IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

Application of Machine Learning Approach to Cloud Security: A Review

Trupti S. Ghegade[1], Dr. Monika Rokade[2]

ME Research Scholar, Department of Computer Engineering [1]

Research Guide & Assistant Professor, Department of Computer Engineering [2]

Sharadchandra Pawar College of Engineering, Otur, Pune, India

Abstract: *The current paper is related to the application of the machine learning approach to cloud security. This is a review article. Some important papers are chosen to identify research gaps in the cloud security field. Cloud computing is becoming increasingly popular and widely used. Several businesses are investing in this field, either for their own use or as a service to others. The emergence of various security issues for both industry and consumers is one of the consequences of cloud development. Machine Learning is one method for securing the cloud (ML). ML techniques have been used in a variety of ways to prevent or detect Cloud attacks and security gaps. The primary goal of this research is to perform a thorough examination of the machine learning approaches used to address, identify, and prevent cloud security concerns and vulnerabilities. Finally, it is concluded that SVM is the most commonly used ML in both hybrid and standalone models. KDD and KDD CUP'99 are the most used datasets by previous researchers.*

Keywords: Machine Learning, Cloud Security, Cloud Computing, Cloud Development, KDD, KDD CUP'99

I. INTRODUCTION

Cloud computing is a technical advancement that provides information technology facilities, platforms, and applications as Internet services. It is seen as the realization of a long-held ideal known as "Computing for Use," and it is increasingly being adopted by businesses as private, public, or hybrid Clouds. Its major goal is to allow consumers to use and pay for what they want, providing on-demand software and infrastructure services. Although cloud computing is widely seen as a huge and positive IT infrastructure transition, substantial security work is still to be done to mitigate its flaws. Cloud security concerns and vulnerabilities must be recognized and addressed since a considerable quantity of personal and business data is stored in cloud data centers.

Cloud infrastructure is vulnerable to assaults since it employs common Internet protocols and virtualization methods. Traditional sources of these attacks include Address Resolution Protocol, IP spoofing, Denial of Service (DoS), and so on. They might potentially originate from somewhere else. Zero-day attacks, also known as unknown assaults, are regarded as a serious issue in the cyber security area. Traditional detection and prevention strategies are inefficient in dealing with these threats while simultaneously dealing with a big data flow. Machine Learning (ML) approaches are extremely useful for detecting attacks, whether traditional or zero-day. Machine learning is a collection of algorithms that can learn patterns from data and predict them.

To improve prediction, machine learning combines computer science and statistics. There are three forms of learning in machine learning: supervised, unsupervised, and semi-supervised. The classification model for supervised machine learning is built using categorized data that has been trained. Unsupervised learning techniques allow you to train a model without using any instructions. Nearest Neighbor, Nave Bayes, Decision Trees, Linear Regression, Support Vector Machines (SVM), and more techniques are available for each. Unsupervised algorithms like K-means clustering are an example. Deep Learning (DL) allows multi-layered computer models to learn data representations at different levels of abstraction. It has made substantial progress in a variety of areas, including image analysis, speech recognition, and text recognition.

The primary goal of this research is to perform a thorough examination of the machine learning approaches used to address, identify, and prevent cloud security concerns and vulnerabilities.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

II. LITERATURE SURVEY

2.1 Types of Attacks in Cloud Computing

In order to protect the Cloud from those threats and prevent any damage, the attacks that can be launched need to be identified and understood. The attacks most often discussed in cloud computing are the following: [1], [2], [3]

(a) Denial of Service (DoS) attack: This is an attempt to affect service availability for users. Distributed Denial of Services (DDoS) is used to launch DoS using multiple computers.

(b) Zombie attack: When an attacker floods the victim with requests from innocent hosts in the network. Such an assault interrupts Cloud's anticipated behavior, influencing the accessibility of Cloud services.

(c) Phishing attack: This is an attempt to manipulate and gain personal information from innocent people by redirecting them to a false link. At Cloud, an attacker may be hosting a Cloud service to hide the accounts and services of other cloud users via a phishing attack site.

(d) Man-in-the-Middle attack: Where an attacker is able to access the communication path between two users. An intruder can access information interactions between data centers in the Cloud There are other attacks such as Cloud malware injection attacks, breach of confidentiality, authentication attacks, attacks on virtualization, etc.

There are other attacks such as cloud malware injection attacks, breaches of confidentiality, authentication attacks, attacks on virtualization, etc.

2.2 Application of Machine Learning Methods of Attack Detection for Cloud Computing

We discovered that ML is utilized in a variety of methods for detecting Cloud attacks in the articles we collected. Traditional detection, which detects and warns users when an attack occurs, is one of the most used methods. Another method of detection is to evaluate the Cloud security for any gaps or weaknesses before the assault begins.

Table 1. List of Researchers Work on Similar Topics

Researchers	Year	Work Done by Researcher
1. H. Tabrizchi and M. Kuchaki Rafsanjani[4]	2020	This paper examines the numerous Cloud infrastructure components and emerging security and privacy challenges raised in cloud platforms. Besides, they provide a new classification of the current security technologies in this field
2. L. Alhenaki, A. Alwatban, B. Alamri, and N. Alari[5]	2019	This paper provides a review of the significant attacks against Cloud computing. Adding more, it offered solutions and possible countermeasures for comparative analysis
3. R. Kumar and R. Goyal[6]	2019	This paper provides an extensive survey that defines a coherent security taxonomy, risks, vulnerabilities, and counter measurement criteria. Furthermore, it emphasizes security issues in other relevant fields, such as trust-based security architectures, large-scale, IoT, SDN, and Network Function Virtualization (NFV)
4. L. B. Bhajantri and T. Mujawar[7]	2019	This paper addresses problems related to security in Cloud at the conceptual level, data level, and reviews the cloud identities and Cloud access control. Further, the paper discusses several methods applied to prevent or mitigate Cloud security threats.
5. K. V. Uma and E. S. Blessie[8]	2019	This research paper reviews malware on mobile devices and compares data mining algorithms to find the most accurate one among them.
6. D. Dave, N. Meruliya, T. D. Gajjar, G. T. Ghoda, D. H. Parekh, and R. Sridaran[9]	2018	This paper reviews Cloud computing deployment models, their issues, and the issues of service models as well.
7. I. Avdagic and K. Hajdarevic[10]	2018	This research study identifies current attack detection and prevention strategies based on ML, in addition to analyzing algorithms to get the most accurate algorithm.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

8. D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim[11]	2017	This paper provides an overview of deep learning techniques with a focus on network anomaly detection. It provides a background to the topic along with a literature review and an analysis to compare accuracies.
9. G. Nenvani and H. Gupta[12]	2016	This survey discusses Cloud architecture, types, risks and threats.
10. B. Nelson and T. Olovsson[13]	2016	The purpose of this research is to categorize and analyze, both in a quantitative and a qualitative way, big data papers related to security or privacy.
11. S. G. Kene and D. P. Theng[14]	2015	This research study presents an overview of Cloud intrusion attacks, types of systems, and analysis of existing techniques. One of those techniques is ML.
12. C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan[15]	2013	This paper discusses vulnerabilities, threats and attacks to Cloud computing in detail.
13. C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan[16]	2013	In this research paper, forms of intrusion that can threaten the integrity, confidentiality and availability of Cloud services, are discussed, along with the techniques for solving these threats.
14. A. Patel, M. Taghavi, K. Bakhtiyari, and J. C. Júnior[17]	2013	This is a thorough review of intrusion detection along with the methods and comparative analysis of their features.
15. M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi[18]	2012	This research study focused on Cloud computing gaps and identifying types of attacks and related solutions. ML techniques are used, and their accuracies are compared to determine the best choice.
16. M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi[19]	2011	This paper is divided to two parts. The first is a survey on Cloud computing, including the gaps and threats. The second proposes solutions using ML techniques.

III. RESULTS AND DISCUSSIONS

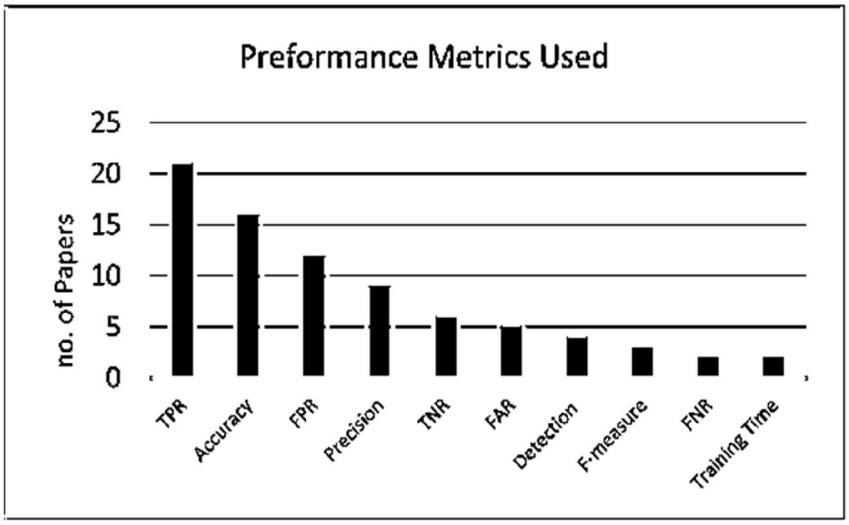

Figure 1. Performance Metrics Employed by Researchers

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

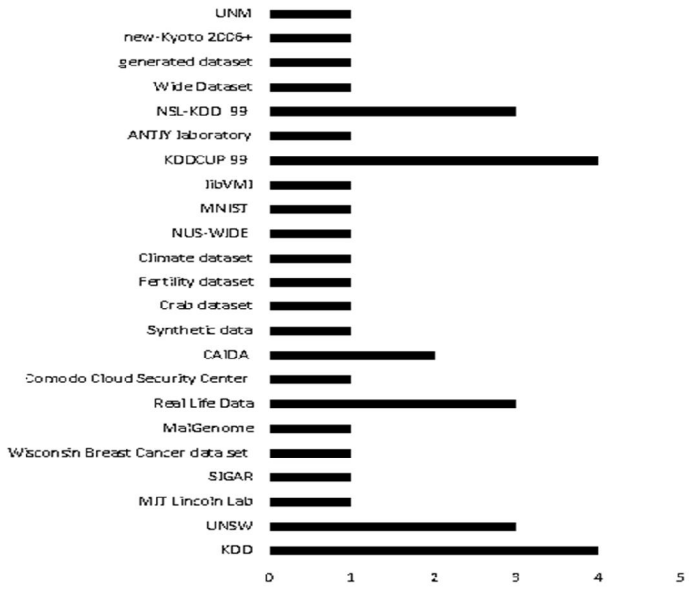

Figure 2. Data Sets Employed by Researchers

In their investigation, 20 articles included performance metrics. We ended up with more than 30 measures after gathering the metrics they utilized. As a result, we'll only include metrics that are utilized several times. True Positive Rate is the most commonly utilized of the 13 metrics we obtained. It's also known as detection rate, recall, or sensitivity. TPR is the value of accurately anticipated or categorized normal data. Accuracy is used by 16 studies to evaluate performance because it demonstrates the effectiveness of their machine learning model. Twelve studies utilize the False Positive Rate to describe when a value of normal data is wrongly predicted or categorized. The frequency with which a model properly predicts a good outcome is known as precision. Six articles employed the True Negative Rate, also known as specificity, to calculate the value of anticipated normal data. False Alarm Rate comes in second in terms of frequency of use, with five research using it. Following that are Detection value and Specificity, both of which are employed in four investigations. F-measure, also known as F-score or F-value, is the average of recall and accuracy. Only twice is the False Negative Rate utilized to compute the data that was incorrectly anticipated. Figure 1 shows the performance metrics employed by researchers.

Data sets are critical in arriving at a model evaluation and in achieving the optimal result. As a result, we undertook an assessment of the data sources utilized in the relevant publications, resulting in 36 datasets. The datasets that were used, as well as their frequency, are shown in Figure 2. KDD and KDD CUP '99 are the most popular datasets, with four research publications per dataset being utilized for assessment. The NSL KDD '99 dataset and Real datasets, which are utilized in three studies, are next. CADIA and UNSW each have two papers. The remaining datasets are utilized only once.

IV. CONCLUSION

The primary goal of this research is to perform a thorough examination of the machine learning approaches used to address, identify, and prevent cloud security concerns and vulnerabilities. Finally, it is concluded that SVM is the most commonly used ML in both hybrid and standalone models.KDD and KDD CUP'99 are the most used datasets by previous researchers.

ACKNOWLEDGMENT

The main author is wishing thank to the research guide Assistant Professor, Dr. Monika Rokade fromSharadchandra Pawar College of Engineering, Otur for her co-operation and technical help during the survey phase of my research work.

REFERENCES

[1] G. Nenvani and H. Gupta, ``A survey on attack detection on cloud using supervised learning techniques," in Proc. Symp. Colossal Data Anal. Netw.(CDAN), Mar. 2016, pp. 1_5, doi: 10.1109/CDAN.2016.7570872.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

[2] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, ``A survey on security issues and solutions at different layers of cloud computing," J. Supercomput., vol. 63, pp. 561_592, Oct. 2013, doi: 10.1007/s11227-012-0831-5.
[3] M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi, ``A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing," Future Gener. Comput. Syst., vol. 28, no. 6, pp. 833-851, Jun. 2012, doi: 10.1016/j.future.2012.01.006.
[4] H. Tabrizchi and M. Kuchaki Rafsanjani, ``A survey on security challenges in cloud computing: Issues, threats, and solutions," J. Supercomput., vol. 76, no. 12, pp. 9493-9532, Dec. 2020, doi: 10.1007/s11227-020-03213-1.
[5] L. Alhenaki, A. Alwatban, B. Alamri, and N. Alari, ``A survey on the security of cloud computing," in Proc. 2nd Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS), May 2019, pp. 1-7, doi: 10.1109/CAIS.2019.8769497.
[6] R. Kumar and R. Goyal, ``On cloud security requirements, threats, vulnerabilities and countermeasures: A survey," Comput. Sci. Rev., vol. 33, pp. 148, Aug. 2019, doi: 10.1016/j.cosrev.2019.05.002.
[7] L. B. Bhajantri and T. Mujawar, ``A survey of cloud computing security challenges, issues and their countermeasures," in Proc. 3rd Int.Conf. I-SMAC (IoT Social, Mobile, Anal. Cloud) (I-SMAC), Dec. 2019, pp. 376-380, doi: 10.1109/I-SMAC47947.2019.9032545.
[8] K. V. Uma and E. S. Blessie, ``Survey on Android malware detection and protection using data mining algorithms," in Proc. 2nd Int. Conf. I-SMAC (IoT Social, Mobile, Anal. Cloud) (I-SMAC)I-SMAC (IoT Social, Mobile, Anal. Cloud) (I-SMAC), 2nd Int. Conf., Aug. 2018, pp. 209-212, doi: 10.1109/I-SMAC.2018.8653720.
[9] D. Dave, N. Meruliya, T. D. Gajjar, G. T. Ghoda, D. H. Parekh, and R. Sridaran, ``Cloud security issues and challenges," in Advances in Intel-ligent Systems and Computing, vol. 654. New York, NY, USA: Springer, 2018, pp. 499-514, doi: 10.1007/978-981-10-6620-7_48.
[10] I. Avdagic and K. Hajdarevic, ``Survey on machine learning algorithms as cloud service for CIDPS," in Proc. 25th Telecommun. Forum (TELFOR), Nov. 2017, pp. 1-4, doi: 10.1109/TELFOR.2017.8249467.
[11] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, ``A survey of deep learning-based network anomaly detection," Cluster Comput.,vol. 22, pp. 949_961, Sep. 2017, doi: 10.1007/s10586-017-1117-8.
[12] G. Nenvani and H. Gupta, ``A survey on attack detection on cloud using supervised learning techniques," in Proc. Symp. Colossal Data Anal. Netw.(CDAN), Mar. 2016, pp. 15, doi: 10.1109/CDAN.2016.7570872.
[13] B. Nelson and T. Olovsson, ``Security and privacy for big data: A systematic literature review," in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2016, pp. 3693-3702, doi: 10.1109/BigData.2016.7841037.
[14] S. G. Kene and D. P. Theng, ``A review on intrusion detection techniques for cloud computing and security challenges," in Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS), Feb. 2015, pp. 227-232, doi: 10.1109/ECS.2015.7124898.
[15] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, ``A survey on security issues and solutions at different layers of cloud computing," J. Supercomput., vol. 63, pp. 561-592, Oct. 2013, doi: 10.1007/s11227-012-0831-5.
[16] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, ``A survey of intrusion detection techniques in cloud," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 42-057, 2013, doi: 10.1016/j.jnca.2012.05.003.
[17] A. Patel, M. Taghavi, K. Bakhtiyari, and J. C. Júnior, ``An intrusion detection and prevention system in cloud computing: A systematic review," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 25-41, Jan. 2013, doi:10.1016/j.jnca.2012.08.007.
[18] M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi, ``A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing," Future Gener. Comput. Syst., vol. 28, no. 6, pp. 833-851, Jun. 2012, doi: 10.1016/j.future.2012.01.006.
[19] M. T. Khorshed, A. B. M. S. Ali, and S. A. Wasimi, ``Trust issues that create threats for cyber attacks in cloud computing," in Proc. IEEE 17th Int. Conf. Parallel Distrib. Syst., Dec. 2011, pp. 900-905, doi: 10.1109/ICPADS.2011.156.