CSE 543: Information Assurance and Security

Arizona State University, Spring 2022

FINAL PROJECT REPORT-GROUP 14

PROJECT TITLE:

Using Blockchain and Machine Learning to significantly improve the security of Cloud-based Systems.

GROUP MEMBERS:

Name	Email Id
Dheeraj Neela (Leader)	dneela@asu.edu
Sai Greeshma Gandham (Deputy Leader)	sgandha2@asu.edu
Disha Prakash Bhukte	dpbhukte@asu.edu
Meghana Amirineni	mamirin1@asu.edu
Aditi Shashank Joshi	ajoshi64@asu.edu
Sujana Duvvuri Venkateswarlu	sduvvur7@asu.edu
Akhil Kumar Gudipoodi	agudipoo@asu.edu

SUMMARY:

Organizations are actively looking towards the use of Cloud systems rather than on-premises systems due to factors such as affordability, scalability, etc. Due to this rise in demand, there has been increasing concern for data security, as the world is witnessing a rise in data breach incidents. Hence, by utilizing revolutionary technologies such as Blockchain and Machine Learning, security can be improved to prevent data breaches.

Table of Contents

1. Introduction……………………………………………………………………………………4

1.1. Motivation and background…………………………………………………………..………4

1.2. Goals and scope of the study…………………………………………………..………...5

1.2.1. Cloud Computing…………………………………………………….....5

1.2.2. Benefits of cloud computing……………………………………..........5

1.2.3. Types of cloud computing…………………………………………......7

1.2.4. Deployment models in cloud computing…………………...………………7

2. Overview……………………………………………………………………………….....9

2.1. Security in Cloud Computing and its importance…………………………………..........9

2.2. Machine Learning in Cloud Computing……………………………..................……9

2.3. Blockchain in Cloud Computing…………………………………………………….........10

2.4. Team members responsibilities……………………………………………………….….11

3. Detailed Results………………………………………………………………………………15

3.1. Network security in Cloud Computing……………………….……………………...15

3.2. Threats in Cloud Computing Network………………………………………………........15

3.3. History of security solutions for network threat detection……………….…………….…..16

3.4. Security solutions network threat detection using Machine Learning algorithms………....17

3.5. Security solutions network threat detection using Blockchain……………………………,20

3.6. Malware Detection to prevent data breach…………………………...……………........22

3.6.1. Different types of Malwares…………….…………………………........22

3.6.2. Most frequent types of Malwares ……………………………………………23

3.6.3. Approaches to detect malware…………………………….……………...24

3.6.4. Methodology………………………………………............................26

3.6.5. Data breach in Block chain………………………….………………........29

3.6.6. How can blockchain help in reducing Data Breach…………………...........31

3.7. SQL Injection………………………………………………………………………………32

3.7.1. Types of SQL Injections……………………………………………………32

3.7.2. Threats in Cloud……………………………………….……………………33

3.7.3. Enhancing SQL injections detection with Machine Learning ..……………33

3.7.4. Methodology……………………………………………….…………...……34

3.8. DDOS Attacks………………………………………………………...………………...37

3.8.1. DDOS attack process…………………...…………………….……………38

3.8.2. DDOS attack in Cloud Computing…………………..……………….……......38
3.8.3. Types of DDOS attacks…………………………………………….…….…….39
3.8.4. Volume/Network based attacks…………………………………….…………..39
3.8.5. Protocol based attacks…………………………………………………………..40
3.8.6. Application based attack……………………………………………….……..41
3.8.7. Fragmentation Attack……………………………..……….……………42
3.8.8. DDOS mitigation using Machine Learning and Blockchain technologies….42
3.9. Account Hijacking………………………………………………………………………………...44
3.9.1. What is Cloud account hijacking………………………………………………44
3.9.2. Cloud account hijacking risks…………………..……………………………45
3.9.3. Simple solutions for Cloud account hijacking protection……….……………46
3.9.4. More secure solutions for cloud account hijacking defense………………...47
3.9.5. Machine Learning techniques to prevent Cloud account Hijacking...……...48
3.9.6. Blockchain solutions to prevent Cloud account hijacking……………..……49
4. Conclusions and recommendations…………………………………………………50
4.1. Conclusions and recommendations for Malware Detection………………………………50
4.2. Conclusion and recommendations for SQL injection …………….……………………….…51
4.3. Conclusion and recommendations for DDoS attacks ………………………………..……51
4.4. Conclusion and recommendations for Account Hijacking……………………………….........52
4.5. Conclusion and recommendations for Network Security…………………………………………53
5. References……………………………………………………………………………...55

1. Introduction

Nowadays, most businesses are migrating to the cloud rather than on-premises or standalone servers because of the lower cost and higher availability, as it can be scaled based on traffic to the server, reducing cost and maintenance. However, because more businesses are migrating to the cloud and relying entirely on the cloud for product and service delivery, there are numerous security problems associated with the cloud. The report focuses on identifying, overcoming, and mitigating security issues using cutting-edge technologies such as machine learning and blockchain to propose feasible solutions.

1.1 Motivation and Background

Modern cloud computing systems primarily offer two types of services which are Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS). IaaS offers basic computational and storage resources over the internet on-demand whereas PaaS offers a complete platform which is a combination of infrastructure and software required to support organizations' applications. In today’s world, organizations are actively looking towards the use of Cloud systems rather than on-premises systems due to various factors such as less cost of setup and maintenance, better security, scalability etc.

With the increase in demand for Cloud-based systems, there has been a similarly increasing concern corresponding to the data security within these systems, as the world is witnessing a rise in data breach incidents over recent years (namely Yahoo, Twitter, LinkedIn, Facebook, etc.). These data breaches not only affect an organization through huge penalties imposed by the regulators but also millions of individuals whose personal and financial identities are out in the open for misuse.

Considering all these factors, a lot of research is being conducted on the improvement of security in Cloud-based systems. Here we provide a literature study on utilizing revolutionary technologies such as Blockchain and Machine Learning technologies that have enormous potential to improve security in the Cloud environment. Hence, by utilizing revolutionary technologies such as Blockchain and Machine Learning, security can be improved to prevent data breaches.

1.2 Goals and Scope of study

1. Identifying the threats to the data integrity in cloud computing.
2. Explore the technologies related to the integration of Blockchain and Machine Learning.
3. Determining feasible machine learning models for detection of the threats.
4. Determine the use of Blockchain to mitigate the threats.
5. Determining the integration of Blockchain and Machine Learning models to mitigate the threats.

6. Determining measures for preventing the threats to data security in cloud computing.

Ensure data integrity by improving security in the cloud services such as SaaS, PaaS, IaaS using machine learning and Blockchain approaches from cyber-attacks like Malware, Adware, etc. o The proposed model will monitor the activities of the cloud using Blockchain and will assist with the identification of suspicious patterns, and logins proactively with the help of Machine Learning approaches.

1.2.1 Cloud Computing

When a computer system is available on-demand, as and when required, the system is large enough to store terabytes to petabytes of data, the processing speed of the system is quite large, and there is almost less to no human intervention than such a system is called Cloud Computing. This term can be seen in data centers available across the internet which are used by millions of people daily.

Cloud storage stores and transfers data via the internet to cloud service providers. These services are accessible on-demand based on capacity and pricing requirements, and they allow for convenient and remote data assistance from any location at any time. These systems' locations are diverse, resulting in rugged hardware that can handle a huge quantity of storage and compute efficiently. Despite their apparent power, these systems include vulnerabilities that allow for information exploitation and information conveyance.

1.2.2 Benefits of cloud computing

High Speed: The capability to deploy instances or servers in fraction of seconds has changed entire software development cycle and deployment speed. Without depending on standalone servers or lengthy server deployment process, server administrators may rapidly deploy new contents and construct application architectures.

Automatic software Updates and Integration: Continuous Integration and Continuous Delivery are since new software program variations can be easily tested and deployed within the cloud environment, allowing for faster product innovation and the release of more functions to end-customers on a monthly, weekly, and in some cases, daily basis. Cloud environments also interact with commonplace DevOps equipment and logging structures, making it easier to monitor and detect issues in production**.**

Efficiency and Cost Reduction: You won't have to spend a lot of money on buying and maintaining equipment if you use cloud infrastructure. This significantly lowers CAPEX and Total Cost of Ownership

(TCO). To grow your business, you don't need to spend money on hardware, buildings, utilities, or building a big data center. You don't even need large IT teams to handle your cloud data center operations because you can rely on the expertise of your cloud provider's workforce. Downtime costs are also reduced by using the cloud. Because downtime is infrequent in cloud systems, you won't have to spend money or effort dealing with capacity issues caused by downtime.

Scalability: Different organizations have different IT requirements — a large corporation with 1000+ employees would not have the same IT requirements as a small business. Using the cloud is a great solution since it allows businesses to scale up and down their IT departments fast and efficiently, depending on their needs. Cloud-based complete solutions are ideal for businesses with changing or variable bandwidth requirements. If your business requires expansion, you may easily expand your cloud capability without having to spend money on physical equipment. This level of adaptability might provide businesses that use cloud computing with a competitive advantage. This scalability reduces the risks associated with in-house operational issues and maintenance.

Back-up and Restore Data: The fact that records can be saved in the cloud without any restrictions also helps with backup and recovery. Older software program versions may be retained for later stages, in case they are needed for restoration or rollback, as end-customers record changes over time and want to be tracked for policy or compliance reasons.

Disaster Recovery: Having previous versions of software program saved within the cloud, as well as manufacturing times walking across multiple cloud availability zones or areas, allows for faster disaster recovery: if your application is deployed across multiple locations and one location goes down for some reason, the visitors can automatically failover to the running areas with no interruptions to end-users. In other cases, if the software program release contains a high-quality worm, a short rollback can be performed to restore a previously released, more secure version and reduce harm.

Control: Any business has to be able to manipulate sensitive facts. You never know what can happen if a file falls into the wrong hands, even if it's just the hands of an uneducated employee. The cloud gives you complete view and control over your data. You may easily establish which clients are at what stage of admittance to which facts. This allows for modification, but it also streamlines work because a team of workers will easily recognize which files are assigned to them. It could even develop into a simple collaboration. Because one version of the file can be worked on by multiple users, there's no need to keep multiple copies of the same file in circulation.

1.2.3 Types of cloud computing

Cloud computing is a type of service that provides the consumer’s capabilities such as compute, storage, and network capabilities on demand-based on the user’s requirement.

There are three types of cloud computing, which are given below:

- IAAS – Infrastructure as a service
- PAAS – Platform as a service
- SAAS – Software as a service

IAAS-Infrastructure as a Service

It is a type of cloud computing service that provides the client the ability to control, access, and manage the operating system, the runtime, and libraries and applications. The hardware resources can be accessed but can’t be controlled and managed. One of the best examples of this kind of service is Amazon AWS Elastic cloud computing (EC2). Which allows the user to create compute instances with the desired storage and network capabilities.

PAAS – Platform as a Service

This type of cloud computing service doesn’t provide the access to the hardware components and the operating system to the user but allows the user to access, manage and control the runtime and libraries for the application and it also provides the ability to deploy and control the applications that run to perform the certain task. One of the best examples of this kind of service is AWS Lambda. Which performs the task when triggered with a condition?

SAAS – Software as a Service

The user of the SAAS Service is not given the privilege to access, modify or control any of the hardware, OS, and library components but allows to access and manage the application that runs on the operating system of the server. The best example for this kind of service includes Microsoft 365 components such as MS Word and Excel.

1.2.4 Deployments in cloud computing

There are three cloud deployment models as shown below.

- Public Cloud
- Private Cloud
- Hybrid Cloud

PRIVATE CLOUD

This is the type of cloud model, where the infrastructure such as hardware components, the other parts of the software stack are solely owned and managed by the private body such as an organization or the government. The main motive of using the private cloud is security with the cloud and cost-effectiveness over a long time. The owner of the cloud has full control over all the operations within the private cloud. The reliability is comparatively less compared to the public cloud as the total no of computing resources is limited and there exists an initial upfront cost. The hosting of the resources of the cloud is done on the premises of the cloud owner. For example, all the IT Companies such as IBM and Oracle have their private cloud, whose resources are used by the employees of the organization.

PUBLIC CLOUD

The public cloud is a network that allows the users to access the cloud resources abundantly and the reliability is very high due to the availability of the resources globally. This type of cloud service provides the users with a variety of cloud functionalities. There is no upfront cost to set up the public cloud and security is one of the main concerns. The wastage of the resources is limited as the users of the public cloud use them only upon their demand. Some of the key features of the public cloud are the scalability that allows the resources of the cloud to scale in and scale out upon the user's demand. The resources are abundant due to the key factor called resource pooling.

Hybrid Cloud

This type of cloud deployment model combines the functionalities of both the private and public cloud. It brings in the advantages of the private cloud such as cost-effectiveness and security and provides a variety of cloud resources and reliability due to the merging of the public cloud functionalities. Both the features of the private cloud and the public cloud are used parallelly based on the demand. It can generally run using the private available resources and can use the public resources in times of requirement such as traffic. Whenever the traffic is high, we can move it to the public cloud and vice versa when the load is low. For Example, Amazon uses both public and private networks to perform its daily responsibilities of the organization

2. Overview

2.1 Security in Cloud Computing and its importance

The cloud is subject to attacks, and a cloud-based record can be hacked [22]. Data security is a major worry for everyone, given the widespread use of cloud services for data storage. Consider the case of a financial institution that chooses to host its application on the cloud. If a cloud provider utilizes third-party tools to enhance the capabilities of the cloud services, the data of the financial firm is at danger if the third-party tools are in any way insecure. Because the cloud relies on servers, the highly secured servers are unprotected to threats like SQL injections and DDoS attacks, data breach etc. If data integrity is violated, the organization may suffer significant financial losses, as well as the loss of consumers and a tarnished reputation. As a result, ensuring the data integrity of data hosted on cloud platforms is critical. In the sections that follow, we'll go through a variety of threats and ways for increasing data integrity in cloud computing.

2.2 Machine learning in Cloud computing

Since few years, the companies increased the utilization of cloud computing and cloud storage services. The massive quantity of data involving sensitive and critical data stored in the cloud possesses a high risk, reminding the need for increased security in Cloud computing. In improving cloud security, Machine Learning algorithms come to aid. They minimize the human job by deriving the patterns from the provided data that helps them to perform their job accurately. Machine Learning can predict the occurrence of a threat, it can detect intrusions and block its access to cloud networks and data without any human intervention. ML learns the normal system behavior and if it detects any abnormal behavior, it flags it as a potential threat and makes it compulsory to pass through additional levels of security phases. Machine Learning algorithms are divided into four types:

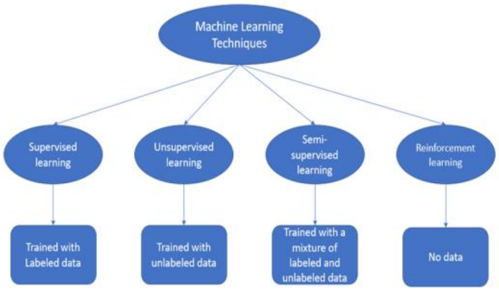

Machine Learning Techniques
Supervised learning
Unsupervised learning
Semi-supervised learning
Reinforcement learning
Trained with Labeled data
Trained with unlabeled data
Trained with a mixture of labeled and unlabeled data
No data

Figure 1: Machine Learning Techniques- types

1) Supervised Learning: Supervised learning techniques train on labeled datasets. This learning model is trained with labeled data with the aim to identify the data and make predictions accurately. This type is especially beneficial today as there are so many standard datasets available related to security threats from a variety of sectors. Examples of this type of Learning are Decision Tree, Naïve Bayes, Support Vector Machine etc.

Supervised algorithms can be further divided into two types.

a) Classification: Classification algorithms are used to train the machine to precisely categorize the data into specific classes.

b) Regression: Regression algorithms learn the relation between dependent and independent variables and are used for the prediction of continuous variables.

2) Unsupervised Learning: Unsupervised learning techniques train on unlabeled, unclassified, and uncategorized data to reveal hidden patterns. This type is beneficial where labeled data is not available, and the algorithms learns the hidden structure of the available data and make informed decisions. Examples of this type of Learning are K- means clustering, Hierarchical Clustering, etc.

Unsupervised algorithms can be further divided into two types:

a) Clustering: The technique for adjusting unlabeled data into groups based on similarity and difference.

b) Association: It uses a variety of rules to discover correlations between variables in a dataset.

3) Semi-Supervised Learning: This is a machine learning technique that blends a little amount of labeled data with a large amount of unlabeled data during training. It aims to observe how training a system with labeled and unlabeled input affects its learning behavior.

4) Reinforcement Learning: A reinforcement learning algorithm known as an Agent learns by interacting with its surroundings. Each correct action earns the agent a reward, while each incorrect action earns him a punishment. An agent's goal is to earn the highest reward points, so it enhances its performance.

2.3 Block Chain in Cloud Computing

Blockchain is a distributed ledger that consists of a sequence of blocks. Each block has 2 components i.e., block header and block body. Each block has only one parent except for the genesis block which is the first block of the blockchain [13]. The block header contains the block version, merkel tree root hash, timestamp, nBits, Nonce and parent block hash whereas the body contains the records.

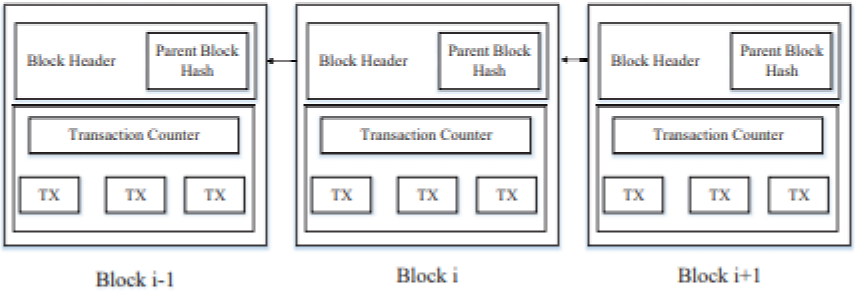


Figure 2. Blockchain Architecture [13]

As seen in the diagram above, the blocks are interlinked using the hash of the parent block. The blockchain network is a p-p system in which each node receives a copy of the blockchain. Without the validation of the blockchain network's nodes, a new block cannot be added to the blockchain. Furthermore, the blockchain employs digital signatures for data verification, adding an additional degree of protection. If an attacker attempts to tamper with the blockchain, the altered chain will be rejected owing to disagreement. As a result, tampering with the blockchain is tough.

Consensus algorithms are used to preserve data consistency in blockchain distributed networks (such as proof of work and proof of stake). A block can't be removed or reversed after it's been put to the blockchain. Users can interact with blockchain using a generated access. Note that blockchain cannot guarantee the perfect privacy preservation due to the intrinsic constraint [13].

As mentioned previously, the cloud is subject to attacks, and a cloud-based record can be hacked [15]. The use of blockchain technology in conjunction with cloud computing improves the security of cloud computing systems. It adds an extra degree of security by encrypting user data and data stored in databases. Blockchain's decentralized structure promotes transparency by recording each operation. Furthermore, because data cannot be altered, it reduces the risk of data loss or tampering. This facilitates audits by making it easier to trace down records

2.4 Team members responsibilities:

Member Name	**Responsibilities**

Dheeraj Neela	● Conducting weekly group meetings and individual meetings to track the status of the report and to identify the challenges. ● Setting up the Weekly goals to achieve safe delivery of the Project. ● Distributing the tasks and setting the deadlines for the timely final delivery of the report. ● Literature review of journal articles and conference papers related to Cloud Computing and its deployments. ● Research and analyze various **DDoS attacks.** ● Research and analyze various types of **DDoS attacks.** ● Review and evaluate existing solutions for detecting DDoS attacks utilizing Machine Learning and blockchain chain integrated frameworks. ● Conclude feasible solutions that detect and mitigate DDoS attacks using machine learning and blockchain. ● Review everyone's tasks and provide the necessary changes in the document.
Greeshma	● Literature review for existing Cloud Computing security issues ● Compiling cloud computing applications and security challenges and solutions ● Worked on various SQL attacks and its types. ● Coordinating with the group to organize meetings whenever required ● Handle the responsibilities of the leader in her absence, if required

Akhil	● Exploring various threats to cloud computing which could lead to data breaches. ● Investigate existing solutions proposed using Machine Learning to prevent data breaches, specifically Malware Attacks in Cloud Computing environments. ● Research on the efficacy of existing/proposed ML solutions to prevent Malware Attacks in Cloud environments. ● Recommend the best possible solution to proactively detect Malware attacks in Cloud environments.
Sujana	● Research and analyze various security solutions using Machine Learning and Blockchain Techniques to detect **Cloud Account Hijacking attacks.** ● Review and Analyze secure solutions to prevent cloud Account Hijacking by intruders. ● Summarize feasible solutions that detect and mitigate Cloud Account hijacking.
Aditi	● Compiling relevant information related to **Cloud computing and threats to Cloud Computing** ● Research and analytical methods to detect data breaches and analyze privacy concerns in Blockchain. ● Propose current or any possible solutions to prevent a breach of data in Blockchain
Meghana	● Literature review for existing Machine Learning techniques for intrusion detection in Cloud computing. ● Gather information on the use of Machine Learning to improve security in cloud computing.

	● Researched on network security and its importance in cloud computing.
Disha	● Researched about Network Security Threats in Cloud Computing. ● Gather information about the use of Blockchain in Cloud Computing. ● Researched Intrusion Detection systems in the cloud using Blockchain.
Everyone	● Actively participating in weekly progress reports and submission. ● Actively participated in contributing to the final project report.

3. Detailed Results

3.1 Network security in Cloud Computing

The qualities that make the cloud so powerful also make it difficult to safeguard. Adding additional assets to a cloud network is simple. The IT and security departments monitor all new infrastructure in an on-premises network. It means that all the new infrastructure is configured by the security professionals and the network expansion can be complex. This may not be the same scenario in adding infrastructure to the cloud network. In a cloud network, adding new infrastructure doesn't require any interference from the IT or security personnel and can be done by anyone or any machine with the required access and authorized credentials. It makes expanding the network much easier, but it also increases the likelihood that new infrastructure will not be set up securely, leaving it vulnerable to attack.

In a Hybrid environment setting where both on-premises and cloud networks environments are involved, the data is stored across different systems and protected by multiple security methods. It makes it difficult for security teams to keep a clear image of their cloud environment and obtain an accurate view of overall environment security. Due to the involvement of a hybrid environment, it becomes impossible to detect and track malicious actors moving across the cloud and on-premises networks.Another difficulty in securing cloud networks is the velocity of change in cloud settings. A weekly or even daily scan cannot detect a vulnerability that exists only for a few minutes which can provide enough time for a hostile actor to discover and exploit it.

Below are some of the principles that helps to improve network security in cloud computing environments:

1) To ensure isolation between several zones, the layers of firewalls must be activated.
2) All applications hosted in the environment should use end-to-end encryption.
3) While deploying a virtual private cloud like Secure Shell Protocol (SSH), Internet protocol security (IPsec), Secure Socket Layer (SSL) should be utilized.
4) Application of strong network controls to traffic flowing between different applications in the environment.

3.2 Threats in Cloud Computing Network

Data Breach: When any external person gets **access** to the sensitive information in cloud storage, it results in a breach of information. The damage depends on the sensitivity of the data. The data may include some financial information too which may result in huge damage. Data breach is caused due to multiple reasons such as theft. It is one of the major security concerns of the engineers who are working with cloud software’s. Cloud service providers are trying their best to ensure that data breach should not happen, and the data should be protected, however, that stage has not been achieved yet. This can be avoided by encrypting the data at the client and server side.

Data Loss: When data which is stored on cloud is not available as and when required, it results in loss of data. This can happen by accidental deletion of data while accessing cloud storage, overwriting the data from the client’s side, malicious attack on the server by hackers or untrustworthy employees, not keeping backups at data centers, server crash etc.

Insider Threats: This is a hazard that is extremely tough to counter. This threat includes the insider employee modifying, deleting, updating, or disclosing sensitive data. Cloud providers make certain that all of their personnel are trustworthy. [1] One safeguard that cloud customers should employ is to keep their encryption keys on their premises rather than on the cloud. "The system is still vulnerable to a hostile insider attack if the keys are not kept with the customer and are only exposed when data is used." A hostile insider poses a significant threat to systems that rely primarily on the cloud service provider for security. [3]

Data Location: The cloud's data should not be kept in a single location. Data should be stored in many locations with adequate backups. People should not be aware of the location of data; only higher authorities should be aware of it. [1] Although cloud providers should take responsibility for ensuring the security of systems (including data) and providing robust authentication to protect customers' information, what is required in this situation is that providers not only inform consumers, but also provide the necessary information that the consumer may not be aware of. [2]

Account Hijacking: This threat involves the leakage of credentials in order to hijack cloud services and steal sensitive information, enter fake information, and engage in other transactions, all of which can result in legal concerns for both the user and the cloud service provider. Hijacker has access to all services and information that were previously only available to authorized users. [1]

3.3 History of security solutions for network threat

Machine getting to know algorithms were contributing to each detection and vulnerability analysis. Researchers have addressed problems and helped create devices getting to know answers for danger detection. In this section, we are able to talk the vintage fundamental improvements withinside the discipline of danger detection the usage of device getting to know that helped the brand-new researchers to reach at what they may be now. In general, that research use most important kinds of device getting to know algorithms: clustering and type algorithms, each of that are powerful for danger detection. First, clustering algorithms are a procedure that calls for unsupervised device getting to know. It involves locating herbal grouping in facts mechanically without counting on education facts or type models. The k-means, DBSCAN, expectation-maximization, and car elegance algorithms are a number of the clustering algorithms studied in early community visitor's research. Second, in assessment to clustering, type algorithms offer supervised device getting to know for community danger detection.

3.4 Security solutions network threat detection using Machine Learning algorithms

In this section, we will review the machine learning algorithms and techniques used to identify and prevent network threats in a cloud environment. The problem statement of Network Threat detection can be identified as both a classification problem and an Outlier detection problem in Machine learning paradigms.

A hybrid machine learning approach for malicious behavior detection and recognition in cloud computing" [10], provides a classification approach for network threat detection in cloud systems. In this approach, the researchers used a novel technique of incorporating user behavioral features like location, number of times the user logged in, number of attempts for successful login, and many other such features rather than using the traditional features like credential checks and firewalls. Since this model is based on behavioral observations from multiple users rather than from a single user, this approach is considered reliable. A few high-level behavior examples are explained in Table 1 below.

Behavioral type	Evidences Collected
File Operations	No. of operations such as Read, Write, Delete, Open, Move etc.
Folder operations	Number of folder operations including create, read, write, open and delete.
Network operations	Number of different source/destination IPs, ports, query hosts/dead hosts/domain/DNS/DNS servers by different protocols.

Table 1: User Behavior Feature Examples [10]

This approach is further divided into pre-processing and network threat recognition phases. The data preprocessing phase is regarded as the important step in any machine learning as it crafts the data and cleans it for identifying threats in our model. Firstly, network traffic data generated by the user are collected using software like Argus and Netmate, which are cleaned by only selecting required features. These selected features are transformed into quantitative data using techniques like Onehot encoding, Label encoding, etc since machine learning algorithms can only work on numbers. The transformed features are reduced to improve the accuracy of predictions by using dimensionality reduction techniques like Principal Component Analysis (PCA) and also feature normalization is done using the min-max method. The diagrammatic representation of how network data is collected is shown below.

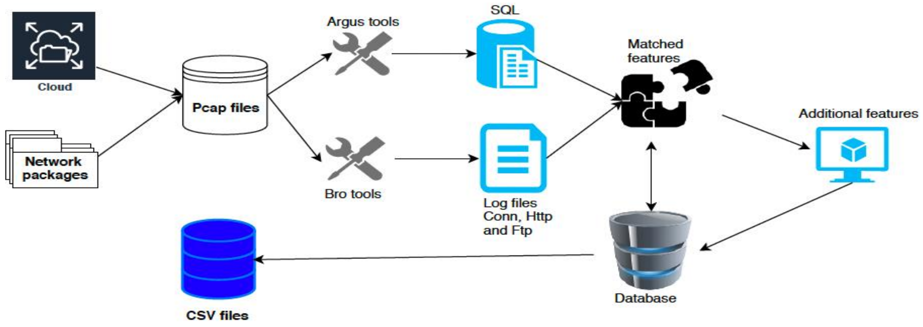


Figure 3: The flow process of feature generation from Pcap files to CSV files [10]

This research used a Probabilistic Neural Network model to detect whether a given network activity of a user is malicious or not. In the training phase of the ML model, the features are computed from the UNSW-NB15 dataset, and class category (whether malicious or not) is fed into the model and the neural network is trained.

In the testing phase, the incoming network data is collected, converted, reduced, and normalized and then fed into machine learning models for predictions, and required activity is taken. The conceptual view of the above-proposed system is depicted below.

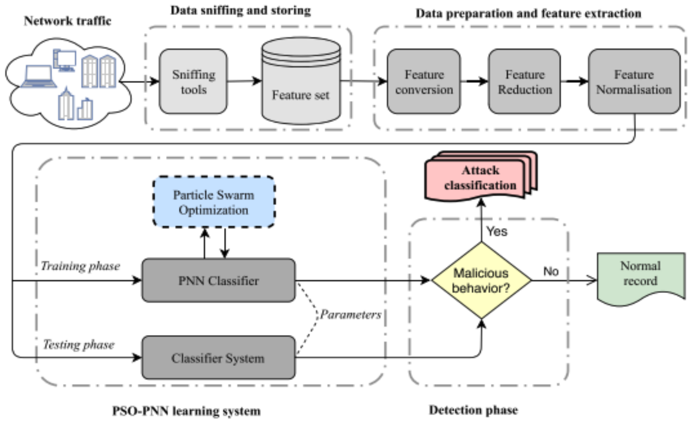
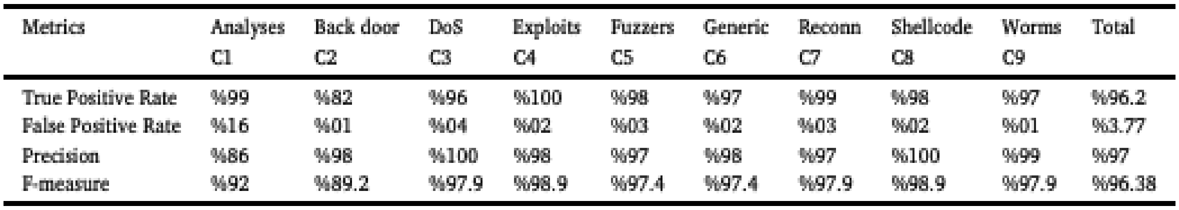


Figure 4: Conceptual View of Network Threat Detection [9]

The results from the above proposed system on the UNSW-NB15 dataset is shown in Table 2.



Metrics	Analyses	Back door	DoS	Exploits	Fuzzers	Generic	Reconn	Shellcode	Worms	Total
	C1	C2	C3	C4	C5	C6	C7	C8	C9	
True Positive Rate	%99	%82	%96	%100	%98	%97	%99	%98	%97	%96.2
False Positive Rate	%16	%01	%04	%02	%03	%02	%03	%02	%01	%3.77
Precision	%86	%98	%100	%98	%97	%98	%97	%100	%99	%97
F-measure	%92	%89.2	%97.9	%98.9	%97.4	%97.4	%97.9	%98.9	%97.9	%96.38

Table 2: Evaluation metrics of PNN [9]

Classification methods are most suitable for classifying existing types of attacks. But with the advancements in technology, hackers are coming up with new ways to get into the system, which is sometimes unidentifiable by the above methods. Outlier detection can be helpful in such cases, where any action which is not near to the normal actions is flagged and reported. The approach in Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach [12], discusses the technique to detect intrusions in the system using outlier detection approach.

The below figure illustrates the proposed system architecture of an NOF outlier detection approach. In this, intrusion packets that are received from the internet will be sent for feature extraction. After the features are extracted, they are sent to Intrusion Detection System (IDS). Large datasets are present in the training model to increase the IDS system performance. The distance between the extracted features and the trained model is then computed using the suggested IDS. As a result, false alarm will be if any of the outlier value exceeds the set threshold.

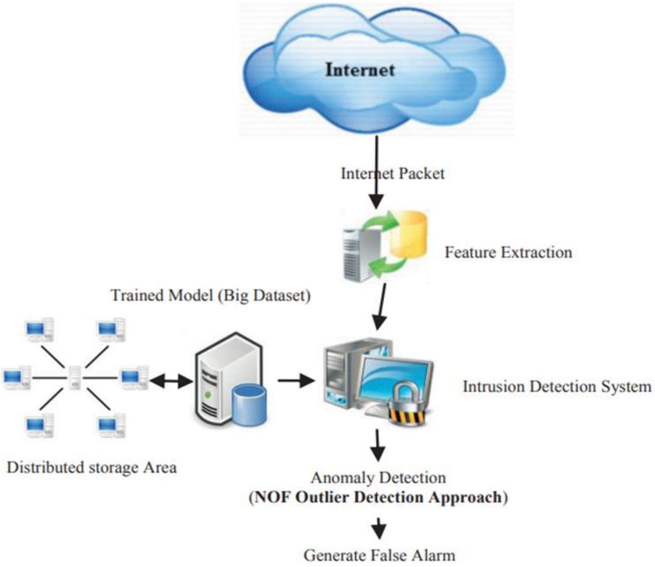


Figure 5: Proposed system Architecture [11]

Outliers are data points that are present at an unusual distance from the other data points. The aim of this model is to use the Neighborhood Outlier Factor (NOF) to designate a data example as an outlier and look for rare data with unique behavior compared to normal data in higher volumes.

Let's assume we have C1 with more examples in its cluster and C2 has less examples in its cluster compared to C1. In Figure 4, we can observe that C2 has higher cluster density when compared to C1. Inside the C1 cluster, Let's assume there is one object q for each example. P2 is considered as an outlier since, the distance between p2 and the nearest neighbor from C2 is less than the distance between q and its nearest neighbor. Whereas, by taking the nearest neighbor distances, P1 will be considered as an outlier. [11]

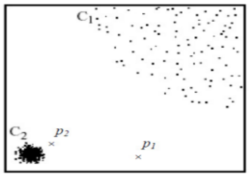


Figure 6- NOF Outlier Detection Approach [11]

Since NOF considers the density of all the points, it can find both p1 and p2 as outliers. In addition to this, irrespective of the size of dataset, NOF approach utilizes very less CPU compared to various other machine learning approaches. It also identifies almost all types of attack. Depending on the testing data's outlier value, the intrusion detection rate depends. The dataset is considered as an intrusion dataset when the outlier value increases. The below figure shows the comparison between the NOF outlier detection approach and the existing approaches in Anomaly Detection. In the below figure, violet colored line represents the NOF Outlier Detection approach.

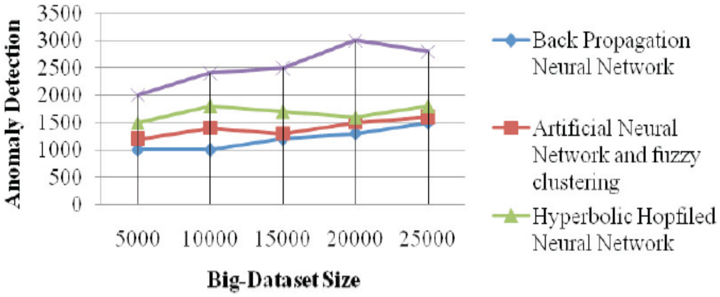


Figure 7: Big Data Size vs Anomaly Detection [11]

3.5 Security solutions network threat detection using Blockchain algorithms

We'll look at how blockchain can be utilized in cloud computing to detect network threats in this part. Intrusion detection systems are one of the most prevalent security techniques for detecting network threats [42]. Network Intrusion Detection Systems (NIDS) and Host Intrusion Detection Systems (HIDS) are the two types of intrusion detection systems (HIDS). The network intrusion detection system (NIDS) is primarily concerned with capturing network packets and analyzing them for hostile activity detection. It may be installed in network backbones, servers, switches, and gateways [42]. In contrast, the Host Incursion Detection System (HIDS) focuses on the system, evaluating system files, illegal access, and process patterns to identify intrusion [42]. The network intrusion detection system (NIDS) will be the center of our attention.

When large networks come into the picture, distributed intrusion systems will be required. In distributed IDS, Network Intrusion Detection System (NIDS) are deployed over a large network. All of the IDS in the distributed configuration exchange logs and alert information [42]. In this case, the network administrator must periodically setup the network in order to facilitate threat analysis and network monitoring. It allows managers to gain a more comprehensive picture of the network assault [43]. As the patterns of the assaults change, this network needs to be reconfigured from time to time. The attackers are always devising new techniques to dodge detection. This system has provides a centralized platform for threat analysis. Following are the developments necessary in this system:-

- The logs created by this system should be tamper-proof.
- There should be a common consensus among the nodes on the quality of alerts generated by the IDS.
- To meet the demands of a centralized node, the system must be scalable.
- Individual IDS should have rights and control over how alert data is disclosed and accessed.

Manish Kumar and Ashish Kumar Singh presented a Blockchain-Based Intrusion Detection in their paper[25] to address the above-mentioned criteria of DIDS.

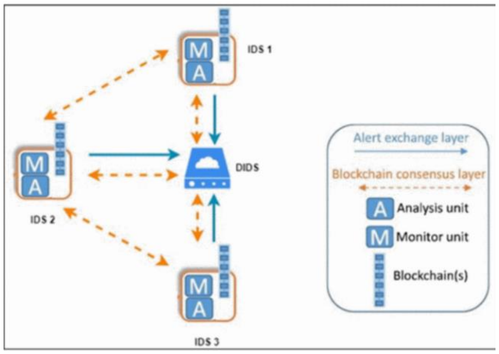


Figure 8. Blockchain based DIDS architecture [41]

The authors propose that IDS alerts could be securely shared over the network using blockchain. The IDS will store each alert it generates as a transaction. Before adding the new alert to the blockchain, the centralized server will start the consensus procedure which will update the blockchain only after validation of the new alert . The issue of scalability and performance will be handled by the cloud infrastructure whose security needs to be ensured using this system. Furthermore, because massive amounts of data will be created, the cloud system will be useful for managing real-time data. The paper uses Amazon EC2 instances as a centralized node and distributed nodes where the IDS is installed as an example.

3.6 Malware Detection to prevent data breach

[44] Cloud computing is a rapidly growing technology that allows users to access scalable on-demand resources and services at a lower cost of infrastructure. [46] The number of services and features offered by various cloud service providers (CSP) has recently increased dramatically. Infrastructure as a Service is the practice of renting out access to servers to clients for computing and storage reasons (IaaS). The rise in popularity of IaaS has sparked major and pressing concerns about cyber security and privacy. Malicious entities frequently use malware to compromise sensitive data or hinder cloud services' functionality. The vulnerability of Infrastructure as a Service (IaaS) clouds to malware is a big issue.

Cloud automation solutions have become the standard, allowing IT professionals to supply resources in the cloud automatically. Such automation is accomplished using tools (such as Puppet1 and Chef2) that allow you to write configuration scripts that can build, change, and destroy cloud resources. Similar to how orchestration technologies aid DevOps teams, they can increase the security threat surface.

VMs are frequently produced using automatic setup methods, resulting in a large collection of VMs that are similarly configured, if not exact replicas. Malware might quickly spread between VMs due to the inherent redundancy in these virtual machines, especially if the configuration scripts contain weaknesses. Compromise of a group of VMs has significantly more serious consequences than a single compromised VM. Due to its intrinsic complexity and dynamic environment, where threats are continually changing and expanding, cloud infrastructure necessitates significant security implementations. As a result, it is critical to design malware detection systems that are both accurate and quick.

3.6.1 Different types of Malwares:

Some malware types have evolved into hybrids as they employ similar malware attack methods, such as using logic bombs, which are pre-programmed attacks that are sometimes triggered by victims themselves; phishing and social engineering tactics to deliver malware directly to victims; or mobile malware, which targets mobile devices.

The most frequent types of malwares to be aware of include the following – Fileless Malware, Viruses, Trojan Malware, Adware, Ransomware, Worms, Bots, Spyware, and Rootkits. However, we will look into different types of malware attacks specific to cloud computing environments.

3.6.2 The most frequent types of malwares:

When it comes to cloud virus attacks, there are basically five different varieties to choose from. You may learn about each one individually here.

DDoS Attacks: In a cloud-based system, a distributed denial of service (DDoS) attack is a popular sort of attack. Large-scale botnets are used by cybercriminals to flood a network with malicious traffic, effectively shutting down or drastically slowing down the cloud computing system. Because of the millions of hacked devices, botnets have become more accessible to attackers. DDoS assaults are most common in public clouds, where they disrupt an entire neighborhood's infrastructure. Furthermore, if a DDoS is left unchecked or untreated for an extended period of time, attackers may be able to use cloud computing resources for criminal purposes by modifying cloud computing characteristics.

Hypercall Attacks: The hypercall assault is carried out in a way that is intrusive. The attacker poses as a guest who requests domain access from the host via the hypercall interface provided by the hypervisor. The attacker here compromises the hypercall handler-enabled VMs of a company. Once hypercall attacks are launched, they might make it difficult for hosting hypervisors to identify and prevent them using traditional network security techniques.

Hypervisor DoS: In a hypervisor attack, an attacker takes advantage of the hypervisor, which manages numerous virtual computers on a virtual host. If the hypervisor is compromised, malware can infect any of the VMs running on the host. As a result of an affected hypervisor, virtual machine resource usage may escalate, resulting in a denial of service to the entire host or potentially multiple hosts. Because most hosts are networked and do not require authentication from other hosts, they can simply infect more hosts, exacerbating the problem.

Hyperjacking: The attacker must have control of the hypervisor in order to infect a cloud computing-based system with a hypervisor DoS. To acquire control of the hypervisor, the attacker employs a rootkit planted on a VM (Virtual Machine). Hyperjacking is the term used to describe such cyber-attacks. An attacker who successfully hyperjacks the hypervisor's power can take control of the entire hosting environment. As a result, attackers can alter the functionality of virtual computers and do damage to them.

Exploiting Live Migration: Moving to the cloud or transferring between clouds provides a big opportunity for attackers. When a business does an automatic live migration, attackers can hack the cloud management system and change it in a variety of ways: Create a denial-of-service attack by making a large number of fraudulent migrations. Migrate resources to a virtual network or cloud subscription controlled by the attacker. Change the transferred systems to make them less vulnerable to future attacks.

3.6.3 Approaches to detect Malware

We begin with a brief explanation of malware detection strategies in this section. The challenge of malware detection can be approached in a variety of ways. We'll look into signature-based, behavior-based, and statistical-based detection in this section.

Signature Based Detection

[45] The most extensively utilized method of detection is signature-based. A signature is a string of bytes that can be used to identify a particular piece of malware. Signatures are scanned using a number of pattern matching algorithms. Systems that rely on signatures must keep a store of known malware signatures up to date when new threats emerge.

The signature of an executable is evaluated and compared to a database of known malware signatures in static malware detection [45], [46], [47]. Attackers have used strategies like obfuscation and packing to try to reduce the usefulness of the static analysis. Furthermore, static malware analysis is confined to known malware executables and is unable to detect zero-day malware, which is constantly changing. Because of these two significant drawbacks, there has been a lot of study into behavioral malware detection approaches.

Behavior Based Detection

[46] The acts that the virus takes during execution are the focus of behavior-based detection. Behavioral-based solutions include dynamic and online virus detection. Malware executables are launched in a protected environment, such as a sandbox, and their activity is analyzed using dynamic malware detection methods. Because it is not based on previously known signatures, but rather on the actual

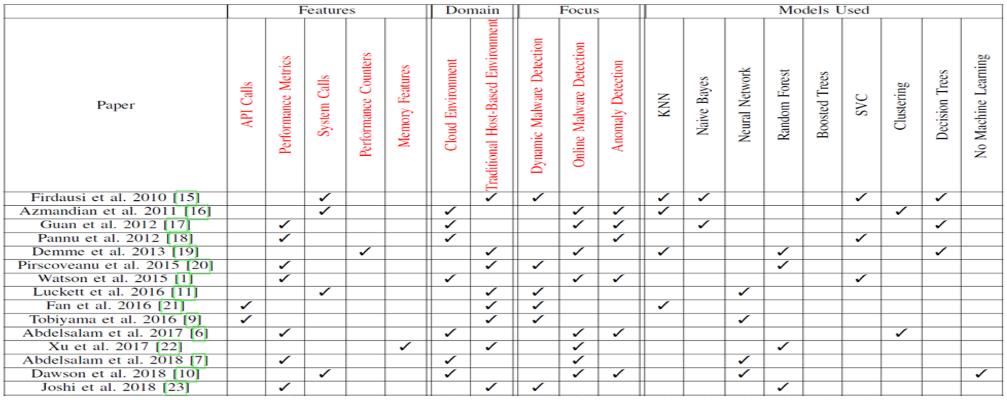
behavior of the executable, the detection system may evaluate novel zero-day malware. However, attackers have developed malware that can detect the presence of sandboxes and stop behaving aggressively in order to avoid discovery. The detection system focuses on identifying malware in the delivered executables before they are run on actual computers in both dynamic and static techniques. Malware is, nonetheless, very frequent.

Malware, on the other hand, frequently enters a system via weaknesses, evading these crude detection methods. The behavior of a system that it is trying to safeguard from malware is the topic of online malware detection [5]– [8]. Online approaches, rather than evaluating executables and their activity, monitor the performance of the complete virtual system and trigger an alarm if any signs of malicious behavior are discovered at any time. As a result, online malware detection methods are regarded as continuous real-time detection system that overcomes the drawbacks of static and dynamic malware detection methodologies.

Online Malware Detection

Online malware detection methods concentrate on continuously monitoring entire systems, assuming that malware would infiltrate the system at some point. Machine Learning (ML) and neural network approaches are frequently used to accurately and efficiently record malware behavior [9]. This is owing to the models' ability to evaluate a large amount of data provided by a virtual machine in order to identify executables as dangerous or benign. The parameters chosen to capture the behavior of the virus installed in a given system have a substantial impact on online malware detection procedures.

For example, some works [2], [10], [11] use system calls (the most extensively used); nevertheless, system calls are resource intensive, and data can only be retrieved using an on-host collecting agent. Because resource consumption measures (also known as performance metrics) are less expressive than system calls in terms of catching low-activity malevolent conduct, only a few works [1], [5]–[8], [12], [13]



	Features					Domain		Focus			Models Used								
Paper	API Calls	Performance Metrics	System Calls	Performance Counters	Memory Features	Cloud Environment	Traditional Host-Based Environment	Dynamic Malware Detection	Online Malware Detection	Anomaly Detection	KNN	Naive Bayes	Neural Network	Random Forest	Boosted Trees	SVC	Clustering	Decision Trees	No Machine Learning
Firdausi et al. 2010 [15]			✓				✓	✓			✓	✓				✓		✓	
Azmandian et al. 2011 [16]			✓			✓			✓	✓	✓						✓		
Guan et al. 2012 [17]		✓				✓			✓	✓		✓						✓	
Pannu et al. 2012 [18]		✓				✓				✓						✓			
Demme et al. 2013 [19]				✓			✓		✓		✓			✓				✓	
Pirscoveanu et al. 2015 [20]		✓					✓	✓						✓					
Watson et al. 2015 [1]		✓				✓			✓	✓						✓			
Luckett et al. 2016 [11]			✓				✓	✓					✓						
Fan et al. 2016 [21]	✓						✓	✓			✓								
Tobiyama et al. 2016 [9]	✓						✓	✓					✓						
Abdelsalam et al. 2017 [6]		✓				✓			✓	✓							✓		
Xu et al. 2017 [22]					✓		✓		✓					✓					
Abdelsalam et al. 2018 [7]		✓				✓			✓				✓						
Dawson et al. 2018 [10]			✓			✓			✓	✓			✓						✓
Joshi et al. 2018 [23]		✓					✓	✓						✓					

Figure-9: The differences between related works are discussed in this table.

use them. Performance metrics, on the other hand, are better suited to cloud systems since they are less expensive and can be easily retrieved from the hypervisor. The red sections indicate a difference in the features used, the environment in which the solution was tested. A Tick means this attribute or model is present in this paper, while a blank cell means this attribute or model is not present.

Using Machine Learning models that use process-level performance measures, we examine and compare the effectiveness of various online virus detection methods.

3.6.4 Methodology

Experimental Setup: Creating reliable malware detection models necessitates simulating real-world data. A cloud environment with traffic to simulate real-world cloud activity is required to do this. The testbed had a single control node and four compute nodes, and was equipped with tools such as OpenStack, a prominent cloud platform. The dashboard, storage, network, identification, and compute are all handled by the control node. The compute nodes are exclusively responsible for computational services, and each compute node is equipped with networking, polling, and collection agents. In the data collection process, it's also important to let malware act naturally.

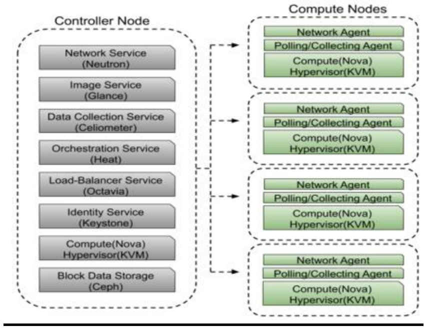

Controller Node
Network Service (Neutron)
Image Service (Glance)
Data Collection Service (Celiometer)
Orchestration Service (Heat)
Load-Balancer Service (Octavia)
Identity Service (Keystone)
Compute(Nova) Hypervisor(KVM)
Block Data Storage (Ceph)
Compute Nodes
Network Agent
Polling/Collecting Agent
Compute(Nova) Hypervisor(KVM)
Network Agent
Polling/Collecting Agent
Compute(Nova) Hypervisor(KVM)
Network Agent
Polling/Collecting Agent
Compute(Nova) Hypervisor(KVM)
Network Agent
Polling/Collecting Agent
Compute(Nova) Hypervisor(KVM)

Figure 10: Node Structure

Malware Samples: VirusTotal provided us with the malware that we used in our experiments. A total of 113 samples were collected and picked at random from a variety of malware families, including DoS, Backdoor, Trojan, and Virus, among others.

Experimental Deployment: To simulate a real-world environment, a three-tiered web architecture is used. This architecture included web servers, application servers, and a database server. An internal load balancer connects the web servers to the application servers, distributing requests among the application servers, and the application servers are all connected to a single database server. There is also an auto-scaling strategy that is based on average CPU utilization and applies to both the web and application servers individually. New VMs are created and attached to the corresponding load balancer if the average CPU utilization of all VMs in the web or application layer hits 70%.

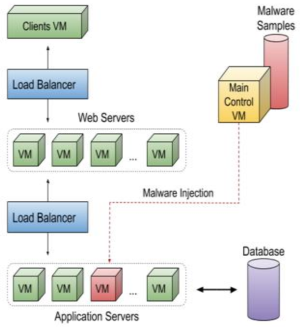
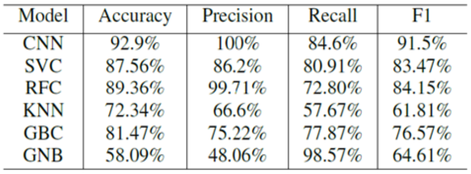


Figure 11: Server Architecture

Data Collection: A total of 40,680 samples were collected from 113 trials, each utilizing a distinct malware executable. As illustrated in Figure 3, each experiment lasted one hour and was divided into benign and infected periods. The first 30 minutes are the benign phase; no virus is injected into the PC during this time. A single piece of malware is introduced into one of the application servers between minutes 30 and 40. The malware injection and execution periods vary, giving the tests a more dynamic feel and ensuring that a strict injection and execution would bias the results. Minute 40 is known as the nefarious phase. Malware is openly running on the machine at this time. Every 10 seconds, data samples are collected, resulting in a total of 360 samples being recorded in the database for each experiment.

Evaluation

We utilize five standard metrics to assess the performance of different models: accuracy, precision, recall, and F1, where TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively (False Negative).



Model	Accuracy	Precision	Recall	F1
CNN	92.9%	100%	84.6%	91.5%
SVC	87.56%	86.2%	80.91%	83.47%
RFC	89.36%	99.71%	72.80%	84.15%
KNN	72.34%	66.6%	57.67%	61.81%
GBC	81.47%	75.22%	77.87%	76.57%
GNB	58.09%	48.06%	98.57%	64.61%

Table 3: Performance results for the different ML models

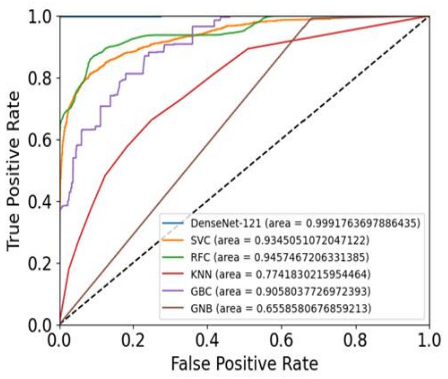
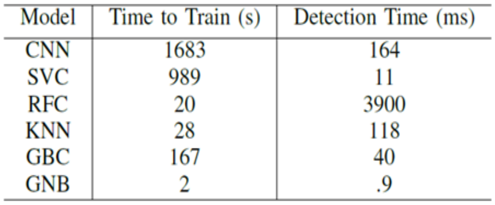


Figure 12: Receiver Operating Characteristic (ROC) Curves



Model	Time to Train (s)	Detection Time (ms)
CNN	1683	164
SVC	989	11
RFC	20	3900
KNN	28	118
GBC	167	40
GNB	2	.9

Table 4: Time Cost for the Models

_**Analysis:**

With an F1 score of 91.5 percent, the CNN model beat all other models in terms of overall performance. We use the F1 score to describe overall performance since it considers the precision and recall metrics and is thus better than the accuracy measure at capturing a model's overall performance. With 84.15 percent and 83.47 percent, respectively, the RFC and SVC models obtained the next highest F1 scores. The F1 scores for the GBC, GNB, and KNN models were 76.57 percent, 64.61 percent, and 61.81 percent, respectively.

The GNB model, with a score of 98.57 percent, was the model that scored the highest in the recall metric, which quantifies the percentage of contaminated samples that were detected. However, there is one caveat: its precision score of 48.06 percent, which counts how many of the samples identified as infected were truly infected, was extremely low. The CNN model had an 84.6 percent recall rate and a 100 percent precision rate.

When we look at the accuracy scores, we see that the CNN model outperforms the others, with the RFC and SVC models not far behind. When we take a look at the ROC Curves and AUC scores for each of the models, the best performing classifier is the CNN model with an AUC score greater than 99%.

The CNN model's superior metrics clearly show that it is the greatest fit for our use case. The CNN algorithm correctly identified 84.6 percent of all infected samples while not mislabeling healthy samples as infected. While it is preferable to detect every occurrence of malware, a large number of false positives can cause just as much disruption as infection. As a result, the GNB model's high recall rate is not as promising as it appears. The model's exceptionally low precision score of 48.06 percent suggests that it produced a large number of false positives. In reality, 52 percent of the samples that were identified as infectious turned out to be healthy. In a business case, a model that generates this many false positives could obstruct day-to-day operations by mislabeling important, non-malware processes as malware.

The amount of time it takes to train these models, as indicated in Table III, can influence which use cases each model is appropriate for. The more effective models have a clear pattern of taking longer to train. By far the most time-consuming to train, the CNN model also outperformed every other model by a significant margin. The RFC model outperforms the SVC model in terms of training time, as the SVC model took over 900 seconds to train while producing similar outcomes to the RFC model. In general, sacrificing some effort to train models that are as precise as feasible is usually worthwhile. As a result, using a deep learning method like the CNN model rather than rapidly taught models like SVC and RFC is favored.

3.6.5 Data Breach in Blockchain

Blockchain is a ‘decentralized’ technology - in a network before a new transaction is added the system requires all parties to give consent. As blockchain is a shared network, transactions one written in a block of information in the system cannot be altered. Every user can add information to the blockchain and every other member across the network is responsible for verifying the data being added is authentic, preventing removal of existing data. [4]

Blockchain is a self-auditing system in which the blockchain network automatically “updates” and keeps a check on its data every 10 minutes or so. Due to this there are many benefits which make it difficult for a hacker to breach the data.

In a centralized system, when a hacker occupies one version of the data, the rest of the system can be hard to hack. But with blockchain, even if the hacker gets a hold of one block of information in the network, they will have to get to the next block and the next and next to successfully take over the system.

Some of the Largest, most recent cyber hacks include the 2013/14 breach of Yahoo's database by what is thought to have been a state-sponsored cyberattack, impacting over 3 billion users. The hackers collected consumers' names, email addresses, telephone numbers, dates of birth, hashed passwords and unencrypted answers to security questions. [6]

The majority of the data gathered and stored is under the hands of governments and corporations, which have collected large amounts of personal data to protect. These organizations may be monetizing these datasets at the same time, either by using them to better their own operations and offers or by selling them to third parties. The amount of data generated and gathered is growing at an exponential rate, extending users' footprints. Data consolidators can link data items across data sources and combine data in ways that neither the parties who acquired the data nor the users who contributed it could have predicted. [6]

Below figure, which uses data provided by Statista, shows the cost of amassing these large databases. Statista, a statistical research firm, tracks cybersecurity failures and trends. A recently published Statista report reveals that these events are increasing, especially in the past five years, underscoring the need to improve how data is secured. [6]

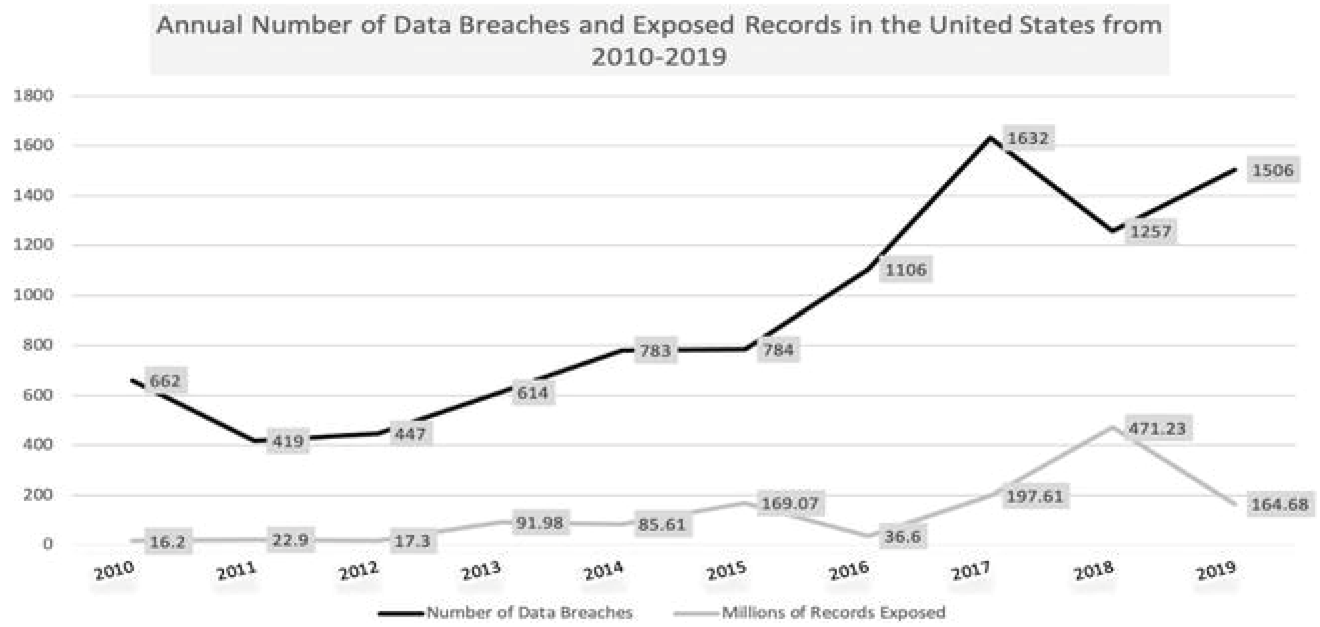


Figure 13: Cybersecurity breaches and record exposure. [6]

3.6.6 How can Blockchain help in reducing the data breach?

The numerous advantages of blockchain technology can help to reduce the risk of data breaches caused by a hostile hacker or a deceptive insider in a system. [6]

Establishing a decentralized sort of data storage with blockchain can protect sensitive data. Using this mitigation method, hackers would find it more difficult, if not impossible, to breach data storage systems. Many storage service companies are investigating how blockchain technology may protect data from hackers. Apollo Currency Team is a great example of a business that has already used blockchain technology into its operations (The Apollo Data Cloud). [7]

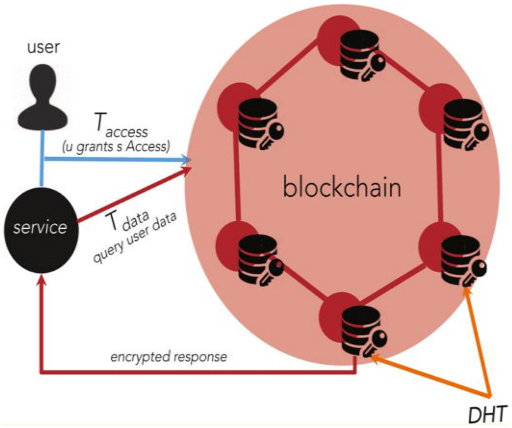


Figure 14: Overview of the decentralized platform. [5]

Similarly, access to private data is controlled on a public blockchain using a variety of approaches. Some of these approaches include hashing sensitive data, encrypting it, or using zero-knowledge proofs to make precise guarantees about the underlying data without revealing the data itself. [8]

Blockchain technology is built on encryption and distributed throughout a network of read-only computers, which keeps a record safe and improves security by functioning as an impenetrable wall of gatekeepers rather than a single company that may be targeted for security vulnerabilities. [8]

The primary benefit of blockchain is complete transparency. Any end-user can conduct a thorough audit of the system to confirm that it is in working order. There's no need to rely on faith when this is the case. Individual privacy is vital for self-sovereign identification, and blockchains allow entities to keep it. Users of a system or group of systems have traditionally had a federated identity, which can be defined as a single identity used by individuals to access services or information platforms provided by various parties, with single sign on (SSO) authentication enabling and determining a single identity. [6]

3.7 SQL Injection

SQL INJECTION is a code injection method that is used to attack the Data driven application inserted into the input field for execution of harmful SQL statements (for example, to cover the attacker's content). For example, SQL injection is selectively entered by user input, and if the user input is incorrectly lost if the user input is input to SQL statement or user input unexpectedly, the application software You need to exploit the vulnerability. SQL injection is almost known as an attack vector for a website but can be used to display any type of SQL database. SQL Injection Attacks allow you to deliver issues such as identifying, operation using existing data, and changes to the back of the transaction, and changing database servers.

3.7.1 Types of SQL Injections:

SQL injections typically fall under three categories: In-band SQLi (Classic), Inferential SQLi (Blind) and Out-of-band SQLi.

In-Band SQLi: In-band performance The most prevalent SQL injection attack is SQL injection. For attackers, this SQLi attack type is simple and effective. In this situation, the attacker uses the same communication channel to launch the attack and collects the results. This form of SQLi attack comes in two flavors:

Error-based SQLi: Due to the attacker's actions, the database generates an error message. Based on the data provided by these error messages, the attacker next gathers knowledge about the database infrastructure.

Union-based SQLi: Association-based SQL insertion occurs when an attacker utilizes the UNION SQL operator to retrieve the necessary data by combining numerous select statements into a single HTTP response. conclude.

Inferential SQLi: Attackers exploit behavior patterns and respond to the server's post-submitted data payload to learn about its structure in this sort of SQLi. The attacker cannot see the attack information in the band because the data is not sent from the website database to the attacker. The subclasses of reference SQLi are as follows:

a) Time-based SQLi: In this attack, the attackers send a SQL query to the database, which causes the database to wait (in seconds) before returning a yes or false response to the query.

b) Boolean SQLi: The attacker sends a SQL query to the database, which allows the program to react with true or false answers.

SQLi out of band:

When an attacker can't use the same channel to launch the attack and gather information, or when a server is too slow or unreliable to make these shares, this type of attack is used.

3.7.2 Threats in Cloud:

Cloud computing performs a critical position withinside the modernization of IT structures. As the speed, reliability, and flexibility of the cloud increase, companies maintain to increase new programs and pass current programs to cloud-primarily totally based offerings. Some associations pick out to increase their very own cloud technologies, are blind to the dangers involved, and will divulge their programs to attackers. SQL injection exploits used to assault conventional facts middle environments also can be used to assault cloud environments. This means that vulnerabilities might also additionally exist, and attackers try and take advantage of them, even though the cloud computing software program is going for walks totally withinside the cloud. Unauthorized use of cloud offerings can result in elevated malware infection or facts poisoning because it cannot shield belongings unknown to the association. In addition, it limits attention and manipulate of membership facts and information. Associations use those APIs to organize, manipulate, coordinate and music sources and customers. These APIs might also additionally contain programming weaknesses just like APIs which include operating frameworks and libraries. The API from person executives of the cloud API for registering localhost is to be had 24 hours a day, 7 days every week at the Internet, making it extra susceptible to capability exploits and assaults. Some of the above SQL injections can result in facts deletion. Attackers search for vulnerabilities withinside the control API. If those vulnerabilities are discovered, they may be a success assault and endanger the related sources. From this point, the attacker can use the cloud to release similarly assaults towards different cloud clients.

3.7.3 Enhancing SQL injections detection with Machine Learning:

One of the first and most widely used approaches for identifying SQL Injection threats is pattern matching. AMNESIA [1] is an example of a gadget that uses healthy sample procedures to categorize dangerous queries. The person entered is matched with the registered patterns. A likely attack is recognized if there is a mismatch. It incorporates both static and real-time assessment. Dynamic Tainting [2] is a similar approach that is dependent only on the health of the sample. Before being executed, the individual who enters is processed. When tainted records are utilized for a critical activity, such as recovering a password or gaining access to a file, Dynamic Tainting compares them to the previously registered records.

The Parsing Approach can also detect malicious web code. The person entered is parsed alongside SQL language in this method. SQL Check is unquestionably one of the Parsing Approaches. Every individual who enters is treated as a controversial question, and the SQL parser is used to evaluate the syntax of SQL queries. SQL Check flags argument queries as malicious if they don't match the registered statistics type. SQLrand [3] is another parsing method for preventing SQL injection. It employs an intermediary proxy that translates SQL to a more familiar language. The database parser becomes stuck and terminates the queries inserted by the attacker. SQL Guard [4] is another parsing method that works similarly to SQL Check. It does static verification by isolating a few points that ship SQL queries such as AMNESIA and parses non-malicious SQL queries from each point. When SQL queries are created, they're sent to the safety element, where they're dynamically exchanged. If a dynamically examined syntax does not match the equivalent statically analyzed syntax, SQL Guard flags the code as harmful. A mix of sample matching with a Role Based Access Control system [5] is one option among several.

It is a two-layered system, and in order for a question to be answered through the safety aspect, it must pass through all of the security levels. The inquiry in the main layer must be compatible with the static registered queries. The query must be compliant with the statistics type, function, and type for a suit to occur. RBAC provides a second layer of defense if the original layer fails. RBAC is founded on the fact that consumers are granted access to resources depending on their roles. If a certain operation or pastime is not permitted for a particular individual, that operation will no longer be possible.

For the identification of SQLIAs, another technique employs devices that learn algorithms [6]. The functions are extracted and policies are established as a result of learning about phase. Malicious and non-malicious code are included in the education set. The functions are recovered from the education set using any function extraction technique, such as the TF-IDF approach. Based only on the function vectors, the classifier ranks the provided code as malicious or non-malicious. The selection of functions is the most crucial component of this method. To improve efficiency, it's critical to extract functions as effectively as possible. One of the advantages of using devices to learn algorithms is that it gives you access to a wider range of SQL queries.

3.7.4 Methodology:

We propose a method for detecting SQL injection in this report. Using the Nave Bayes system and algorithm knowledge, our method detects SQLIAs. Machine learning algorithms provide reasonably accurate predictions on check records. Large records units benefit from their use as well. This is critical in our instance because there are several unique types of assault for which a specific sample cannot be collected.

Then, for each harmful and non-malicious query, we calculate the probability separately. A range of capabilities matching the specified check scenario and the total amount of harmful inquiries are used to calculate the likelihood of an inquiry being malicious. The probability of a query being non-malicious is

estimated using a number of factors of capabilities matching the given check case and the full number of non-malicious queries.

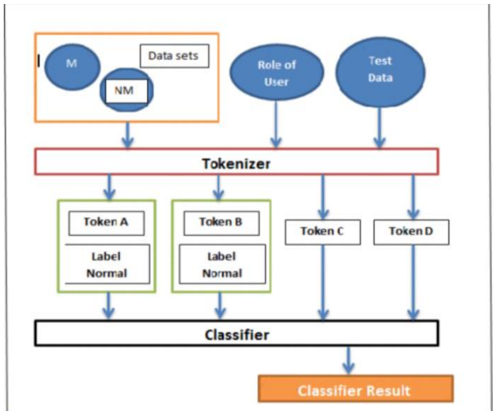


Figure 15: Token Classifier

Malicious and non-malicious inquiries are included in the data set. The capabilities are retrieved from the records units, and those capabilities are utilized to build policies for the category. The person's position is entered into the gadget, and it serves as one of the capabilities for the category. The check case, which is the question to be labeled, is also offered as enter. The tokenizer turns queries into capability tokens and categorizes them as harmful or non-malicious. This entry is sent to the classifier, which subsequently sorts the requests into harmful and non-malicious categories.

Feature Extraction:

Feature extraction is the method of remodeling the entered facts into the set of capabilities so that you can carry out the preferred venture. The usage of this decreased illustration in preference to the total length entered is referred to as function extraction. Each time period is extracted from datasets through any extraction approach. The maximum not unusual place approach is the Blank Separation Method. The clean separation approach extracts phrases with appreciation to every clean space. Number of prevalence of every time period is likewise calculated. For example, while the subsequent report is separated:

‘AND’’e’ != ‘e’ /**/;

Output is:

Name	**Terms**
AND	1
e	2
‘	6
!=	1
/**/	1
;	1

Table 5

Tokenization Method:

Breaking the query into elements is called tokenization. The list of tokens becomes input for further processing which is classification. Some of the tokens are:

Name	Token
Single Line Comment	- -
Multi Line Comment	/**/
Logical Operator	NOT AND OR &&
Punctuation	‘ ; “ [] () ,
Literal	‘string’ “string”
Operator	<> => >= == =! | & - + % ^

Table 6: Tokens of SQLIA

The output as shown in:

Name	Number of Terms
Punctuation	7
Numeral	2
Logical Operator	1

Multi Line Comment	1
Operator	1

Table 7

3.8. DDOS ATTACK

Due to COVID-19 and the prevalence of WFH (work from home) today, people spend a significant portion of their day on the internet, often without taking specific precautions to ensure a secure session. In addition, organizations worldwide that host data and conduct business over the internet are always vulnerable to DDoS attacks. These DDoS attacks are getting more extreme with hackers getting easy access to botnet farms and compromised devices as can be seen in the graph three of the six strongest DDoS attacks were launched in 2021 with the most extreme attack occurring just last year in 2020.

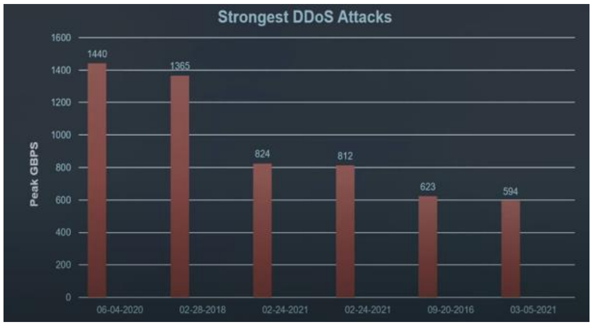


Figure 16: Strongest DDoS Attacks

The DDoS is an acronym for ‘distributed denial of service’. Which is a subdivision of the Denial of service. DDoS is a type of attack where there is a sudden heavy load of traffic on a network at a particular time aiming to use up all its bandwidth resulting in the denial of service of the server. Due to this the rate of speed of the Web Server or Domain response might decrease and sometimes the server shutdowns after a period which results in the restart of the server during the business time without any proper information to the customers. This DDoS attack can be achieved by attackers by simply clogging the available bandwidth of the network by continuous web requests to a particular domain or a server or by a huge number of continuous pings to the server.

One of the strongest attacks to date was on GitHub. It is one of the platforms for software developers to host their repositories available to the public or can keep in private. In the year 2018, GitHub went down for 20 minutes after which the servers or systems had been bought to a stable stage.

3.8.1 DDOS attack process

The DDoS attack is a two Phase. During the first phase, the hacker identifies which Web Server or Domain needs to be targeted. Once the Web Server or the Domain is decided. The hacker handles a few devices or computers in the network and injects them with malware or some sort of virus in the systems allowing them to be controlled anytime remotely by the attacker. The group of the devices is known as the botnet.

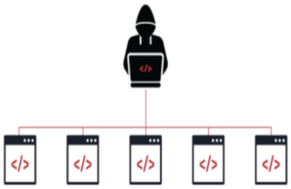


Figure 17: Hacker injecting malware into systems

Once the botnet is ready for the attack here comes the Second Phase. When a hacker identifies the right time to attack the Web server or domain. The attacker activates all the affected systems in the botnet network to send multiple API requests to the target server or continuous numerous ping to the server decided in the first phase, clogging the available bandwidth of the server resulting in the denial of service of the server. These multiple API requests or attacks can be SYN or UDP Flooding.

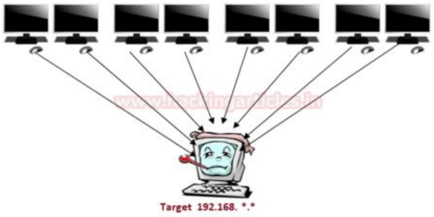


Figure 18: Bonnets targeting the server

3.8.2 DDOS attack in cloud computing

Unfortunately, even the fully secured cloud computing models such as IAAS, PAAS, and SAAS were also affected by hackers with DDoS attacks either internal or external cloud environments. The victims of the internal DDoS attacks are IAAS and PAAS layers. The reason for the internal cloud-based attack is due to the free trial provided by AWS or any other cloud-providing platforms. The cloud providers are providing free tails to make the user's environment friendly, but this had become the loophole for the attackers to penetrate the environment. Due to this the attackers log as an authorized users and infect the victim machines with the virus resulting from the bonnet phase of the DDoS attack.

Secondly in the external DDoS cloud-based environment attack, the victims of the external DDoS attack are SAAS and PAAS layers. In this type of attack instead of penetrating the user environment, the attacker tries to spread the virus to the entire Virtual machine hosted in the cloud. If the attacker is successful in spreading the virus it will spread to almost all the machines connected in the cloud and becomes a first phase of the DDoS attack as a botnet and once the best time is available, the attacker implements the second phase of the DDoS attack.

Whenever cloud security is ignored, it will become the origin of the many internal and external cloud-based DDoS attacks. This is briefly explained in the upcoming sections

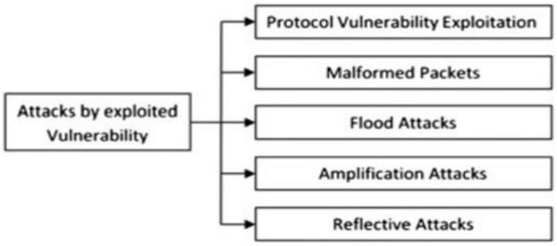

Attacks by exploited Vulnerability
Protocol Vulnerability Exploitation
Malformed Packets
Flood Attacks
Amplification Attacks
Reflective Attacks

Figure 19: Attacks by exploited Vulnerability

3.8.3 Types of DDOS attacks

There are four different types of DDoS Attacks. Which are given below.

1. Volume/Network-Based Attacks
2. Protocol Based Attacks
3. Application-Based Attacks
4. Fragmentation Attacks

3.8.4 Volume/network-based attacks

The term Network-based attacks itself states that the attacker attacks the available network i.e., by clogging the bandwidth of the target server thereby increasing the load on the server. This type of attack generally sends multiple short requests simultaneously thereby increasing the traffic over the network which results in the server down thereby affecting the system shutdown. Example UDP Floods, ICMP Floods.

- **UDP Floods**

This type of attack involves sending a large amount of spoofed data at a high packet rate to multiple random ports of a target server using the available IP range. Due to such an attack, all the available applications become inaccessible.

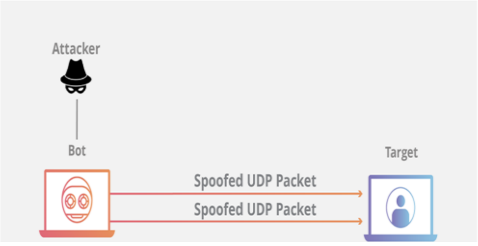


Figure 20: UDP Floods Attack

- **ICMP Floods**

This type of attack involves sending ping packets to the Web server as soon as possible without excepting any replies from the server. Due to repeatedly sending the requests as per the server design responsibility is to reply to each request. Hence due to multiple requests to the server and bandwidth is fully occupied resulting in a high load of the server and results in the server being down. Significantly overall system shutdown.

3.8.5 Protocol-based attacks

This type of attack tries to acquire the entire resource or middleware resources such as load balancers or Firewalls of the target server. These types of attacks typically exhaust the firewalls, windows defenders which are typically the software that is designed to protect against attacks such DDoS attack is one among them attacks. Examples of Protocol-based attacks are SYN Floods and Ping of Death.

- **SYN Floods:** The attacker sends many sync requests to a target web server through a fake Ip address. The target machine sends back the SYN-ACK as a response to the received request and waits for the acknowledgment to end the connection port. As the attacker uses the fake IP address, the acknowledgment never comes back. SYN attack takes the advantage of a flaw in the way most hosts implement a TCP 3 Way handshake when the host receives the SYN from the other for at least 75 seconds, the host must keep track of the partially opened connection in a listen queue. A malicious host can explore the small size of the listen to queue by sending multiple scene requests to the host but never replying to the SYN acknowledgment request. Hence the server is quickly filled up this ability to hold each incomplete connection for 75 seconds can cumulatively use a denial-of-service category and results in the SYN Flood attack.

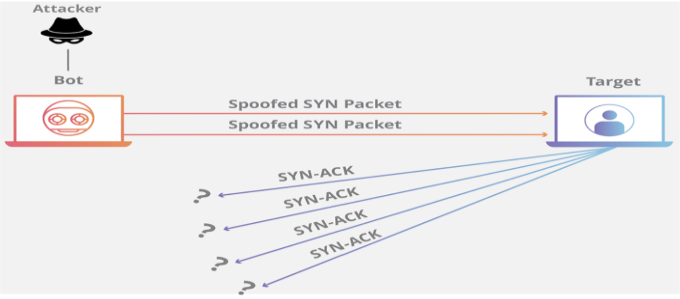


Figure 21: SYN Flood Attack

3.8.6 Application-based attack

The term Application based attack illustrates that the attack is on the application level or operating system level. To perform this attack, the attacker needs a high level of knowledge related to the overall application from the operating system to the end-to-end product. The main goal of the attack is to prevent the information to the users from the target machine by employing a heavy load on the bandwidth till the application/ system crash. Examples of the Application based attack are BGP Hijacking.

- **BGP HIJACKING:** The acronym of BGP is Border Gateway Protocol. The Border gateway protocol is a protocol, which is used to maintain the internet routing tables i.e., all the requests from the customer will be sent to the BGP at first. Based on the routing tables defined the request will be sent to the corresponding application. Here the attackers instead of penetrating the application, attack the BGP and culprit the internet routing information present in it. Thereby diverting the customer requests to the inappropriate application resulting from the loss of information. The BGP Hijacking is dangerous because even the owner will be unable to predict the BGP is Hijacked or not. This will result in a great loss of traffic towards the application and ultimately results in the revenue loss of the company.

There are many names for BGP Hijacking such as prefix hijacking, route hijacking, or IP hijacking. As all the names deal with the IP (internet protocol), routing, etc.

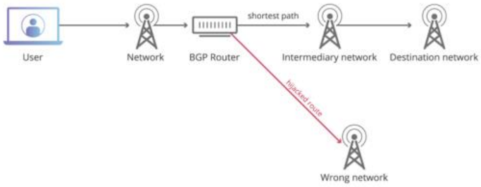


Figure 22: BGP Hijacking Attack

3.8.7 Fragmentation attack

The word fragmentation is defined as dividing the object into multiple tiny objects. In the concept of the network, whenever the data transmission is too large to travel between the source to destination, the data will be divided into multiple fragments and travel across the network. Once all the fragments had been reached the destination, they will combine to form single data. Similar to the attackers used this concept as a fragmentation attack i.e. The attackers send the multiple subparts of the data as fragments by having the same header to the destination usually slower than the regular transmission rate. Due to the very slow transmission rate, the server takes a long time to process the data than the usual rate which takes up all the resources indefinitely. Examples of the Fragmentation attack are the Teardrop attack.

- **TEARDROP ATTACK:** In this type of attack, the attacker sends multiple fragments of the data to the targeted server i.e., the header will be the same for all the fragments. Due to the Internet protocol vulnerability, sometimes the server will be unable to assemble the tiny fragments at the server. Due to multiple fragments, the load server gets increased and resulting in the system crash. Significantly overall system shutdown.

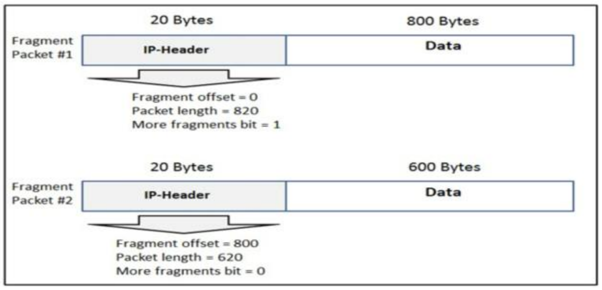


Figure 23: TEARDROP Attack

3.8.8 DDOS mitigation using machine learning and blockchain technologies

As technology advances, application functionality improves in parallel, applications become less secure, and attackers are more likely to attack them. For example, the IoT has revolutionized application processing by allowing attackers easy access to private data wirelessly. DDoS attacks are also increasing rapidly. Recently, the GitHub platform was attacked by DDoS and was unavailable to users
for a few minutes

Mahjabinetal. [18] Describes the rapid increase in DDoS attacks and their motives. The author conferred an analysis of DDoS attacks. Describes different mitigation techniques for handling DDoS attacks and how to use BOTNETS to execute similar attacks. The author outlined the different types of attacks, techniques of packet filtering, and methods for detection are performed.

Ni, et al. [20] conferred a method of using time series analysis to detect DDoS attacks. They proposed an architecture for DDoS attacks that would run out of bandwidth and resources. Their post contains an HTTP GET request for each system address to distinguish a DDoS attack. Detection methods using SVM classifiers provide high efficiency and flexibility in attack detection.

Using smart contracts and the blockchain, a collaborative approach to capturing the attacker's IP [22]. All blacklisted or malicious IP addresses are kept on the blockchain and made public to ISPs (ISPs). Before the attacker reaches the victim, the ISP is responsible for blocking the attacker's request.

Rashidi et al. [23] presented a method for reducing DDoS assaults that involved sending over-the-top traffic to other domains for filtering. NFV-enabled domain networks give additional system assets, allowing other users to have domains on the system as needed, thanks to network functions virtualization technology.

Random forest, logistic regression, KNN, and gradient boosting methods were used to detect intruders in the NSLKDD dataset by Dobson et al. [24]. They compared the outcomes of the machine learning and DL algorithms and found that the DL method predicts a higher level of correctness.

[25] Othman et al. To minimize processing time, we used SVM in the intrusion detection benchmark dataset using the chi-square feature selection method.

DVVS Manikumar, B. Uma Maheswari [34] were used to prevent DDOS attacks by employing machine learning and a blockchain of malicious or real packets to trace harmful requests. A system was proposed by me. The steps below have been presented as a model for the research findings above.

1) The target server sends real malicious requests to nearby auxiliary nodes, who then send them to distribute network partners [34].
2) The helper node must then determine whether the packet is malicious or authentic [34].
3) Machine learning techniques are used to classify packet maliciousness, with a classification model being performed on each node in the web [34].

4) The neighbor node tells the server, and the server establishes a connection with the client [34] when it finds that the packet is authentic.

5) If the neighbor node recognizes the request as malicious, then the time stamp and state (open / closed) of the blacklisted IP are stored in a smart contract on the Ethereum blockchain and distributed ledger.

6) Additionally, the malware IP is blocked for a certain duration after obtaining the blacklisted IP information recorded in the blockchain (value set by administrator). The server will unblock the IP [34].

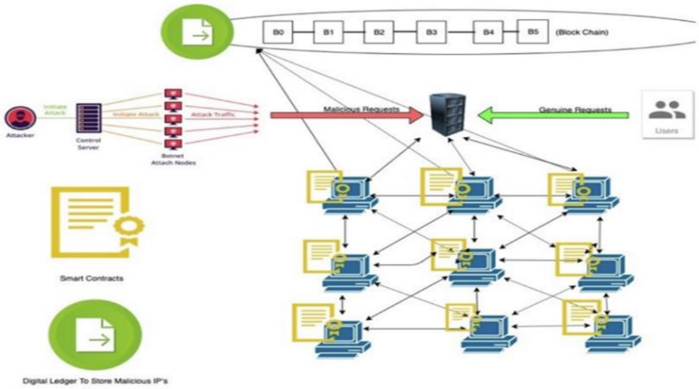


Figure 24: flow for DDoS Mitigation using Blockchain and Machine Learning

Account Hijacking

3.9.1 What is Cloud Account Hijacking?

Cloud computing is one of the most important technological breakthroughs in computing. It offers a variety of solutions to perform various computing tasks with benefits like scalability and affordability. With an increase in cyber-attacks involving malicious activities, it becomes imperative to monitor the cloud environments as well.

Cloud account hijacking is a prevalent method in identity theft schemes and is one of the most rapidly growing types of cyberattacks. It is a procedure in which an attacker or intruder steals an individual's or

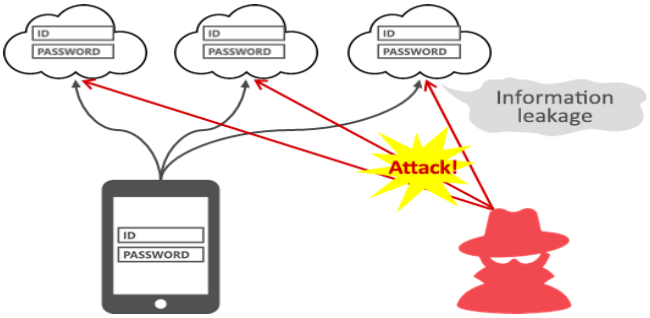


Figure 25: Representation of account hacking by an intruder.

organization's cloud account. It is a typical identity theft technique used by criminals to carry out unlawful behavior. The attacker uses an email account or other credentials to impersonate the account owner during cloud account hijacking. The most common scenario for cloud hijacking is when a corporation does not change its default cloud service password.

Because a huge amount of data is housed in one location, it becomes an attractive target for cybercriminals. There are numerous advantages to cloud computing, including the ability to create on-demand resources and the reduction of capital expenditures and expenses.

<u>Typically, a cloud hijacking process would look like this:</u>

- A firm does not change the cloud service’s default password.
- Someone from the firm ends up falling victim to the phishing attack.
- The cloud access credentials are exposed somewhere publicly.

For nefarious purposes, the account credentials of a lawful user are compromised during account hijacking. The accountability, confidentiality, and integrity of the cloud services are compromised with the stolen credentials fetched for a particular account. A wide range of techniques like fraud, phishing allows the attackers to hijack the credentials of a user account. A few possible ways of preventing account hijacking could be to allow users to enable multi-factor authentication and should be taught to mitigate the sharing of account-related passwords or credentials across users and cloud services. As cloud services are becoming popular by the day, it does add a new layer of threats to the aspect of service or account hijacking.

3.9.2 Cloud Account Hijacking Risks

One of the worst scenarios could be losing access to the company data. It could be held ransom for a hefty amount. There is no guarantee of the services returning even after having paid the ransom amount, the firms are completely locked out of their data and applications. The services have to be restored afresh and data must be properly recovered from existing backups.

In a recent survey, it is seen that more than 69% of American IT professionals believe that the risks associated with cloud-related services outweigh the benefits. One of the main reasons quoted by them was the need or concern for data security. Similarly, in a 2013 analysis, the Cloud Security Alliance associated service cloud traffic hijacking as one of the greatest cloud computing security risks. These

breaches happen when intruders hijack cloud accounts by capturing security credentials or eavesdropping on transactions, conversations, or activities.
Intruders generally try to redirect clients to illegitimate sites, manipulate data, and try to inject falsified information.

In less extreme situations, a firm could again lose up to few hours or days of their work if their working cloud is found to be hijacked. It becomes a tedious process to update the servers once the access is restored. Along with the loss of money, a company could also lag behind in deployments, orders, and production-related processes, suffer huge reputational damage, loss of confidential data, data leakage, or falsification of information and as a result, lose out on orders.

At enterprise levels, cloud account hijacking can be extremely devastating. There could be the legal implications for organizations and firms in a highly regulated industry, such as banking if clients' confidential data is exposed by intruders during the incident of cloud account hijacking.

3.9.3 Simple Solutions for Cloud Account Hijacking Protection

Few proactive measures can be taken by organizations to ensure their data is kept securely on the cloud. As we know, the users of cloud computing networks generally witness account hijacking. The intruders somehow find their way into hijacking the user account even though the cloud owners/providers take all the preventive measures in this regard.

Here are some proactive measures that can be taken by organizations to prevent exposure to security threats caused by hackers: and to keep their data secure on the cloud:

- All sensitive information like credentials/passwords should be well encrypted before it goes to the cloud.
- Multi-factor authentication should be set up as and when possible. It basically requires users to enter static or dynamic passwords that are delivered via SMS, biometrics, hardware tokens, or other schemes.
- The IP address range should be restricted. There are certain tools that allow only specified IP ranges, which ensures that the users access the firm's resources only via VPNs or corporate networks.
- Backing up of data must be done securely, in case of data loss in the cloud
- The authentication method must be strong for cloud-related apps

- The background verification of the employees must be done prior to giving them physical access to the servers in the data centers.

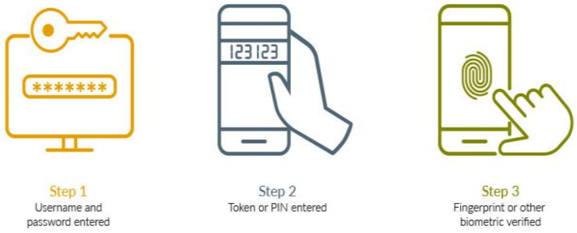

****
123123
Step 1
Username and password entered
Step 2
Token or PIN entered
Step 3
Fingerprint or other biometric verified

Figure 26: Depicting MFA flow

- A protocol must be set up so that account credentials between employees and services are shared securely.
- A method to understand the policies and SLAs of the cloud provider, before signing in a cloud network. They can combat the effects of exposure of client account details and data leakage to external intruders. The firms must be made aware of these techniques as well as common defence in-depth protection strategies in order to ensure the damage is contained as a result of the breach.
- proactive monitoring techniques should be employed to detect unauthorized activity.

3.9.4 More secure solutions for Cloud Account Hijacking Defense

A more secure technique to protect data is to enable end-to-end encryption, continuous data monitoring, and the ability to restrict risky data activity based on contextual or behavioral parameters, notably the event, user, and data access type. This comprehensive method enables firms to address cloud-related security problems while simultaneously maximizing the benefits of cloud computing.

Users must also be vigilant enough to understand and not click on random links attached to the email messages. There might be a high chance of these links injecting a malicious virus into the computer system and using the sensitive information for illegitimate purposes.

One must ensure that prior to submitting any personal information, the security certificate of the website must be thoroughly checked and verified. The system must be kept up to date with the latest anti-virus software and one must avoid entering any personal information into pop-up windows. Moreover, Users must be trained to click on random links attached to email messages which might inject a virus into the computer and the sensitive information of the user might be hijacked for illegitimate purposes. Several times attackers might set up a faux pop-up window and direct the users to the original website. On clicking these pop-up windows, the sensitive information of the user gets stolen. [38]

3.9.5 Machine Learning Techniques to prevent Cloud Account Hijacking

Account Hijacking attacks have become extremely troublesome for the cyber world. It poses a great threat to the privacy and financial issues of internet users. Scammers, think of ways to create fake websites to deceive people. They ensure that emails are spoofed to steal the identity of legitimate users.

Certain ML algorithms have been devised to detect phishing performed through Machine Learning. Fishers always find various strategies to intrude the system. They tend to gather credentials from trustworthy people with the help of a technique called Social Engineering [39]. Phishers try various manipulative ways to create false websites or spoof emails that look very real and similar to the real company. The user's information is highly exploited, and for these reasons, it is of utmost importance to ensure that phishing in today’s world is highly challenging, critical, and urgent. Based on the characteristics of a domain, several studies have been carried out to study the websites like website URL, source code of the website, website content, and the screenshot of the website.

However, the users are always prone to attackers stealing their information as there is a lack of developed anti-phishing tools to detect malicious URLs in a firm. If a malicious code is seen to be injected into the website, then it makes it easier for hackers to steal information, which poses a high risk to user privacy. Using Machine Learning, the malicious or suspicious URLs of the websites can be easily identified by accurate analysis. The approach taken is mainly based on the blacklist is mainly used to verify an URL and this list is often updated. The count of the list keeps increasing every day significantly and these cybercriminals continue to use various algorithms to ensure the blacklist is circumvented by generating new URLs. Therefore, it becomes an exhaustive list of malicious URLs making it extremely difficult to identify the malicious URL. The new malicious URLs cannot be identified with the already existing techniques. Based on computer learning, the researchers have suggested a few methods to capture or identify malicious URLs. In compared to the blacklist approach, our proposed solution has a greater speed and optimization capability for identifying unknown malicious URLs.

RNN (Recurrent Neural Network) - LSTM (Long Short-Term Memory) is a machine learning technology that provides the best answer for real-time complicated situations. The use of LSTM and RNN together allows for the extraction of a large amount of data from a small number of data. With the help of LSTM, RNN can retain inputs for a longer period of time. Storage is very similar to the concept of computers.

The features that are present will be treated in a uniform manner.It primarily focuses on detecting harmful URLs among a large number of URLs. The research focuses on an RNN-based URL detection method. Its major goal is to do the following:

- A innovative method for detecting dangerous URLs and alerting user
- Implementation of the RNN concept, which is a well-known Machine Learning technique capable of handling large amounts of data.
- The suggested solution uses machine learning techniques to analyze real-time URLs and produce useful results

3.9.6 Blockchain solutions to prevent for cloud account hijacking

Many aspects of blockchain, also known as distributed ledger technology, make it impossible to hack a well-designed system. In business networks, this technology can be used to fight and prevent fraud. It has the potential to share data in an exceptionally rapid, dependable, and secure manner without requiring any one institution to assume responsibility for data security. The capacity of blockchain technology to provide greater security is one of the key advantages of employing it. Due to the major properties of this technique, the Advance Encryption Standard (AES) is also utilized to encrypt and decrypt the data.In a business network, blockchain can be used to fight and prevent fraud. One of the key features that determines the usefulness of blockchain is its capacity to share data quickly and securely without requiring any one institution to accept responsibility for data security. One of the most significant advantages of blockchain technology is increased security.

To begin with, as the term "distributed ledger" suggests. There is no single point of failure in the blockchain. Once a node or server rack has been infiltrated, no single node or server rack can reveal data to attackers. Attackers cannot truly modify or replace a blockchain network since any incorrectly updated node would be challenged by the rest of the network. Second, the blockchain is based on a highly complex encryption algorithm.

4. Conclusion

4.1 Conclusion and recommendations for Malware Detection:

In this study, we looked at several machine learning approaches to see which one is the most effective for detecting malware in the cloud. The DenseNet-121 (CNN) model has the best overall performance, even though it takes the most time to train. The SVC and RFC models generated promising results that were not far behind those of the CNN model and were considerably easier to train.

When it comes to malware detection, however, spending more effort to train a more accurate model is generally favored. KNN, GBC, and GNB, the remaining models, were just unable to compete with the others. The success of the CNN model suggests that deep learning models are better at detecting malware.

Blockchain technology offers a new form of trading that is built on important technologies including password security, decentralized coherence, shared public accounts, and visibility of necessary controls and permissions. By registering and transferring real and virtual, tangible and intangible assets, it has the potential to profoundly transform the way our society creates and lives.

The potential to decentralize is a critical feature of blockchain technology, regardless of how it is used. This feature eliminates the one point of vulnerability that could be exploited. As a result, infiltrating systems or facilities where access control, data storage, and network traffic are no longer in a single area becomes nearly impossible.

Personal and sensitive data should not be entrusted to third-party tools and technologies, as they are vulnerable to hacking and misuse. Instead, people should own and control their data without jeopardizing security or limiting the ability of businesses and governments to provide individualized services. Making legal and regulatory judgments about acquiring, storing, and sharing sensitive data should also be easier with a decentralized platform. Furthermore, laws and regulations might be built directly into the blockchain and enforced automatically. [48]

Because blockchain technology can withstand typical cyber-attacks so successfully, fraudsters are investing extra effort into developing a new vector of attack designed specifically to destroy the technology in question. On the contrary, the prospective costs of launching a cyber-attack are likely to be much lower than the costs of defending against such attacks or developing long-term solutions based on Blockchain and dealing with the aftermath. Additionally, the lack of globally supported standardization and regulation of the technology, as well as the dearth of parallel processing and many ethical and administrative issues that Blockchain raises with the unknown identity of data traffic, should be identified in order for the technology to attain widespread adoption. [47]

4.2 Conclusion and recommendations for SQL injection:

This study investigates the potential outcomes of applying AI to recognize destructive dangers. In the field of network protection, AI holds a ton of guarantee. The exactness of most classifiers is around 98%. The irregular timberland classifier outperforms all others on the dataset and arrives at 99.8% exactness, which is a conspicuous pattern. The three best elements found are the length of the info, the quantity of accentuation letters, and the quantity of various bytes. Vindictive composes are more risky than pernicious peruses. Therefore, our models have been customized to recognize perused and compose activities.[44]

4.3 Conclusion and recommendations for DDOS:

D.V.V.S. Mani Kumar, B Uma Maheshwari [35] proposed a system that is used to mitigate the DDOS attack using Machine learning to malicious or genuine packets and blockchain to keep track of the malicious requests. Based on the CICDDoS2019 Dataset, generated by the canadian institute of cybersecurity which consist of numerous current shortcomings.

Based on the dataset using the KNN machine learning algorithm i.e we can detect the malicious or genuine packets at an accuracy of 87.34% whereas using Decision Tree Classifier at an accuracy as 93.83% finally the best algorithm is Random Forest Algorithm which provides and accuracy of 95.19%. After identifying the malicious IP’s and blocking list over a period time using the block chain framework i.e. the public Ethereum blockchain which is used to hold the reported IP addresses in the blockchain running over the smart contracts would be an effective to detect and blacklist all the malicious IP address so that the server or system can’t netwget attacked by DDoS further more. But due to increasing the technology, the attackers finding the many loopholes in the technology and using as them as advantage to attack the server.

As this study proves the defense system had been more strongly developed in detecting the malicious IP using ML Algorithms with the accuracy as stated above and adding the blockchain framework as an advantage. Can also be improved by adding few Deep Algoritoms to get more defensive by getting more accuracy in the detecting the malicious IP Address.

4.4 Conclusion and recommendations for Account Hijacking:

As cloud services are increasing in importance, they also pose unique challenges and risks. Cloud service providers must ensure strict policies are followed to prevent the hijacking of accounts by intruders. They must be taught how to use strict authentication mechanisms if they are necessary. The user authentication process should not rely heavily on publicly available data. In conclusion, regardless of the nature of the incident, having an incident response strategy in place is critical.

Additionally, most cloud systems are a multi-vendor environment, and the Virtual Machines can be migrated live across systems. The attacker or intruder sees this as an opportunity to attack, disrupt, steal confidential information, or gather information on the VM in transit. These also relate to the fact that most of the cloud-related implementations are at a large-scale, highly interconnected making them vulnerable to attacks and use multiple technologies that may be internally exploited or vulnerable.

This makes it equally challenging in maintaining both sovereignty and allow full client access. Blockchain technologies and Machine Learning techniques provide a secure method to prevent attackers from intrusion the cloud servers. Machine learning with the help of RNN and LSTM provided an optimal way of filtering malicious website URLs. Blockchain technology coupled with ciphertext access control technology seems like a promising technology that would be implemented in solutions in this space to trust cloud nodes and intrusion detection engines. There is scope to work further in mitigating cloud-related account hijacking.

For future implementations of Ciphertext access control technology, it is highly recommended to use Ciphertext access control technology as a Blockchain technique for cloud-related attacks by intruders. By Data Confidentiality, we mean that only verified and authenticated users can access or obtain data explicitly. Encryption is one of the most common ways to protect data. Users, generally encrypt data before it is transferred to the cloud. One of the important strategies to ensure the legality of information use is Access Control. It also deals with network security and system resource protection. Based on different access control policies, the subject implements access to its resources depending on the access control model. The data stored in the cloud is usually in the ciphertext state, therefore, the user’s access to it becomes a ciphertext access control problem.

The ciphertext [40] access control technology encrypts the key information and controls the user’s access and the access rights of the key information. It is an important means to help ensure the confidentiality of user data in the cloud untrusted environment, it can not only improve the privacy of user data but at the same time ensure that the risk of user data being leaked is reduced to a great extent.

ABAC (attribute-based ciphertext access control)[40] is a combination of an attribute-based encryption algorithm (ABE) with access control technology that permits data to be accessed only if the user requesting access meets the attribute's judgment rules. ABAC is more suited for cloud storage environments with multiple tenants and frequent permission changes because it has more flexibility and finer access control granularity. The user can only decrypt in the KP-ABE and CP-ABE schemes if the attribute set matches the access tree.

CP-ABE is more suitable for cloud storage access control systems than KP-ABE because of the high resource sharing, high system openness, and high data dynamics in the cloud storage environment.

4.5 Conclusion and recommendations for Network Security:

For network threat detection, we have researched two separate approaches, one using machine learning and the other using blockchain. The traditional systems can only detect the attacks that have occurred before. But, the proposed systems involving machine learning and blockchain techniques can detect the occurrence of completely new attacks. The machine learning approach is concerned with the analysis of the network traffic data to determine if it is a malicious record. The machine learning algorithms can be used to learn traffic patterns, analyze encrypted traffic data elements, etc., which can then be used to detect malware, threats, suspicious behaviors, and cluster security events. The blockchain approach proposes the deployment of an intrusion detection system over the network to form a distributed system of IDS.

We recommend a solution wherein we use a combination of these above-mentioned two approaches. The proposed machine learning IDS system can be deployed on the individual nodes in the distributed network using the cloud infrastructure. This system would be capable of capturing the nodes, cleaning the data, and preprocessing it to get it ready for analysis, additionally identifying which packets are malicious and which ones are normal, thereby blocking the malicious records from getting access to the system.

Furthermore, the logs of this system will be handled by the blockchain. The alerts generated by this system will be stored as a transaction using the blockchain-based infrastructure wherein the nodes will generate the alerts, and the alert will be added to the blockchain only if the centralized nodes validate the alerts.

However, the only modification that we suggest here would be that instead of keeping just a centralized validating node we will have multiple nodes for validation of the alerts to keep the system secure. This will result in better security of cloud environments as we are making the process of network analysis smarter and more secure using machine learning and block-chain. The future scope of this system would be the analysis of the performance, cost of implementation, etc. of the system.

References

1. H. Gupta and D. Kumar, "Security Threats in Cloud Computing," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1158-1162, doi: 10.1109/ICCS45141.2019.9065542.
2. A.Annie Christina, "Proactive Measures on Account Hijacking in Cloud Computing Network," Asian Journal of Computer Science and Technology ISSN2249-0701 Vol.4 No.2,2015,pp.31-34
3. Shrestha,Prajwal. (2019). Impact of Blockchain in Data Security. 10.13140/RG.2.2.13472.10242.
4. G. Zyskind, O. Nathan and A. '. Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data," 2015 IEEE Security and Privacy Workshops, 2015, pp. 180-184, doi: 10.1109/SPW.2015.27.
5. S. Heister, and K. Yuthas, "How Blockchain and AI Enable Personal Data Privacy and Support Cybersecurity", in Advances in the Convergence of Blockchain and Artificial Intelligence. London, United Kingdom: IntechOpen, 2021 [Online]. Available: https://www.intechopen.com/chapters/75936 doi: 10.5772/intechopen.96999
6. Yaqoob, I., Salah, K., Jayaraman, R. et al. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-020-05519-w
7. Julien Legrand, The Future Use Cases of Blockchain for Cybersecurity, Cyber Management Alliance, September 2020, https://www.cm-alliance.com/cybersecurity-blog/the-future-use-cases-of-blockchain-for-cybersecurity
8. Tal Kol, Why blockchain is the necessary safeguard against data breaches, January 2020, https://www.securityinfowatch.com/cybersecurity/article/21122947/why-blockchain-is-the-necessary-safeguard-against-data-breaches
9. Rabbani, Wang, Y. L., Khoshkangini, R., Jelodar, H., Zhao, R., & Hu, P. (2020). A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing. Journal of Network and Computer Applications, 151, 102507–. https://doi.org/10.1016/j.jnca.2019.102507
10. Rabbani, Wang, Y., Khoshkangini, R., Jelodar, H., Zhao, R., Ahmadi, S. B. B., & Ayobi, S. (2021). A review on machine learning approaches for network malicious behavior detection in emerging technologies. Entropy (Basel, Switzerland), 23(5), 529–. https://doi.org/10.3390/e23050529
11. Jabez, & Muthukumar, B. (2015). Intrusion Detection System (IDS): Anomaly Detection Using Outlier Detection Approach. Procedia Computer Science, 48(C), 338–346. https://doi.org/10.1016/j.procs.2015.04.191

12. Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," 2017 IEEE International Congress on Big Data (BigData Congress), 2017, pp. 557-564, doi: 10.1109/BigDataCongress.2017.85.
13. A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, “Hawk: The blockchain model of cryptography and privacy-preserving smart contracts,” in Proceedings of IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 2016, pp. 839–858.
14. N. Sanghi, R. Bhatnagar, G. Kaur and V. Jain, "BlockCloud: Blockchain with Cloud Computing," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 430-434, doi: 10.1109/ICACCCN.2018.8748467.
15. Swan M. Blockchain: Blueprint for a New Economy. — O‘Reilly Media, Inc.‖; 2015.
16. N. Sanghi, R. Bhatnagar, G. Kaur and V. Jain, "BlockCloud: Blockchain with Cloud Computing," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 430-434, doi: 10.1109/ICACCCN.2018.8748467.
17. Tasnuva Mahjabin, Yang Xiao, Guang Sun, Wangdong Jiang, "A survey on distributed Denial-of-service attack prevention and mitigation Techniques”, International Journal of Distributed Sensor Networks, vol. 13, no. 12, 2017.
18. M. obinson, J. Mirkovic, S. Michel, M. Schnaider and P. Reiher, "DefCOM: defensive cooperative overlay mesh," Proceedings DARPA Information Survivability Conference and Exposition, Washington, DC, USA, 2003, pp. 101-102 vol.2.
19. T. Ni, X. Gu, H. Wang and Y. Li, "Real-time detection of application-layer DDoS attack using time series analysis", J. Control Sci. Eng., vol. 2013, pp. 4, Aug. 2013
20. DDoS Open Threat Signaling (DOTS) Authorized licensed use limited to: Cornell University Library. Downloaded on September 03,2020 at 11:47:19 UTC from IEEE Xplore. Restrictions apply. http://ftp.kfki.hu/documents/iana/drafts/draft-doron-dots-telemetry00.pdf
21. Rodrigues Bruno et al., "A Blockchain-Based Architecture for Collaborative DDoS Mitigation with Smart Contracts", IFIP International Conference on Autonomous InfrastructureManagement and Security, pp. 16-29, 2017.
22. B. Rashidi, C. Fung and E. Bertino, "A Collaborative DDoS Defence Framework Using Network Function Virtualization," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 10, pp. 2483-2497, Oct. 201.
23. Dobson, Anthony, Kaushik Roy, Xiaohong Yuan, and Jinsheng Xu. "Performance Evaluation of Machine Learning Algorithms in Apache Spark for Intrusion Detection." In 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), pp. 1-6. IEEE, 2018.

24. S.M. Othman, F.M. Ba-Alwi, N.T. Alsohybe and A.Y. Al-Hashida, "Intrusion, detection model using machine learning algorithm on Big Data environment", Journal of Big Data, vol. 5, no. 34, 2018.
25. G. Oikonomou, J. Mirkovic, P. Reiher and M. Robinson, "A Framework for a Collaborative DDoS Defense," 2006 22nd Annual Computer Security Applications Conference (ACSAC'06), Miami Beach, FL, 2006, pp. 33-42, doi: 10.1109/ACSAC.2006.5.
26. J. Dheeraj et al., "DDoS Mitigation Using Blockchain",International Journal of Research in Engineering Science and Management, vol. 1, no. 10, October 2018
27. Z. Martinasek, "Scalable DDoS Mitigation System for Data Centers", Advances in Electrical and Electronic Engineering, vol. 13, no. 4, pp. 325-330, Nov. 2015.
28. Prasad, M & V, Prasanta & Amarnath, C. (2019). Machine Learning DDoS Detection Using Stochastic Gradient Boosting. International Journal of Computer Sciences and Engineering. vol.7, no. 4, pp. 157-166, 2019.
29. Li, Qian & Meng, Linhai & Zhang, Yuan & Yan, Jinyao. (2019). DDoS Attacks Detection Using Machine Learning Algorithms. 10.1007/978-981-13-8138-6_17.
30. A. A. T. Innocent and G. Prakash, "Blockchain Applications with Privacy using Efficient Multiparty Computation Protocols," 2019 PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS), Bangalore,India, 2019, pp. 1-3, doi: 10.1109/PhDEDITS47523.2019.8986954.
31. Gopalakrishnan, Prakash & B., Uma Maheswari. (2018). “Keyless Cryptosystem for Secure Primitive Pairing of Mobile Devices,” Journal of Computational and Theoretical Nanoscience, vol.15, no. 5, pp. 1607-1614. 10.1166/jctn.2018.7349.
32. Mugunthan, S. R. (2019). “Soft Computing Based Autonomous Low Rate DDoS Attack Detection And Security For Cloud Computing”. Journal of Soft Computing Paradigm (JSCP), 1(02), 80-90.
33. DDoS Evaluation Dataset (CICDDoS2019) https://www.unb.ca/cic/datasets/ddos-2019.html
34. D. V. V. S. Manikumar and B. U. Maheswari, "Blockchain Based DDoS Mitigation Using Machine Learning Techniques," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 794-800, doi: 10.1109/ICIRCA48905.2020.9183092.
35. A, Pasumponpandian & S., Smys. (2019). “DDoS Attack Detection in Telecommunication Network Using Machine Learning”. Journal of Ubiquitous Computing and Communication Technologies. 01. 33- 44. 10.36548/jucct.2019.1.004. [18] Mugunthan, S. R. (2019). “Soft Computing Based Autonomous
36. A.Annie Christina, “Proactive Measures on Account Hijacking in Cloud Computing Network”, Asian Journal of Computer Science and Technology ISSN 2249-0701 Vol. 4 No. 2, 2015.

37. S. S. Tirumala, H. Sathu and V. Naidu, "Analysis and Prevention of Account Hijacking Based INCIDENTS in Cloud Environment," 2015 International Conference on Information Technology (ICIT), 2015, pp. 124-129, doi: 10.1109/ICIT.2015.29.
38. M. N. solu, D. Sarma, F. F. Lima, I. Saha, R. -E. -. Ulfath and S. Hossain, "Phishing Attacks Detection using Machine Learning Approach," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 1173-1179, doi: 10.1109/ICSSIT48917.2020.9214225.
39. Dutta AK (2021) Detecting phishing websites using machine learning technique. PLoS ONE 16(10): e0258361. https://doi.org/10.1371/journal.pone.0258361
40. Peng Cheng Wei, Dahu Wang, Yu Zhao, Sumarga Kumar Sah Tyagi, Neeraj Kumar,Blockchain data-based cloud data integrity protection mechanism, Future Generation Computer Systems,2020.
41. M. Kumar and A. K. Singh, "Distributed Intrusion Detection System using Blockchain and Cloud Computing Infrastructure," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 248-252, doi: 10.1109/ICOEI48184.2020.9142954.
42. S. Ghribi, A. M. Makhlouf and F. Zarai, "C-DIDS: A Cooperative and Distributed Intrusion Detection System in Cloud environment", 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 267-272, 2018.
43. D. Tripathy, R. Gohil and T. Halabi, "Detecting SQL Injection Attacks in Cloud SaaS using Machine Learning," 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 2020, pp. 145-150, doi: 10.1109/BigDataSecurity-HPSC-IDS49724.2020.00035.
44. https://www.opengrowth.com/article/cloud-malware-types-of-attacks-and-security-measure
45. https://cloud.netapp.com/blog/blg-cloud-malware-5-types-of-attacks-and-3-security-measures
46. https://doi.org/10.48550/arXiv.2105.09268
47. https://doi.org/10.48550/arXiv.2203.09938
48. http://doi.org/10.22214/ijraset.2020.6015