JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

A SURVEY OF MACHINE LEARNING-BASED SECURITY FOR CLOUD COMPUTING

Dr. Pankaj Kawadkar[1], Sunny Mohite[2]

[1]Asso. Prof., Sri. Satya Sai University of Technology and Medical Science, Sehore MP.

[2]Research Scholar, Sri. Satya Sai University of Technology and Medical Science, Sehore MP

kawadkarpankaj@gmail.com, sunnymohite2684@gmail.com

Abstract

A computing paradigm known as "cloud computing" offers end users on-demand, scalable, and measurable services. Nearly every firm in the modern day has huge reliance on this computing technology for cost-savings, infrastructure, development platforms, and data processing analytics of data, etc. The services that the cloud service offers end users can use providers (CSP) whenever they choose. using a web application and the internet, from anywhere. The cloud infrastructure's security must be top priority. Furthermore numerous technologically based research projects are used to give a better and more precise defence system to prevent cloud attacks. A type of machine learning technology that has demonstrated to yield better outcomes in in recent times, protecting the cloud environment. To develop models that can automate the process of identifying cloud threats with higher accuracy than any other technology, machine learning algorithms are trained on a variety of real-world datasets. This article examines some of the most recent studies that have used machine learning as a defence against cloud threats.

Keywords - Cloud Computing, Machine Learning, Intrusion Detection System, Datasets, Supervised Machine Learning, Unsupervised Machine Learning, Reinforcement Learning, Deep Neural Network.

1 INTRODUCTION

The main worry in the modern day is cloud environment security. Even large cloud service providers like Amazon, Google, and others that have adequate security procedures are subject to a number of cloud assaults that are periodically reported on a regular basis. Information security, identity security, network security, infrastructure security, and software security are the five basic categories into which cloud security can be divided. Cloud computing uses the service paradigm known as Machine Learning as a Service (MLaaS) to improve its defence against various cloud assaults. With the use of machine learning algorithms, a number of intrusion detection systems have been created, improving the accuracy of identifying attacks and permitting the continuation of efficient corporate operations.

2 Theoretical Setting

This section provides a quick overview of both machine learning and cloud computing by describing their historical development.

Submitted: 24/11/2022 **Accepted**: 09/12/2022

2.1 Cloud Computing

A computing paradigm known as "cloud computing" offers end customers on-demand, scalable, quantified, and secure services over the internet. These advantages are the reason why the cloud computing paradigm finds a very broad range of application cases. Today's market is filled with cloud service providers who provide a wide range of cloud services to their clients. Amazon Web Services (AWS), Microsoft Azure, IBM Cloud, Google Cloud, Oracle Cloud, Alibaba Cloud, and others are a few of them.

2.1.1 Characteristics of Cloud Computing

According to NIST, cloud computing primarily has five key characteristics. [1, 2]

- **On-demand self-service**- Cloud services are made available to end users on demand and without the involvement of the cloud service provider, according to the characteristic known as on-demand self-service.
- **Rapid elasticity:** Cloud-based applications can adapt to sudden changes in service demand without running out of resources or experiencing business interruptions.
- **Measured Service:** The services offered by the cloud service provider are billed on a metered basis, meaning that customers who utilise the services are only charged for the services they actually use and are free to quit using them at any time.
- **Broad network access**-Cloud services are accessible and can be used with a variety of thin clients, including smartphones, laptops, desktop computers, PDAs, and more.
- **Resource Pooling**-The cloud service provider pools resources such as memory, storage, processing power, network bandwidth, etc. in order to fulfil the requests of numerous clients.

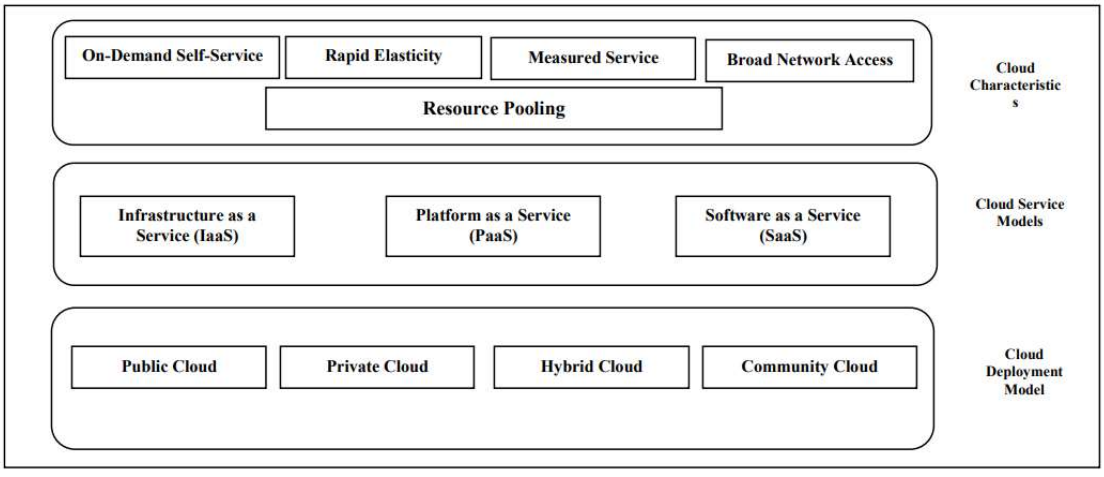

Fig. 1: Working Model of Cloud Computing

2.1.2 Service Models

IaaS: This service model provides the essential hardware needed to run the cloud application. These resources include virtual machines, storage, networks, and computing units. Because the

service provider covers the full cost of setup and maintenance, this service layer helps save money on the high cost of setting up and maintaining these resources. [3, 4]

PaaS-Using the underlying cloud architecture, PaaS provides the platform for developers to create apps. PaaS offers several technologies, programming languages, and other tools needed to create a cloud application. Although the end user has full control over the programme, they have no control over the underlying cloud infrastructure**.**

SaaS-This paradigm enables the end user to access programmes deployed to the cloud via the internet using a variety of clients. End users can only use a programme; they have no control over the cloud infrastructure or the application itself.

2.1.3 Deployment Model

Public Cloud: All end users who only need to use the public cloud-deployed applications are given access to it. Amazon EC2, Google App Engine, Microsoft Azure, and others are examples. [3, 5]

Private Cloud: The private cloud is entirely dedicated to the use of individual private companies for conducting business in a very private and secure manner without interference from third parties. Examples include the Microsoft ECI data centre, the Amazon VPC, the Ubuntu Enterprise Cloud, and Eucalyptus, among others.

Community Cloud- When all of the firms share cloud infrastructure, this deployment technique is used. Community cloud models may be in-site or outsourced depending on the need. Google Apps for Government and Microsoft Government Community Cloud are two examples.

Hybrid Cloud- Hybrid clouds are made up of a combination of other deployment models that are now accessible. An illustration may be VMware vCloud, etc.

2.1.4 Cloud Attacks

The cloud computing paradigm is vulnerable to various threats. Depending on the sort of attack, these can happen at IaaS, PaaS, or SaaS cloud service models. [6]. Some of the well-known cloud assaults are shown in Figure 2 at the relevant service models.

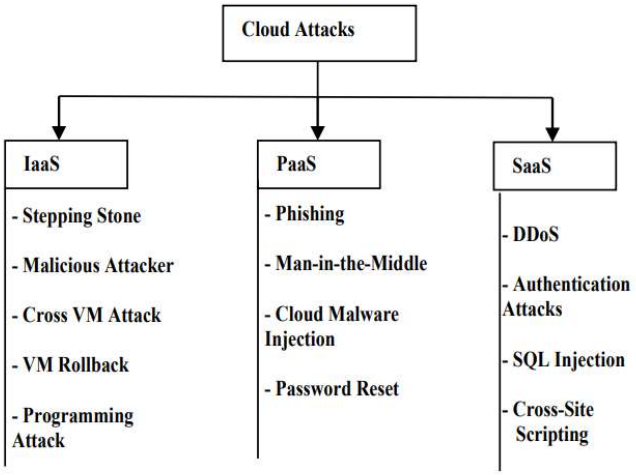

Fig. 2: Classification for Cloud Attacks

Reinforcement Machine Learning: Reinforcement ML uses a feedback-based learning paradigm. The agent is untrained in any area and learns from its own experience of supervised datasets and is rewarded or punished for choosing the right or wrong action in accordance. The categorization of machine learning algorithms is shown in Figure 2.

3 LITERATURE REVIEW

A machine learning-based intrusion detection system for cloud computing was proposed by Chkirbene et al. [13]. Although the classifier is the most crucial part of an intrusion detection system[26], the unbalanced nature of the datasets makes it impossible for it to perform with high classification accuracy. In this proposed solution, the weighted supervised decision tree algorithm is used as the classification algorithm to address this issue. The proposed approach generates low points for negative classification and high scores for positive classification, resulting in high accuracy for the classifier.

In their research paper, Bagga et al. [14] suggest another security framework that combines the SVM machine learning method, Network Function Virtualization, and Software Defined Network. This strategy highlights its significance as it successfully achieves security against various assaults for both NFV and SDN. There are two levels to the suggested structure. First, into "security enforcement plane," which is separated into three components: MA (Monitoring Agent), IB (Infrastructure Block), and CMB, and is in charge of providing security against both internal and external assaults in IoT. (Control and Management Block). Secondly, into the "security orchestration plane," where the security policies are configured in real time. An improved outcome is obtained in terms of accuracy, FRP, detection rate and training time as compared with other existing approaches.

The importance of data security in mobile cloud computing due to the involvement of heterogeneous network is depicted by Dey et al. [15] and an intrusion detection system that can handle such complex security constraints is thus proposed. KMeans and DBSCAN machine learning algorithm lays the foundation for such an IDS, which can guard defence against

heterogeneous attacks such as MITM as well as DDoS. This approach trains the system on cluster basis and does the traffic classification on the basis of distance calculation. Better accuracy results for the proposed IDS is achieved as their is a reduction in the complexity due to the non requirement of updates in the rules regularly.

Dey et al. [15] illustrate the significance of data security in mobile cloud computing due to the participation of heterogeneous networks, and a proposal is made for an intrusion detection system that can manage such complicated security limitations. Such an IDS can protect defence against heterogeneous threats like MITM and DDoS thanks to the K Means and DBSCAN machine learning algorithm. This method uses clusters to train the system and bases traffic classification on distance calculations. Due to the suggested IDS's reduced complexity from not needing to change its rules on a frequent basis, better accuracy results are obtained.

The Rabbani et al. [17] present another machine learning-based method for the CSP (cloud service provider) to monitor user activity in the cloud. The method makes use of the hybridization of PSO-PNN for the aim of identifying unauthenticated users in the cloud (particle swarm optimization and probabilistic neural network). By obtaining high accuracy in terms of true positive rate, false negative rate, f-measure, and precision, the results demonstrated the usefulness of the suggested hybrid method.

Machine learning is used by Hesamifard et al. [18] to protect privacy. Using homomorphic encryption, data is used to encrypt information for the neural network's training. Accurate polynomial approximations are used as an alternative to the neural network's typical sigmoid and ReLU (Rectified Linear Unit) activation functions. Comparing the suggested approach to SMC (secure multiparty computation) and HE, it provides privacy with greater accuracy (homomorphic encryption).

Singh et al. [19] use the mutual authentication protocol to secure machine learning-based data sharing over the cloud. Multiple sorts of cloud attacks, including DoS, DDoS, MITM, reply, etc., are readily protected against by the proposed mutual authentication protocol. Schnorr's signature and ECC (Elliptic curve cryptography) are used for the aim of encrypting data with the advantage of short keys, and voting classifiers are used to classify threats or assaults. The ProVerif tool's results demonstrate the excellent accuracy of the proposed methodology.

Salman and others [20] provided a research paper that advocated using an intrusion detection system in conjunction with machine learning to counteract various cloud threats in a multi-cloud setting. The intrusion detection system used in this proposed approach uses linear regression and random forest supervised machine learning algorithms. The key benefit of this strategy, aside from detecting cloud threats, is that it also ensures threat categorization using a cutting-edge step-by-step algorithm[23][24]. In terms of categorization and threat detection, accuracy is 99.0% and 93.6%, respectively.

A deep neural network-based intrusion detection system was proposed by Chiba et al. [21] by combining the genetic and simulated annealing techniques. The SAA algorithm achieves optimization in the search phase of the genetic algorithm while the upgraded genetic algorithm utilised by this approach reduces convergence as well as execution time. These algorithms

enhance DNN properties such as feature selection and activation function, hence boosting the deep neural network's overall performance.

Khilar et al. [22] suggest using machine learning-based authorisation to restrict access to cloud services to only authenticated users. The trust between service providers and end users increases as a result of the suggested approach's improved authorization system for cloud users and restrictions on unlawful access to cloud resources, and overall data security also advances. When compared to the established method for user access to cloud resources, the suggested strategy produced better results in terms of MAE, time, recall, precision, and f1-sccore.

At the cloud hypervisor, a learning algorithm is utilised in hybrid mode to find network anomalies. The hybrid model that is suggested operates on network traffic According to Sethi et alstudy .'s report [24], IDS based on reinforcement learning improves cloud security. The primary weakness of standard IDS for cloud security is inaccurate classification accuracy, which is addressed by this method's low FPR (false positive rate). The proposed model consists of three modules: the host network, which defends against attacks based on virtual machines, the agent network, which distinguishes between legitimate and malicious requests, and the administration network, which enables administrators to disable the offending virtual machines.

In order to provide cloud security, Chkirbene et al. [25] proposed a "EIDS" scheme for traffic analysis, in which past and present decisions are investigated, unwanted features are removed from the dataset, the data is clustered using the K-Means algorithm, and normal and malicious requests are distinguished using SVM and machine learning-based intrusion detection. To improve the efficiency of the intrusion detection system, decisions on the classification of attacks are compared between the present and the past. The detection rate of the supervised learning classifier has grown by 24% overall (almost 90%), which increases the security of IDS. Chkirbene and others [26] provided two models for trust-based IDS: the accelerated model and the classification model (TIDCS) (TIDCS-A). The former model handles dimensionality reduction so that only the necessary features from the UNSW dataset can be handled by the machine learning method, and the latter model handles anomaly detection. The simulation results clearly show that the proposed model, which uses TIDCS and TIDCS-A machine learning techniques, is capable of both attack classification and detection with greater accuracy.

Data tampering becomes a bigger issue when using machine learning in a distributed cloud setting. For the purpose of retraining the integrity of the data, Zhao et al. [27] established a verification approach for the data in distributed cloud environments called DML-DIV (distributed machine learning data integrity verification). Additionally, it is evident from the simulation findings that the suggested DML-DIV strategy outperforms the already used

comparative approaches in terms of privacy protection, forgery, and tampering attack.

Ref	ML Algorithm Used	Proposed Approach	Dataset
[13]	Decision Tree	Intrusion Detecting System based on weight optimization.	UNSW
[14]	Support Vector Machine	AI Framework based on the combination of ML, NFV, SDN.	NSL-KDD
[15]	K-Means and DBSCAN	Traffic filtration via distance calculation and training system via cluster basis.	Multiple datasets
[16]	Reinforcement Learning	Neuro-Fuzzy system for secure data offloading with PSO to select secure fog node in Fog-Cloud-IoT environment.	Multiple datasets
[17]	Multilayer Neural Network	Identification of unwanted user in the cloud with PSO and PNN.	UNSW-NB15
[18]	Deep Neural Network	Training NN with encrypted data and using accurate activation function for the NN.	Crab, Fertility and Climate Dataset
[19]	LR(Linear Regression) and KNN (K-Nearest neighbor)	ECC along with voting classifier for mutual authentication over multi cloud environment.	CICD
[20]	Linear Regression (LR) and Random Forest (RF)	Machine Learning based Intrusion detection system for detection of attacks in multi-cloud environment.	UNSW
[23]	K-Means clustering and SVM classification	The hybrid model is responsible for performing network traffic investigation, unwanted feature reduction from dataset, clustering the data with K-Means algorithm and classification between normal as well as malicious requests via SVM.	UNSW-NB15
[24]	Random Forest , Quadratic Discriminant Analysis, K-Nearest Neighbours, Gaussian Naive Bayes (GNB) and AdaBoost	Deep Reinforcement Learning model which has the host, agent and administrator network that predicts the affected virtual machines and also blocks them.	UNSW-NB15
[21]	Deep Neural Network	IGASAA, i.e. hybridization of genetic algorithm and simulated annealing algorithm for machine learning based network IDS.	CICIDS2017, NSL-KDD version 2015 and CIDDS-001
[25]	Linear Regression (LR)	EIDS for traffic analysis in which the past as well as current decisions are compared with each other using machine learning.	UNSW-NB-15
[26]	Decision Tree, Random Forest	Machine learning based classification TIDCS and detection TIDCS-A models for IDS.	NSL-KDD, UNSW
[22]	K-Nearest Neighbor, Decision Tree, Logistic Regression, Naive Bays	Authorization of the user is increased to provide better security to the cloud resources using machine learning approach.	User Dataset
[27]	DML	DML-DIV for retraining the integrity of the data in distributed cloud environment.	Advertisement Click Prediction

Table 1. Summary of related work

4 CONCLUSIONS AND FUTURE SCOPE

Customer data stored in the cloud is extremely important, and its security must never be compromised. The researchers deploy several new technologies and a variety of security techniques to increase the security of the cloud ecosystem. There is a lot of room for machine learning to improve accuracy and automate defence against both known and unidentified cloud assaults. The primary goal of this survey article is to provide readers with a current picture of machine learning-based research in the area of cloud security.

In order to provide more precise cloud data security in the future, we suggest an intrusion detection system that will use an improved and optimised machine learning algorithm.

REFERENCES

[1] Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W.: A Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies. IEEE Access, 9, pp. 57792-57807, 2021.
[2] Abdulsalam, Y.S., Hedabou, M.: Security and Privacy in Cloud Computing: Technical Review. Future Internet 2022, 14, 11.
[3] George, S.S., Pramila, R.S.: A review of different techniques in cloud computing. Materialstoday proceedings, 46, pp. 8002-8008, 2021.
[4] Attaran, M., Woods, J.: Cloud computing technology: improving small business performance using the Internet. Journal of Small Business & Entrepreneurship. 13. pp. 94-106, 2018.
[5] Basu, S., Bardhan, A., Gupta, K., Saha, P., Pal, M., Bose, M., Basu, K., Chaudhury, S., Sarkar, P.: Cloud computing security challenges & solutions-A survey. Annual Computing and Communication Workshop and Conference (CCWC), 2018.
[6] Dwivedi, R.K., Saran, M., Kumar, R.: A Survey on Security over Sensor-Cloud. In: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 31-37, 2019.
[7] Butt, U.A.; Mehmood, M.; Shah, S.B.H.; Amin, R.; Shaukat, M.W.; Raza, S.M.; Suh, D.Y.; Piran, M.J. A Review of Machine Learning Algorithms for Cloud Computing Security. Electronics 2020, 9, 1379.
[8] Sarker, I.H.; Machine Learning: Algorithms, Real-World Applications and Research Directions. SN computer science volume. 2, 3 (2021): 160.
[9] Alzubi, J., Nayyar, A., Kumar, A.: Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series, Volume 1142, Second National Conference on Computational Intelligence 2018, Bangalore, India.
[10] Baraneetharan, E.: Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey. Journal of Information Technology and Digital World, Vol. 02, pp. 161-173, 2020.
[11] Saranyaa, T., Sridevi, S., Deisy, C., Chung, T.D., Khan, M.K.A.: Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review. Procedia Computer Science, Vol 171, pp. 1251-1260, 2020.
[12] Sen, P.C., Hajra, M., Ghosh, M.: Supervised Classification Algorithms in Machine Learning: A Survey and Review. Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing, vol 937, pp. 99-111, 2019.
[13] Chkirbene, Z., Erbad, A., Hamila, R., Gouissem, A., Mohamed, A., Hamdi, M.: Machine Learning Based Cloud Computing Anomalies Detection. IEEE Network, Vol. 34, pp. 178-183, 2020.
[14] Bagaa, M., Taleb, T., Bernabe, J.B., Skarmeta, A.: A Machine Learning Security Framework for Iot Systems. IEEE Access, Vol. 8, pp. 114066-114077, 2020.
[15] Dey, S., Ye, Q., Sampalli, S.: A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks. Information Fusion, vol. 49, pp. 205-215, 2019.

[16] Alli, A.A., Alam, M.M.: SecOFF-FCIoT: Machine learning based secure offloading in Fog-Cloud of things for smart city applications. Internet of Things, Vol. 7, 2019.
[17] Rabbani, M., Wang, Y.L., Khoshkangini, R., Jelodar, H., Zhao, R., Hu, P.: A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing. Journal of Network and Computer Applications, Vol. 151, 2020.
[18] Hesamifard, E., Takabi, H., Ghasemi, M., Jones, C.: Privacypreserving Machine Learning in Cloud. Cloud Computing Security Workshop, pp. 39-43, 2017.
[19] Singh A.K., Saxena, D.: A Cryptography and Machine Learning Based Authentication for Secure Data-Sharing in Federated Cloud Services Environment. Journal of Applied Security Research, 2021.
[20] Salman, T., Bhamare, D., Erbad, A., Jain, R., Samaka, M.: Machine Learning for Anomaly Detection and Categorization in Multi-Cloud Environments. IEEE 4th International Conference on Cyber Security and Cloud Computing, 2017.
[21] Chiba, Z., Abghour, N., Moussaid, K., Elomri, A., Rida, M.: Intelligent approach to build a Deep Neural Network based IDS for cloud environment using combination of machine learning algorithms. Computers & Security, Vol. 86, pp. 291-317, 2019.
[22] Khilar, P.M., Chaudhari, V., Swain, R.R.: Trust-Based Access Control in Cloud Computing Using Machine Learning. Cloud Computing for Geospatial Big Data Analytics, pp. 55-79, 2018.
[23] Ram Kumar, Manoj Eknath Patil ," Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies", Turkish Journal of Computer and Mathematics Education ,Vol.12 No.14(2021), 6168-6174.
[24] Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, "A Survey Paper on Altered Fingerprint Identification & Classification" International Journal of Electronics Communication and Computer Engineering ,Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209.
[25] Aljamal, I., Tekeoğlu, A., Bekiroglu, K., Sengupta, S.: Hybrid Intrusion Detection System Using Machine Learning Techniques in Cloud Computing Environments. IEEE 17th International Conference on Software Engineering Research, Management and Applications
(SERA), 2019.
[24] Sethi, K., Kumar, R., Prajapati, N., Bera, P.: Deep Reinforcement Learning based Intrusion Detection System for Cloud Infrastructure. International Conference on Communication Systems & Networks (COMSNETS), 2020.
[25] Chkirbene, Z., Erbad, A., Hamila, R.: A Combined Decision for Secure Cloud Computing Based on Machine Learning and Past Information. IEEE Wireless Communications and Networking Conference (WCNC), 2019.
[26] Chkirbene, Z., Erbad, A., Hamila, R., Mohamed, A., Guizani, M., Hamdi, M.: TIDCS: A Dynamic Intrusion Detection and Classification System Based Feature Selection. IEEE Access, vol. 8, pp. 95864-95877, 2020.

[27] Zhao, X., Jiang, R.: Distributed Machine Learning Oriented Data Integrity Verification Scheme in Cloud Computing Environment,IEEE Access, Vol. 8, pp. 26372-26384, 2020.
[28] Ram Kumar, Sarvesh Kumar, Kolte V. S.," A Model for Intrusion Detection Based on Undefined Distance", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1 Issue-5, November 2011