International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-11 Issue-6, August 2022

Proficient Machine Learning Techniques for a Secured Cloud Environment

Majjaru Chandrababu, Senthil Kumar K

Abstract: ***Many different checks, rules, processes, and technologies work together to keep cloud-based applications and infrastructure safe and secure against cyberattacks. Data security, customer privacy, regulatory enforcement, and device and user authentication regulations are all protected by these safety measures. Insecure Access Points, DDoS Attacks, Data Breach and Data Loss are the most pressing issues in cloud security. In the cloud computing context, researchers looked at several methods for detecting intrusions. Cloud security best practises such as host & middleware security, infrastructure and virtualization security, and application system & data security make up the bulk of these approaches, which are based on more traditional means of detecting abuse and anomalies. Machine Learning-based strategies for securing cloud infrastructure are the topic of this work, and ongoing research comprises research issues. There are a number of unresolved issues that will be addressed in the future.***

Keywords: Cloud Computing, Anomaly Detection, Machine Learning Approaches, Supervised Learning and Unsupervised-Learning.

I. INTRODUCTION

Data storage on the cloud has grown in popularity in recent years. These servers are very popular because of their flexibility and cost-effectiveness. High storage capacity is also employed in non-conventional areas, such as social networking and online gaming. [1] Cloud security market will reach $12.64 billion by 2024, because to the increasing usage of large data capacity cloud services and constrained cyber-attack advancements. Products and solutions that concentrate on enforcement, administration, and data security are possible. As predicted by Gartner, the corporation is expected to be responsible for 95% of cloud security breaches. According to industry estimates, 70% of all businesses use the cloud, which means cloud security threats should be a concern for every business. Cloud computing, like many other types of technology resources, has provided several benefits. As an example, it allowed for the storage of a large amount of data and various resources. As a result, service costs were reduced by dispersing resources across a large number of consumers. There are several advantages to using the cloud for IT management, such as the ability to access it from any location, the ability to customize it, and the ability to save money. In order to ensure the tools' dependability and efficiency, a secure platform is required [2]. But as more vital applications are moved to the cloud, privacy and software protection concerns are becoming increasingly prevalent. In recent years, security researchers have been concentrating their efforts mostly on this. This bundle includes protection for the network, software, and data storage. The NIST defines cloud computing as follows: (National Institute of Standards and Technology), "a concept for providing simple, resource pooling, ubiquitous, on-demand access which may be readily offered with diverse forms of service provider interaction."[3] Consumers are charged only for the resources they use in cloud computing, which is referred to as Pay as You Go (PAYG). Pay-as-you-go (PAYG) models allow customers to customize their own software, storage, computing resources, and development platforms. Due to the above-mentioned advantages, the scientific community has attempted to construct a state-of-the-art idea [4]. A specific cloud environment has been defined primarily with its scale, ownership, and accessibility in the cloud development model, which includes cloud computing. It is true that cloud computing is made possible by the sharing of resources amongst individual users and local servers. As the cloud's functionality and presence increase, so does the deployment model's impact. There are three types of cloud implementation models: hybrid, private, and public [5]. As the impact of networks on our everyday lives grows, so does the importance of cybersecurity research for the models described above. Antivirus software, Intrusion Detection Systems (IDSs), and firewalls are some of the most important components of cyber-security technology. These strategies protect networks from both internal and external attacks. As a security tool that monitors network-based software and hardware conditions, an IDS is of course a kind of detection strategy. There are a number of fully-fledged IDS products on the market. A lot of IDSs are still plagued by a high rate of false alert, which causes security analysts to be overworked and may lead to a serious case of neglect. As a result, several scientists have worked to improve the detection rates of IDSs while also reducing the number of false alerts. Unknown assaults are another problem with existing IDSs. Because of the frequent changes in network circumstances, new attack types and risks are always emerging. Because of this, it is able to discover previously unknown incursions. In order to address these problems, researchers have been focusing on the development of IDSs using machine learning methodologies. ML is a kind of artificial intelligence that can automatically identify important information from a large dataset [6].

Manuscript received on 19 July 2022 | Revised Manuscript received on 22 July 2022 | Manuscript Accepted on 15 August 2022 | Manuscript published on 30 August 2022.

***** Correspondence Author

Majjaru Chandrababu, School of Information and Technology, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: majjaru.chandrababu2017@vitstudent.ac.in

Dr. Senthil Kumar K*, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: ksenthilkumar@vit.ac.in

Machine learning algorithms have enough generalisations to detect fluctuations in intrusions and new incursions when given enough training data. IDS powered by machine learning may be able to identify threats to a satisfactory degree. To further simplify the design and construction process, machine learning-based IDS don't need a great deal of domain expertise. Machine learning-based IDSs have already been offered, and this study attempts to identify and recapture them, as well as highlight the most important principles on the use of machine learning in security challenges. This analysis relies on papers produced between 2015 and 2019 to show how far we've come in recent years [24][33][34]. Machine learning strategies would have been used by researchers in previous surveys [7, 8]. For the most part, these surveys are intended to be used by academics interested in machine learning to construct IDS-based algorithms. In contrast, this form of taxonomy approach focuses on specific implementation technologies rather than cybersecurity challenges. Thus, the surveys do not specifically address how machine learning might be used to tackle challenges in the IDS area. To deal with the issue, we suggest a new IDS taxonomy and include relevant papers using this taxonomy. In this work, we address the shortcomings of previous surveys and suggest a thorough examination of the IDS detection mechanism. ' In addition, provide a thorough discussion of threat models, attacks, and IDS techniques in the cloud environment. Here are some of our most significant contributions. Study the different types of cloud security issues and how they can be resolved.

. Describe the Cloud-Based IDS Classification System taxonomy in detail.

. A thorough examination of the most popular Machine Learning Algorithms in IDS.

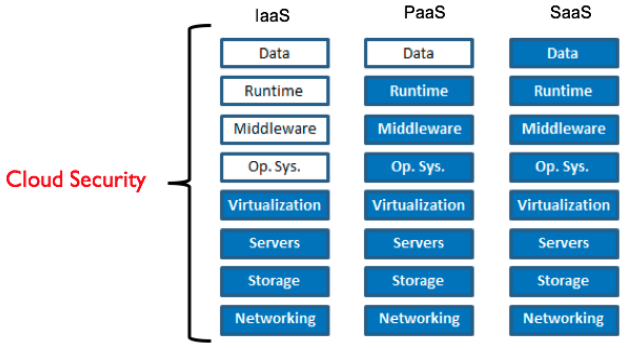
. Research on computer IDS training and challenges.

Cloud-based IDS and machine learning algorithms-based IDS are discussed in sections 3 and 4 of this paper.

This section covers the studies on machine learning-based IDS, Cloud Security challenges, and the final section sums up the papers in question.

II. CLOUD SECURITY ISSUES AND CATEGORIES

Cloud security protects all computer levels in both public and private cloud computing environments. In addition to the application layer, cloud applications are protected by SaaS, PaaS, IaaS, and cloud application security.



As part of this work, we're going to look at the current cloud security issues and the most advanced cloud security solution. 28 security risks (Table 2) are listed in the article and discriminated into 5 categories (Table 1). In addition, the most up-to-date defenses and reactions may be compared.

These five-cloud computing safety-related issues are summarized in Table 1 (see below). Small groups and just some of the four categories of cloud security vulnerabilities are classified using a same manner in [9,10].

Category 1: Regulators and organisations that define cloud security policy are included in the Security Standards category. It necessitates service level agreements, audits, and other agreements between consumers, service providers, and other stakeholders.

Category 2: Consumers use this information to choose which network they should use for cloud computing. It has web browsers, network connections, and data sharing capabilities that are enabled by registering.

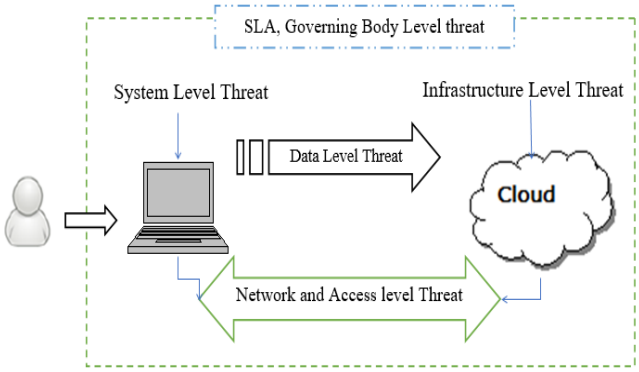
Category3: User-driven access control is concerned with defining, authenticating, and approving access to resources.

Category 4: SaaS, PaaS, as well as IaaS-related Cloud Infrastructure groups focus on security issues, and this is especially true in the virtual environment.

Category 5: Data integrity & privacy are part of the data group.

Table 1. Cloud Security Categories.

Category No	Category	Description
Category 1	Security Standards	Explain the requirements for cloud storage precautionary measures to deter intrusions. It regulates the safety regulations of cloud computing without negotiating reliability and efficiency.
Category 2	Network	It involves intrusions from the network like link access, DDoS, Dos (Denial of service), flooding attack, bugs in the IP, etc.
Category 3	Access Control	Covers connectivity and authentication. It detects problems that affect user privacy and storage of data.
Category 4	Cloud Infrastructure	Deals with cloud infrastructure-specific attacks Such skewed binaries and privileged insiders (IaaS, PaaS, and SaaS).
Category 5	Data	This includes data migration, privacy integrity, and data storage concerns relating to data security.



Components that may be vulnerable to attack are shown in Figure 2. Because cloud infrastructure, clients, and the network are all vulnerable to security risks, there must be measures in place to identify, prevent, and respond to these attacks. Table 2 categorises them according to the criteria outlined above for cloud security (Table 1). Interoperability key management methods, such as XML signatures and XML encryption syntax and processing, are necessary. Cloud computing's security requirements are currently nonexistent [11].

Even if safety standards are expressed correctly, compliance problems seem to be connected to a variety of safety difficulties [13]. It's because of a lack of oversight and a reexamination of company standards. Customers' knowledge of the cloud provider's standards, processes, and operations is limited, especially when it comes to issues like identity management and the division of responsibilities amongst employees. There really is no cloud service provider audit network, which is crucial for Cloud Computing's auditability [14]. Non-transparent services should not be outsourced, and clients should have the ability to audit the whole process. SLAs and regulatory concerns are covered by security standards (C1) or regulating organisations that are not included in cloud computing procedures. Because cloud computing relies on the Internet, it is more susceptible to network-related assaults than conventional computer systems [16]. Cloud computing relies significantly on networking and is tightly integrated [17]. This work's cloud security issues are therefore more prominent than those in other security categories. Though numerous cloud service providers focus on fast and economical performance, QoS (Quality of Service) is a surprise concern [11-15]. In the current study, QoS is seen as a characteristic or action that has a direct or indirect impact on security. Since numerous services might share cloud setups, a little inaccuracy in one or more cloud components can have a significant effect [18]. It has been noted in several case studies that the redundancy [19], loss (and leakage), location (21), recovery (18), confidentiality (21), security (22), and accessibility (22) of data are all critical challenges that must be addressed in order for them to be useful.

III. CLOUD -BASED IDS EVALUATION

The standard IDS framework has been employed in the cloud infrastructure by a number of researchers. ENISA, the "European Union Agency for Network and Information Security," has spent a lot of time focusing on cloud security issues. Stakeholders may use this data to better understand, assess, and manage cloud migration risk. In addition, it provides advise on SLAs for optimising security benefits and tracking. Furthermore, ENISA provides collaborative research to various stakeholders in order to identify critical cloud resources and assess how these conditions affect cloud service failure.

A. Intrusion Detection Systems (IDS)

Attempts to gain unlawful or unauthorised access to the computer system data or disrupt system activities are considered intrusions in the context of IDS.

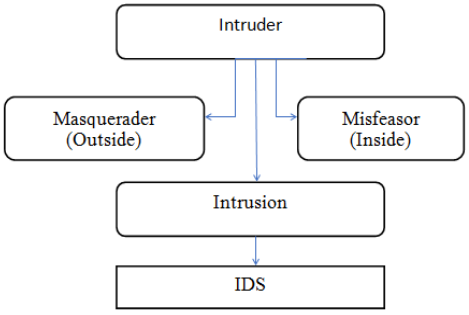
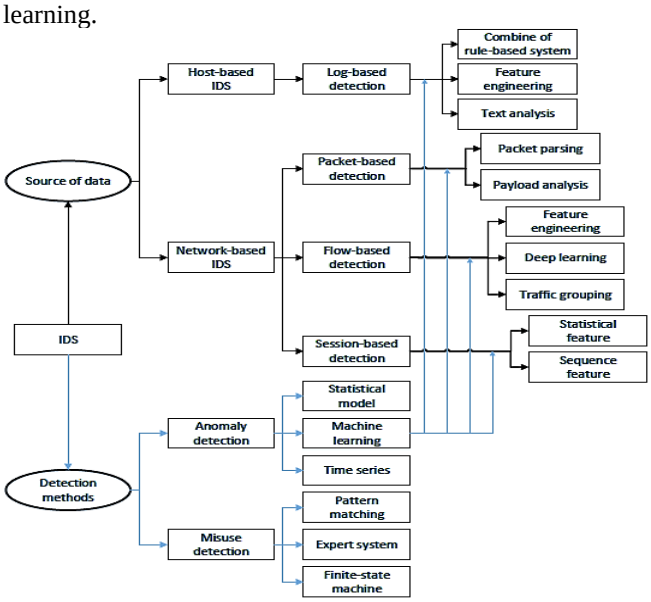


Figure 3. System Structure of IDS

Internal and external security threats may be detected by using an IDS [20], which monitors a wide variety of security threa0ts, as illustrated in Figure 3. Host and network control, computer system activity, warning generation, and responding to suspicious behaviour are some of IDS' most important responsibilities [21]. Due to their proximity to the network's security nodes (such as switches on big network segments), IDDS monitor the related hosts and networks [22]. There are two primary methods for classifying IDS data: One is based on detection, while the other is based on the source of the data. Detection-based IDS approaches include signature-based IDS, misappropriation IDS, and anomaly-driven IDS detection [23]. The IDSs use host-based and network-based data source methodologies. This research uses a primary data source categorization and a secondary detection system classification. Figure 4 shows a hypothetical IDS classification structure. Machine learning techniques to detecting methods are the focus of the study. Section 4 should also provide information on the different data types that may be utilised in IDS to implement machine learning.



B. Classification of Detection Approaches

Signature-based detection is a kind of misuse detection. Conceptual framework for characterising attacks. Sample signatures are used to search the signature database during the detection procedure. System developers have a major challenge in creating abuse detection

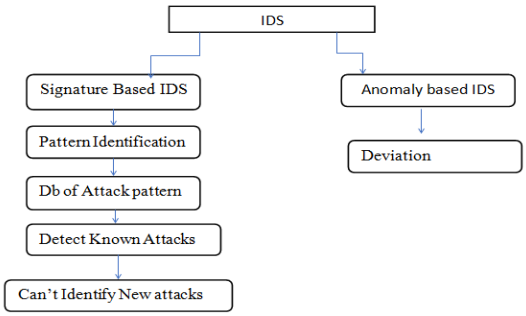


Figure 5. IDS Detection Classifications

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 3. Anomaly detection vs. misuse detection

	Misuse Detection	Anomaly Detection
Efficiency in detection	High, reduction with a scale of signature database	Dependent on the sophistication of the model
Dependence on domain knowledge	Nearly every detection depends on domain knowledge	Low, just the design feature Domain knowledge dependence
Performance in detection	low false rate of alarm; the high missed rate of alarm	low missed rate of alarm; high false rate of alarm
Detection of unknown attack	Known attacks are only detected	Known and unknown attacks are detected
Interpretation	Design based on domain knowledge, high interpretation	Detection results only for outputs, the poor potential for interpretation

Figure 4 depicts the various methods for detecting misuse, including pattern matching, expert system, and machinery-based finite-state techniques. Finding anomalies requires a combination of statistical modelling, time series analysis, and machine learning.

C. Classification by Data Source

The behavior of critical objects may be tracked by a host-based intrusion detection system (IDS) to quickly identify and react to intrusions (for example, ports, as well as programmes sensitive files) [26]. Network assaults cannot be detected by host-based IDS because they use resources on the host computer [27]. Larger switches or hosts are more likely to have an NIDS installed. For the most part, network intrusion detection systems may be deployed on a variety of operating systems [28]. An further benefit of network IDSs is their ability to identify attacks on a specific protocol or network type. Because traffic can only be traced via one network area, it has a drawback [29]. Table 4 shows how network-based IDS and host-based IDS are distinguished [30].

Table 4. Network based IDSs vs Host-based IDSs

	Host-based	Network-basedIDSs
Efficiency of detection	Low, several logs have to be processed	High, attacks can be detected in real-time
Source of data	Operating system logs or application programs	Network traffic
Traceability of intrusions	Follow the intrusion process call paths based on device	Track intrusion location and period based on IP address and timestamps
Deployment	Each host; operating system dependent; hard to deploy	Main network nodes; Deploy quickly
Drawbacks	Can't evaluate network behavior	Track traffic only in a particular segment of the network

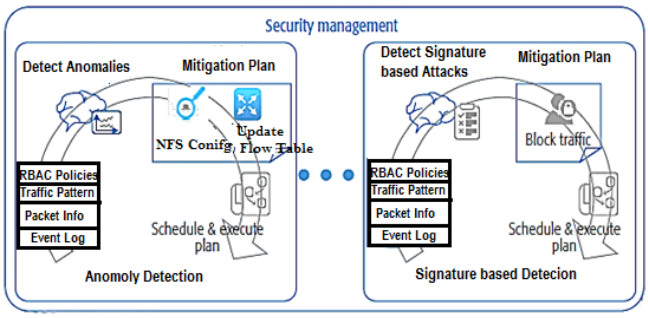


Figure 6: Security management Architecture

Security Management: The most common method of securing a network is to monitor it for signs of previously known dangers. Zero-day attacks may take advantage of this network vulnerability. This is critical, given the frequency with which new assaults are launched. As demonstrated in figure 6, ML's attitude on strict protection measures has been thoroughly examined [37][35]. Machine learning (ML) is used in an effort to identify violence, understand complex patterns of an assault from past data, and establish general rules that allow for the identification of changes in known attacks. Zero-day attacks may be detected using anomaly detection using machine learning (ML) [39]. As a result, it's important to know what's typical and what isn't.

IV. COMMON MACHINE LEARNING ALGORITHMS IN IDS

A. Machine Learning Models

Machine learning is classified as supervised or unsupervised learning. Labeled data is critical to the tracking of learning. While in-class categorization is a popular activity (and is most often employed in IDS), it is more expensive and time consuming to do it by hand [43]. Consequently, supervised learning is hampered most by a lack of appropriately labelled material. Unsupervised learning, on the other hand, uses unlabeled data to infer meaningful knowledge, making it considerably easier to gather training data [44]. However, the output of unsupervised approaches is often lower than that of supervised learning methods [45]. Figure 7 depicts some of the most common IDS machine learning techniques [36].

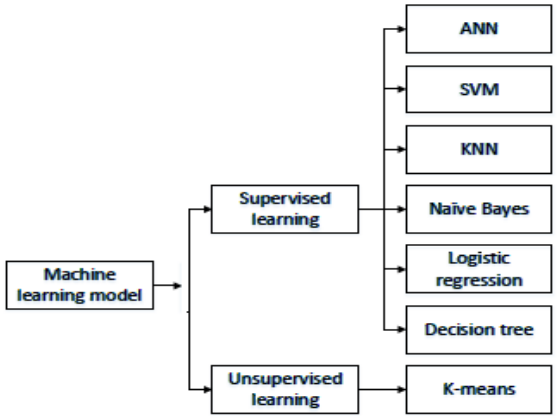


Figure 7. Classification of ML algorithms.

Table5. Advantages and Disadvantages of various ML classification algorithms

Algorithm	Advantages	Disadvantages	Improvement Measures
LR	Simple, quickly trained; Scale features automatically	Don't do well with non-linear data; Suitable foroverfitting	Regularization imported to prevent overfitting [28]
ANN	Capable of dealing with non-linear data;Good fitting capability	Suitable to overfitting; May be limited to the optimum locality; The training model consumes time	Enhanced optimizers, activation functions, and loss functions have been implemented
Decision tree	Choose automatically Functionality; Strong Interpreting	Trends resulting in assignment to the class of majority; Ignoring data correlation	Equalized SMOTE datasets; Latent variables added
KNN	Use for vast volumes of data; Perfect for nonlinear data; trained rapidly; Sturdy to Noise	Low precision on the Class of minority; Long testing periods; Parameter K sensitivity	Comparison time decreased by trigonometric inequality; Improved parameters by PSO (particle swarm optimization) methods [26];To equalized dataset uses SMOTE (synthetic minority oversampling technique) [27]
Naïve Bayes	Noise resistant; Might gradually learn	Don't do well with data relating to attributes	Imported latent stimulation variables Independent assumptions [28]
SVM	Get valuable information from a limited collection of trains; Good potential for generation	Don't do well with bigdata or different classification tasks; Kernel parameters sensitivity	Optimized parameters by PSO (particle swarm optimization) [25]
K-means	Simple, quickly trained; The scalability is strong; Can be adapted to bigdata	Don't do well with non-convex data; Initialization sensitivity; Parameter K sensitivity	Enhanced method of initialization [30]

B. Performance Evaluation Metrics (PEM)

Machine learning algorithms are tested using a variety of measures. The best models are chosen based on these criteria. In order to accurately determine the detection impact, many metrics are used simultaneously throughout the IDS search [40-42].

Accuracy: The percentage of samples correctly categorised and the total number of samples are given. An accurate measure for a balanced dataset is accuracy. However, in practical network setups, standard samples far outnumber aberrant samples, therefore precision will not be an acceptable measure.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

Detection accuracy is measured by the precision (P), which is the ratio of real positive samples to projected positive samples.

$$P = \frac{TP}{TP+FP}$$

The recall (R) statistic measures the percentage of true positive samples among all positive samples. This metric, also known as the detection rate, demonstrates the model's ability to detect attacks, which is a key statistic for intrusion detection systems.

$$R = \frac{TP}{TP+FN}$$

F-measure (F) is a way to say how well and how quickly you can remember the harmonics.

$$F = \frac{2*P*R}{P+R}$$

The false-negative rate (FNR) is the percentage of false-negative & total-positive samples that aren't true-negative samples. An attack detection rate that doesn't go off is called the FNR, or the rate that doesn't go off.

$$FNR = \frac{FN}{TP+FN}$$

The FPR (false positive rate) is the proportion of false-positive samplesandpredicted-positive samples. The false alarm rate (FPR) is a term used to describe how frequently an assault is mistakenly identified as a false positive.

$$FPR = \frac{FP}{TP+FP}$$

TP, FP, TN, FN, and FN are used to denote the true positive, false positive, true negative, and false, respectively. In order to differentiate intrusions, an IDS must be able to distinguish attack samples from non-intruding ones. A number of commonly used measures include accuracy, FNR (or missed alarm rates), recall (or detection rates), and FPR (or false alarm rates) [47].

C. Benchmark Datasets in IDS

Because machine learning output is dependent on input quality, the key aspect of machine learning is the ability to isolate meaningful information. Machine learning relies heavily on data understanding. In order for IDS data to be useful, it must be simple to retrieve and describe the behaviour of hosts or networks. IDSs often use logs, flows, sessions, and packets as data sources. It takes a long time and requires a lot of effort to create datasets. Once a benchmark dataset is developed, it may be used by many researchers. Benchmark datasets provide two more benefits in addition to their simplicity of usage [46].

(1) Benchmark data sets are reliable and provide more convincing experimental findings.

(2) Benchmark data sets are authoritative. Investigations employing conventional benchmark datasets,

such as those from earlier publications, were used in many published studies.

(1) DARPA1998

The MIT Lincoln Laboratory created the DARPA1998 dataset [31], which is widely used as a standard in IDS research. This set also includes labels and raw packages. There are five types of labels: User to root (U2R), denial of service (DoS), and probe. Traditional machine learning models cannot directly utilise raw packets, thus aKDD99 datasets were built to get around this constraint.

(2) KDD99

IDS's most popular dataset at the moment is KDD99[32]. The DARPA1998 mark is quite similar to the KDD99 mark. Statistical features depending on host and time are included, as well as basic features and content features, KDD99 offers four other sorts of options: KDD99's dataset has been shown to include a number of flaws. It's too late to represent the current network situation in KDD data.

(3) NSL-KDD

Using the suggested NSL-KDD [38], the KDD99 dataset has been improved. The KDD99 was used to select the NSL-KDD recordings. The NSL-KDD avoids the problem of classification bias by balancing records from different classes. Due to an error in the NSL-KDD, it has also removed everything except the most vital data from the database. The NSL-KDD mitigates data distortion and data repetition to some degree. Samples of tiny groups remain unidentified and out-of-date since the NSL-KDD does not include recent data.

(4) ISCX 2012

For networks that use FTP, HTTP, SSH, POP3, SMTP, and IMAP, the actual traffic traces are investigated in this dataset (Shiravi et al., 2012). In the data set, you can see what people typically do. The data collection is based on real-world tagged network traffic that encompasses several attack scenarios.

(5) UNSW-NB15

To analyse network traffic and retrieve 409-dimensional features using the Bro algorithm, three virtual servers were created by scientists at the University of South Wales. Compared to the KDD99 dataset, the dataset has a bigger number of attacks and characteristics. IDS built using machine learning need fresh datasets to be developed, despite the fact that UNSW-NB15 currently has less impact than KDD99.

(6) CICIDS 2017

DDoS, Heartbleed, Brute Force SSH, Infiltration, a DoS attack on a website, and a botnet are just some of the new malware attack elements included in CICIDS2017 [25]. In order to identify the dataset's source and destination, the time stamp, protocols and IP addresses are used.

V. RESEARCH ON MACHINE LEARNING-BASED IDSS

As a data-driven method, machine learning focuses on understanding information at the beginning. As a result, we make extensive use of the IDS model's data source form as the primary rating thread in this part, which demonstrates a variety of machine learning approaches to diverse data kinds. The variety of attack activity, including host and network behaviour, is represented by the various data types. Using device logs and network traffic data, you may learn more about the habits of the host and the behaviour of the network at large. There are a wide variety of attacks, each with a distinct pattern. As a result, the selection of appropriate data sources is required in order to identify distinct assaults depending on their features. When it comes to a DDoS attack, the ability to send a huge number of responses in quick succession is critical. Session data must be removed from the IP addresses that are best suited for this purpose [46-50].

A. Packet-Based Attack Detection

Packets, the network's fundamental unit of communication, show every aspect of communication. Binary data packets are difficult to decipher unless they are first digested. A packet contains information about the application and the packet's header. Ports, IP addresses, and other protocol-specific fields are all included in the header. Application layer protocols must provide payload for the application's data part [51].

There are three advantages of using packets as IDS data sources, including the following:

A packet contains information that may be used to identify U2L and R2L attacks.

There are IP addresses and time stamps included in the packet to allow for reliable identification of attack sources. To avoid the need for a cache, packets are examined instantly, allowing for quick identification of suspicious activity. Certain assaults, such as DDOS, are difficult to detect because the whole communication state and the contextual elements of each packet do not reflect the individual packets. Payload analysis and packet parsing are examples of packet-based identification techniques [27].

B. Machine learning Centric Secure Cloud Management

Table 4. Sophisticated, Secure Cloud Management using Machine Learning

Cloud Management Area (CMA)	Cloud Management function (CMF)	ML Techniques
Security	Signature-based Detection	NN,DT,BN,SVM
	Anomaly Detection	(Collaborative) K-means, NN,k-NN, DNN,(Collaborative) Decision Tree, (Collaborative) Support Vector Machine, BN

IDS classes include signature-based and anomaly-based detection methods. Machine learning approaches such as k-means and (Collaborative) Decision Trees are recommended for signature detection and (Collaborative)Support Vector Machine anomaly detection respectively. Cloud management that is machine learning-centric and secure is shown in Table 4.

Table 5 summarises the classification methodologies for IDSs based on machine learning.

Datasets	Classification Methods	Data Source	Machine Learning Algorithms	Authors
Private dataset	Packet parsing	Packet	K-means, SVM	Mayhew et al. [34]
DARPA 2000	Packet parsing	Packet	Fuzzy C-means	Hu et al. [35]
ISCX 2012	Payload analysis	Packet	CNN	Min et al. [36]
ISCX 2012	Payload analysis	Packet	CNN, autoencoder, and LSTM	Zeng et al. [37]
CTU-UNB	Payload analysis	Packet	Autoencoder	Yu et al. [38]
Private dataset	Payload analysis	Packet	GAN	Rigak et al. [39]
KDD99	Statistic feature for flow	Flow	Naive Bayes, SVM, DT	Goeschel et al. [40]
KDD99	Statistic feature for flow	Flow	KNN	Kuttranont et al. [41]
KDD99	Statistic feature for flow	Flow	K-means	Peng et al. [42]
KDD99	Traffic grouping	Flow	SVM	Teng et al. [43]
KDD99 and NSL-KDD	Traffic grouping	Flow	DNN	Ma et al. [44]
CICIDS 2017	Statistic feature for session	Session	DT	Ahmim et al. [45]
Private dataset	Statistic feature for session	Session	K-means	Alseiari et al. [46]
ISCX 2012	Session feature sequence	Session	LSTM, CNN	Yuan et al. [47]
ISCX IDS	Session feature sequence	Session	LSTM	Radford et al. [48]
DARPA 1998 and ISCX 2012	Session feature sequence	Session	CNN	Wang et al. [49]
Private dataset	Rule-based	Log	KNN	Meng et al. [50]
Private dataset	Rule-based	Log	DNN	McElwee et al. [51]

VI. CLOUD SECURITY CHALLENGES

Currently, researchers are focusing on concerns related to cloud security. In light of the near future, the following issues remain unresolved:

Data Classification based on Security: A cloud storage data centre may store data from a variety of different users. Depending on the data's monetary worth, it is possible to gain varying degrees of security via the classification of that data. Access frequency, update frequency, and the number of people who can access the data should all be taken into consideration while creating a database. If the data is detected and tagged, the degree of security associated with this designated data piece may be increased. Depending on the data nature, the degree of security necessitates storage, privacy, integrity, and encryption.

Identity management system: Cloud computing users are identified and then utilised to provide various services. Identity management is essential for both cloud service providers and their customers.

The identity management approach has a number of flaws. A confidential identity and access management process requires that id-generation, storage, distribution, and lifecycle management be handled in a secure manner [52].

- ***Secure trust-based Solution for cloud computing Service:*** Cloud computing services are hindered by the lack of a stable environment and general security concerns. Concentrating and discussing the cloud computing platform is a sure and dependable answer to many of today's problems.
- ***Optimization of resource Utilization:*** Additionally, cloud computing security, virtualization, and efficiency must be explored.

VII. CONCLUSION

This paper discusses the most recent cloud security issues and potential remedies. Among the twenty-eight cloud security issues we've identified are incorrect firewall settings, malevolent insiders and hacked binary systems, multi-tenancy, and unreliable browsers. In terms of addressing a company's security concerns, there are several subcategories of issues to consider. There must be a complete layer of security in place for the data and the cloud infrastructure in order for comprehensive cloud protection. Several studies are being conducted to address cloud security vulnerabilities. However, a solid cloud infrastructure requires the resolution of a number of outstanding issues. At the beginning of cloud computing, connection, network, web services, data secrecy, and applications are some of the most common issues. It is becoming more common for machine learning models to serve as good research advice. In addition, we explore the IDS taxonomy, which makes data sources the focus of several machine learning techniques. On the basis of this taxonomy, we expand and manage IDSs that are applied to a variety of data sources such as logs, session and packets. As attacks evolve, it is vital for IDSs to identify the most appropriate data source.

REFERENCES

1. https://www.prnewswire.com/news-releases/the-global-cloud-security-market-to-reach-usd-1264-billion-by-2024-300558185.html (Accessed on 10thApril 2020)
2. Subramanian N, Jeyaraj A (2018) Recent security challenges in cloud computing. Compute ElectrEng 71:28–42 [CrossRef]
3. Mell P, Grance T (2018) SP 800-145, The NIST Definition of cloud computing | CSRC (online) Csrc.nist.gov. https://csrc.nist.gov/publications/detail/sp/800-145/fnal. Accessed 11 Dec 2018
4. Xu X (2012) From cloud computing to cloud manufacturing. Robot ComputIntegrManuf 28(1):75–86. [CrossRef]
5. Bhamare D, Samaka M, Erbad A, Jain R, Gupta L, Chan HA (2017) Optimal virtual network function placement in multi-cloud service function chaining architecture. ComputCommun 102:1–16 [CrossRef]
6. Michie, D.; Spiegelhalter, D.J.; Taylor, C.(1994) Machine Learning, Neurall and Statistical Classification; Ellis HorwoodSeries in Artificial Intelligence: New York, NY, USA, Volume 13.
7. Buczak, A.L.; Guven, E.(2015) A survey of data mining and machine learning methods for cyber security intrusiondetection. IEEE Commun. Surv. Tutor. 18, 1153–1176. [CrossRef]
8. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.;Wang, C.(2018) Machine learning and deeplearning methods for cybersecurity. IEEE Access, 6, 35365–35381. [CrossRef]
9. Agrawal, S.; Agrawal, J.(2015)Survey on anomaly detection using data mining techniques. Procedia Comput. Sci., 60, 708–713. [CrossRef]
10. Sengupta, S.; Kaulgud, V.; Sharma, V.S.(2011) Cloud computing securityTrends and researchdirections. In Proceedings of the IEEE World Congress on Services (SERVICES),Washington, DC, USA, 4–9; pp. 524–531. [CrossRef]
11. Tripathi, A.; Mishra, A(2011) Cloud computing security considerations. In Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xi'an, China, 14–16 , pp. 1–5. [CrossRef]
12. Morin, J.; Aubert, J.; Gateau, B. (2012) "Towards cloud computing SLA risk management: Issues and challenges". In Proceedings of the 45th Hawaii International Conference on System Science (HICSS), Maui, HI, USA, 4–7; pp. 5509–5514. [CrossRef]
13. Braun, V.; Clarke, V. (2006) Using thematic analysis in psychology. Qual. Res. Psychol. , 77–101. [CrossRef]
14. A Survey on Cloud Computing Security, Challenges and threats|Whitepapers|TechRepublic. Available online: http://www.techrepublic.com/whitepapers/a-survey-on-cloud-computingsecurity-challenges-and-threats/3483757 (accessed on 18 April 2020).
15. Thalmann, S.; Bachlechner, D.; Demetz, L.; Maier, R.(2012)"Challenges in cross-organizational security management". In Proceedings of the 45th Hawaii International Conference on System Science (HICSS), Maui, HI, USA, 4–7; pp. 5480–5489. [CrossRef]
16. Wang, J.-J.; Mu, S.(2011) Security issues and countermeasures in cloud computing. In Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services (GSIS), Nanjing, China, 15–18 ; pp. 843–846. [CrossRef]
17. Lv, H.; Hu, Y.(2011)"Analysis and research about cloud computing security protect policy". In Proceedings of the International Conference on Intelligence Science and Information Engineering (ISIE), Wuhan, China, 20–21; pp. 214–216. [CrossRef]
18. Jain, P.; Rane, D.; Patidar, S.(2011) A survey and analysis of cloud model-based security for computing secure cloud bursting and aggregation in renal environment. In Proceedings of theWorld Congress on Information and Communication Technologies (WICT), Mumbai, India, 11– 14; pp. 456–461. [CrossRef]
19. Behl, A.(2011) Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation. In Proceedings of the 2011 World Congress on Information andCommunication Technologies (WICT), Mumbai, India, 11–14; pp. 217–222. [CrossRef]
20. Mathisen, E.(2011) Security challenges and solutions in cloud computing. In Proceedings of the 5th IEEE International Conference on Digital Ecosystems and Technologies Conference (DEST), Daejeon, Korea; pp. 208–212. [CrossRef]
21. Mahmood, Z. (2011) Data location and security issues in cloud computing. In Proceedings of the International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), Tirana,Albania, 7–9; pp. 49–54. [CrossRef]
22. Denning, D.E(1987) An intrusion-detection model. IEEE Trans. Softw. Eng. 222–232. [CrossRef]
23. Heberlein, L.T.; Dias, G.V.; Levitt, K.N.; Mukherjee, B.; Wood, J.; Wolber, D.(1990) A network security monitor.In Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, Oakland,CA, USA, 7–9; pp. 296–304. [CrossRef]
24. Kuang, F.; Zhang, S.; Jin, Z.; Xu,W.(2015) A novel SVM by combining kernel principal component analysis andimproved chaotic particle swarm optimization for intrusion detection. Soft Comput., 19, 1187–1199. [CrossRef]
25. Syarif, A.R.; Gata, W.(2017) Intrusion detection system using hybrid binary PSO and K-nearest neighborhoodalgorithm. In Proceedings of the 2017 11th International Conference on Information &CommunicationTechnology and System (ICTS), Surabaya, Indonesia,; pp. 181–186. [CrossRef]
26. Pajouh, H.H.; Dastghaibyfard, G.; Hashemi, S.(2017) Two-tier network anomaly detection model: A machinelearning approach. J. Intell. Inf. Syst. 48, 61–74. [CrossRef]
27. Mahmood, H.A.(2018) Network Intrusion Detection System (NIDS) in Cloud Environment based on HiddenNaïve Bayes Multiclass Classifier. Al-Mustansiriyah J. Sci., 28, 134–142. [CrossRef]
28. Shah, R.; Qian, Y.; Kumar, D.; Ali, M.; Alvi, M.(2017) Network intrusion detection through discriminative featureselection by using sparse logistic regression. Future Internet, 9, 81. [CrossRef]
29. Peng, K.; Leung, V.C.; Huang, Q.(2018) Clustering approach based on mini batch kmeans for intrusion detectionsystem over big data. IEEE Access, 6, 11897–11906. [CrossRef]
30. DARPA1998 Dataset. 1998. Available online: http://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusiondetection-evaluation-dataset (accessed on 16 March 2020).
31. KDD99 Dataset. 1999. Available online: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html(accessed on 16 March 2020).

32. NSL-KDD99 Dataset. 2009. Available online: https://www.unb.ca/cic/datasets/nsl.html (accessed on 16 March 2020).
33. Mayhew, M.; Atighetchi, M.; Adler, A.; Greenstadt, R.(2015) Use of machine learning in big data analyticsfor insider threat detection. In Proceedings of the MILCOM 2015-2015 IEEE Military CommunicationsConference, Canberra, Australia; pp. 915–922. [CrossRef]
34. Hu, L.; Li, T.; Xie, N.; Hu, J. (2015) False positive elimination in intrusion detection based on clustering.In Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery(FSKD), Zhangjiajie, China; pp. 519–523.
35. Min, E.; Long, J.; Liu, Q.; Cui, J.; Chen, W.(2018), TR-IDS: Anomaly-based intrusion detection throughtext-convolutional neural network and random forest. Secur. Commun. Netw, 4943509. [CrossRef]
36. Zeng, Y.; Gu, H.; Wei, W.; Guo, Y. Deep (2019) Full Range: A Deep Learning Based Network Encrypted TrafficClassification and Intrusion Detection Framework. IEEE Access, 7, 45182–45190. [CrossRef]
37. Yu, Y.; Long, J.; Cai, Z.(2017) Network intrusion detection through stacking dilated convolutional autoencoders.Secur. Commun. Netw. **2017**, 2017, 4184196. [CrossRef]
38. Rigaki, M.; Garcia, S.(2018) Bringing a gan to a knife-fight: Adapting malware communication to avoid detection.In Proceedings of the 2018 IEEE Security and PrivacyWorkshops (SPW), San Francisco, CA, USA, pp. 70–75. [CrossRef]
39. Goeschel, K.(2016) Reducing false positives in intrusion detection systems using data-mining techniques utilizingsupport vector machines, decision trees, and naive Bayes for off-line analysis. In Proceedings of theSoutheastCon 2016, Norfolk, VA, USA,; pp. 1–6. [CrossRef]
40. Kuttranont, P.; Boonprakob, K.; Phaudphut, C.; Permpol, S.; Aimtongkhamand, P.; KoKaew, U.;Waikham, B.;So-In, C.(2017) Parallel KNN and Neighborhood Classification Implementations on GPU for Network IntrusionDetection. J. Telecommun. Electron. Comput. Eng. (JTEC), 9, 29–33.
41. Peng, K.; Leung, V.C.; Huang, Q(2018). Clustering approach based on mini batch kmeans for intrusion detectionsystem over big data. IEEE Access **2018**, 6, 11897–11906. [CrossRef]
42. Teng, S.; Wu, N.; Zhu, H.; Teng, L.; Zhang, W.(2017) SVM-DT-based adaptive and collaborative intrusion detection.IEEE/CAA J. Autom. Sin., 5, 108–118. [CrossRef]
43. Ma, T.; Wang, F.; Cheng, J.; Yu, Y.; Chen, X(2016) A hybrid spectral clustering and deep neural network ensemblealgorithm for intrusion detection in sensor networks. Sensors **2016**, 16, 1701. [CrossRef]
44. Ahmim, A.; Maglaras, L.; Ferrag, M.A.; Derdour, M.; Janicke, H.(2019) A novel hierarchical intrusion detectionsystem based on decision tree and rules-based models. In Proceedings of the 2019 15th InternationalConference on Distributed Computing in Sensor Systems (DCOSS), Santorini Island, Greece, pp. 228–233. [CrossRef]
45. Alseiari, F.A.A.; Aung, Z. (2015) Real-time anomaly-based distributed intrusion detection systems for advancedMetering Infrastructure utilizing stream data mining. In Proceedings of the 2015 International Conference onSmart Grid and Clean Energy Technologies (ICSGCE), Offenburg, Germany,pp. 148–153. [CrossRef]
46. Yuan, X.; Li, C.; Li, X.(2017)DeepDefense: identifying DDoS attack via deep learning. In Proceedings of the 2017IEEE International Conference on Smart Computing (SMARTCOMP), Hong Kong, China;pp. 1–8. [CrossRef]
47. Radford, B.J.; Apolonio, L.M.; Trias, A.J.; Simpson, J.A.(2018) Network traffic anomaly detection using recurrentneural networks. arXiv:1803.10769.
48. Wang, W.; Sheng, Y.; Wang, J.; Zeng, X.; Ye, X.; Huang, Y.; Zhu, M.(2017) HAST-IDS: Learning hierarchicalspatial-temporal features using deep neural networks to improve intrusion detection. IEEE Access ,6, 1792–1806. [CrossRef]
49. Meng, W.; Li, W.; Kwok, L.F(2015) .Design of intelligent KNN-based alarm filter using knowledge-based alertverification in intrusion detection. Secur. Commun. Netw. 8, 3883–3895. [CrossRef]
50. McElwee, S.; Heaton, J.; Fraley, J.; Cannady, J.(2017) Deep learning for prioritizing and responding to intrusiondetection alerts. In Proceedings of the MILCOM 2017—2017 IEEE Military Communications Conference(MILCOM), Baltimore, MD, USA, pp. 1–5. [CrossRef]
51. Shiravi A, Shiravi H, Tavallaee M, Ghorbani AA (2012) Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Computers& security 31(3):357–374 [CrossRef]
52. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani,(2018) Toward generating a new intrusion detection dataset and intrusion traffic characterization," in ICISSP, pp. 108–116.

AUTHORS PROFILE

Majjaru Chandrababu, is a research Scholar in the School of Information and Technology and Engineering, VIT Vellore. He is Graduated in Computer Science and Engineering and Post Graduated in Information Technology, Hyderabad. And his research area is in Cloud Security using Machine Learning Approach

Dr. Senthil Kumar K, holds PhD with experience in research and teaching. His area of interests are Cloud Computing and Machine Learning. Having published papers in many international conferences and referred journals of repute, He is currently working as Associate Professor, School of Computer Sciences and Engineering, VIT University, Vellore. And Mentors research scholars in various fields of IT.