same framework. An intrusion based on data relating to implementing a computer process rather than defects in the code itself is referred to as a side-channel assault. Any malicious code added to the virtual machine picture will be replicated when the VM is created from that image. Virtual machines (VMs) picture the executive’s structure, which provides separating and filtering for identifying and recovering from security threats.

3. Attacks based on storage: If a stringent monitoring mechanism is not considered, then the perpetrators of the attack will steal the essential data stored on some storage devices. The term "data scavenging" refers to the incapacity to remove data from storage devices so that it cannot be accessed or recovered by the malicious party that placed the data there. "data de-duplication" applies to repeated data copied more than once. This attack can be defended by ensuring that file duplication only occurs after the exact number of file duplicates has been chosen.

4. Attacks based on the application operating on the cloud may be subject to many attacks that can hinder its performance and cause the information to be leaked for nefarious reasons. Malware infusion attacks, stenography attacks, shared designs, attacks on online services, and convention-based attacks are the three most common application-based assaults.

III. ML ALGORITHMS FOR THE CLOUD SECURITY

In this section, we study different ML algorithms used to overcome the security issues in CC.

1. Supervised Learning

Supervised learning is the ML task of learning a limit that maps a commitment to a yield based on model data yield sets. It infers a limit from data involving many planning models. Supervised ML algorithms are those algorithms that require outside help.

2. Supervised ANNs

ANNs are the bits of a computing framework intended to recreate how the human mind analyzes and processes data. They are the establishments of ML that solve issues otherwise impossible or troublesome for humans or statistical principles. Hussin et al. [7] predicted basic distributed computing security issues using ANN algorithms. An ANN algorithm was used to determine security issues in a banking organization. ANNs were used for improving the execution and learning neural capacities. Levenberg-Marquardt (LMBP) algorithms were used to predict the presentation for the cloud security level. LMBP is a nonlinear improvement model used to measure the exactness of the forecasts present and decrease the error between genuine yields and focus for the preparation procedure; the mean square error (MSE) is estimated to decide the presentation. The cloud Delphi procedure was used for informal social events and investigation. The Delphi strategy was used to collect information as qualified sources. The ANN algorithm was used as the measurable information model to forecast distributed computing issues. The LMBP algorithm was utilized for predicting cloud security issues. In the CC security issue with banking organizations, LMBP algorithms have been confirmed to be extremely productive for testing and preparation systems.

3. K-NN

K-NN is likely the most straightforward algorithm among the ML algorithms for relapse and classification issues. The K-NN algorithm uses information and characterizes new information based on similarity measures (e.g., distance). A larger part vote finishes classification to its neighbors. The security of information in the cloud remains challenging. Different frameworks, such as data encryption, are being used to enhance cloud data security. The methodologies of information security cannot be applied. Comprehending the necessities of security is fundamental to the legitimate use of these measures. Zardari et al. [8] proposed a data classification approach based on data confidentiality. The authors described a methodology of information grouping that depends on the security and protection of information. The K-NN method of information arrangement was executed in the cloud administrations and virtual conditions. The target of using K-NN incorporates grouping information based on their security prerequisites. The information was grouped into touchy and non-delicate (or open) information. The order of the information helped in the recognizable proof of the information that is intended to be ensured. Only the touchy and non-open information groups were required to be ensured. The order of security and privacy-based information were proposed using a model for distributed computing. An examination was performed on the arrangement of the information based on security needs. The commitment of this investigation is the information privacy order procedure using a K-NN classifier method of ML. The RSA calculation requires greater security and encryption of delicate and private information. The proposed model of the information order for the security of cloud information has been achieved in the cloud simulation test system.

4. Naive Bayes

In ML, Navies Bayes classifiers are a group of basic “probabilistic classifiers” that apply Bayes’ hypothesis with solid (naive) freedom suppositions between the highlights. They are among the least complex Bayesian system models.

Zekri et al. designed a distributed denial-of-service (DDoS) detection system based on the algorithm to mitigate the DDoS threat. The hidden innovations and legacy conventions contain bugs and vulnerabilities that can enable interruption by

DOI: 10.17148/IARJSET.2023.10309

attackers. Assaults, such as DDoS, cause serious harm and influence the performance of the cloud. DDoS assaults have become one of the fundamental dangers to security. A DDoS attack executes an assault by permitting an interloper to interact with a computerized PC organization. Infected with malware, PCs and different machines (e.g., IoT devices) transform into bots (or zombies). Then, the assailant has remote control over the bots, known as a botnet. The traditional intrusion detection techniques have limitations such as oversized false alarms, noise that reduces the capabilities of the IDS by generating the rate of a false alarm, and constant updating of software to track the new threats. ML methods are acquainted with calling attention to the dangers more productively than traditional IDS. Distinctive ML algorithms are used to identify the threat in a DDoS and perform a calculation that attempts to locate the most miniature decision tree. The decision tree created by C4.5 can be used for the order after investigations are the optimal technique for grouping. From the results of the detection of DDoS using C4.5, it was found that the detection rate is more than 98%; moreover, the greater the DDoS attack duration, the higher the detection rate using this algorithm.

IV. CONCLUSION

In this research, the most difficult problems associated with CC, namely security threats and attacks, were analyzed.

Various machine learning (ML) algorithms, such as ANNs, K-NN, Naive Bayes, SVM, K-Means, and SVD, were investigated as potential solutions to address the security problems that are present in CC. We went over a few different suggested methods for cloud security that made use of machine learning algorithms. We presented an analytical review and analysis of the suggested techniques, focusing on the benefits and drawbacks of each one. In addition to this, we presented several potential new lines of inquiry that require further investigation in the near and distant future.

REFERENCES

[1]. Lim, S.Y.; Kiah, M.M.; Ang, T.F. Security Issues and Future Challenges of Cloud Service Authentication. Polytech. Hung. 2017, 14, 69–89.

[2]. Borylo, P.; Tornatore, M.; Jaglarz, P.; Shahriar, N.; Cholda, P.; Boutaba, R. Latency and energy-aware provisioning of network slices in cloud networks. Comput. Commun. 2020, 157, 1–19.

[3]. Dang, L.M.; Piran, M.; Han, D.; Min, K.; Moon, H. A Survey on Internet of Things and Cloud Computing for healthcare. Electronics 2019, 8, 768.

[4]. Le Duc, T.; Leiva, R.G.; Casari, P.; Östberg, P.O. Machine Learning Methods for Reliable Resource Provisioning in Edge-Cloud Computing: A Survey. ACM Comput. Surv. 2019, 52, 1–39.

[5]. Callara, M.; Wira, P. User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures. In Proceedings of the 2018 International Conference on Applied Smart Systems, Médéa, Algeria, 24–25 November 2018; pp. 1–6.

[6]. Singh, S.; Jeong, Y.-S.; Park, J. A Survey on Cloud Computing Security: Issues, Threats, and Solutions. J. Netw. Comput. Appl. 2016, 75, 200–222.

[7]. Elzamly, A.; Hussin, B.; Basari, A.S. Classification of Critical Cloud Computing Security Issues for Banking Organizations: A Cloud Delphi Study. Int. J. Grid Distrib. Comput. 2016, 9, 137–158.

[8]. Zardari, M.A.; Jung, L.T.; Zakaria, N. K-NN classifier for data confidentiality in cloud computing. In Proceedings of the International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 3–5 June 2014; pp. 1–6.

[9]. Zekri, M.; El Kafhali, S.; Aboutabit, N.; Saadi, Y. DDoS attack detection using machine learning techniques in cloud computing environments. In Proceedings of the International Conference of Cloud Computing Technologies and Applications (CloudTech), Rabat, Morocco, 24–26 October 2017; pp. 1–7.