Deep Science Publishing
https://doi.org/10.70593/978-93-49307-78-0_2

Chapter 2

Background study of DDoS attack

2.0 Background

This chapter is divided into two parts namely the theoretical and experimental background.

2.1 Theoretical background

The section primarily entails an understanding of the following factors: Overview of DDoS attacks, Classification of DDoS attacks, DDoS attacks operating techniques, Cloud computing system, DDoS threats on a cloud system, Owncloud platform, DDoS attack in cloud networks, method of DDoS attacks in the cloud environment DDoS attack tools used in the cloud system, owncloud platform, programing language and machine learning. This is because the project's goal is to develop DDoS detection systems for cloud environments aided by machine learning.

Although the in-depth part of this section has been discussed in the investigation report (CMM512). The section is summarised the listed factors below.

2.1.1 Overview of DDoS Attacks

One of the most popular types of cyberattacks today is the denial of service (DoS) attack, which results in various financial impacts and losses for the target party. Denial of service is a term used in network and computer security to describe an attack that aims to overload or destroy the computer or network resources to prevent legitimate users from using the services that are being offered (Bensalah et al. 2019). Occasionally error messages are encountered when attempting to access a website because the server hosting the website is overloaded. When a server's capacity is exceeded by the number of requests it can handle, the attack occurs (Gaurav et al. 2022). The most common type of DoS attack is called DDoS, or Distributed DoS, which is formally characterized as a coordinated attack due to its capacity to produce more catastrophic impacts simply and quickly. The attacks begin and interrupt attack activities on a big scale by bombarding the target network devices or web services with information requests from thousands of infected host machines (zombies) (Efe 2018).

2.1.2 Classification of DDoS Attacks

Kiourkoulis (2020) suggested that DDoS attacks are split into 3 major classes as illustrated in Figure 2.

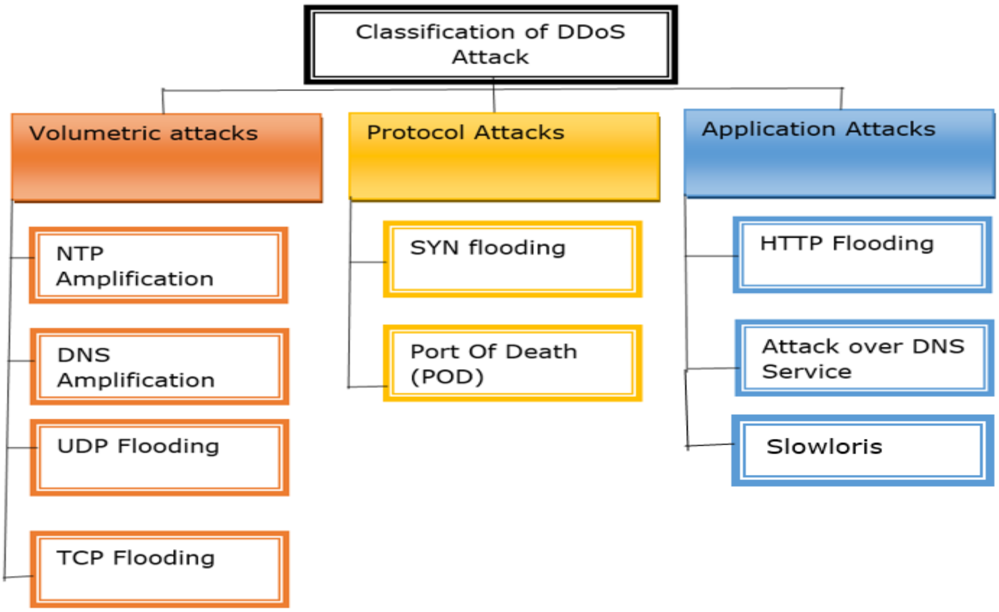

Figure 2: Classifications of DDoS attack.

2.1.3 DDoS attack operations technique

There are two main components of a DDoS attack namely bot and botmaster. Operations of DDoS attacks are divided into three settings namely attacking, handling, and target system which are based on the behaviour and functions of the components performed throughout the attack phase (Lohachab and Karambir 2018).

Attacking Systems: The botmaster initially chooses a botnet depending on its efficiency and structure, based on the sort of attack. The bot, often known as malware, is a package that contains web server files such as SQL templates, graphics scripts, and other files that can be used as a C&C server. The botmaster then selects agents or discovers new infected devices with the help of the bot by exploiting security vulnerabilities caused by hard-coded passwords, essentially weak security, lack of software upgrades, and so on (Cil et al. 2021). Various scanning technologies can detect all these potential security gaps that were left exposed during various stages of the design process. Gaining control of the infected devices during the selecting phase is critical to acquiring sufficient resources and making the attack more potent. Upon successfully gaining control of the device, the bot attempts to forward certain attributes to the report server over a different port, preventing the malware from being identified and deactivated using the rootkit feature. Frequent web scanning by bots aids the attacker in discovering more and more weak machines (Cao et al. 2018).

Handling Systems: These three components Command and Control server, report server, and loader play the role of handlers in the handling system. Most botnets are based on Internet relay chat (IRC), although new research reveals that HTTP-based botnets are gaining popularity (Mahjabin et al. 2017). The HTTP protocol is used for communication between the attacker and the C&C server. The C&C server is a central core component of the botnet that is responsible for sending special commands to carry out the attack. It includes a MySql database of all compromised machines and is a central part of the botnet that is responsible for sending special commands to carry out the attack (Praseed and Thilagam 2021).

Target Systems: Two components that make up the target or attack setting are newly infected devices and the target server. Self-propagating technologies like the Ramen worm and Code Red automate the conversion of target machines into infected machines or zombies after exploiting security holes. Even though they are both outdated worms, they are still useful in attack propagation. When a user executes a crack file or a key generator that contains malicious executable files, the bot is also launched (Lohachab and Karambir 2018).

Figure 3 below illustrates the operation concept of a DDoS attack.

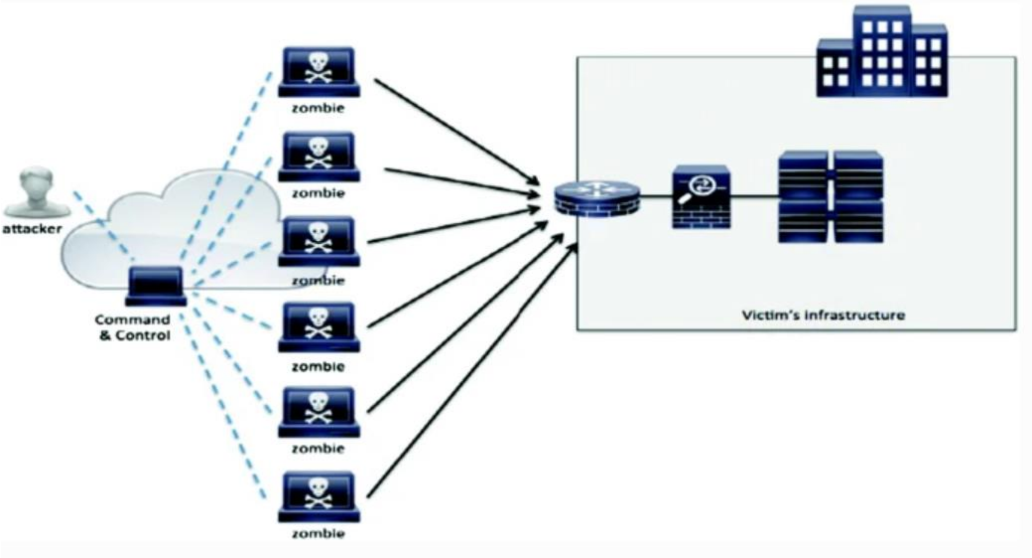

Figure 3: DDoS Attack operating concept (Sharma et al. 2019).

2.1.4 Cloud Computing System

Cloud computing is a utility computing approach that allows users to access resources in a "pay-as-you-go" manner. A Cloud provider is an "Infrastructure as a Service (IaaS)" provider that creates virtual machines on demand while a service provider is a cloud user who has deployed a web service as a virtual machine (e.g., an e-commerce application) in the cloud infrastructure by a cloud service provider (Alarqan et al. 2020).

Cloud providers are concerned that users would exploit their services and launch cyberattacks. The expectation of their companies and the services they will receive from a specific provider are the most important factors in choosing a Cloud provider (Khoda Parast et al. 2022).

There are major vulnerabilities in Cloud computing that can pose substantial dangers. For instance, virtual machine flaws, internet protocol weaknesses, unauthorized access to a company network, injection security flaws, and vulnerabilities in browsers and APIs (Kushwah and Ranga 2022). These flaws have ramifications, like permitting network attacks, providing invaders access control, enabling unwanted service access, and exposing sensitive information are flaws that expose cloud systems to dangers directly or indirectly. Nassif et al. (2021) suggested some adjustments in a business model that can restrict the use of Cloud computing services which are:

i. Abusive use of Cloud computing

ii. Insecure interfaces and APIs

iii. Malevolent employees

iv. Data loss and spillage

v. Security weaknesses

vi. Unclear risk level

A typical cloud computing environment is depicted in Figure 4 displaying where VMs are run on a huge number of servers.

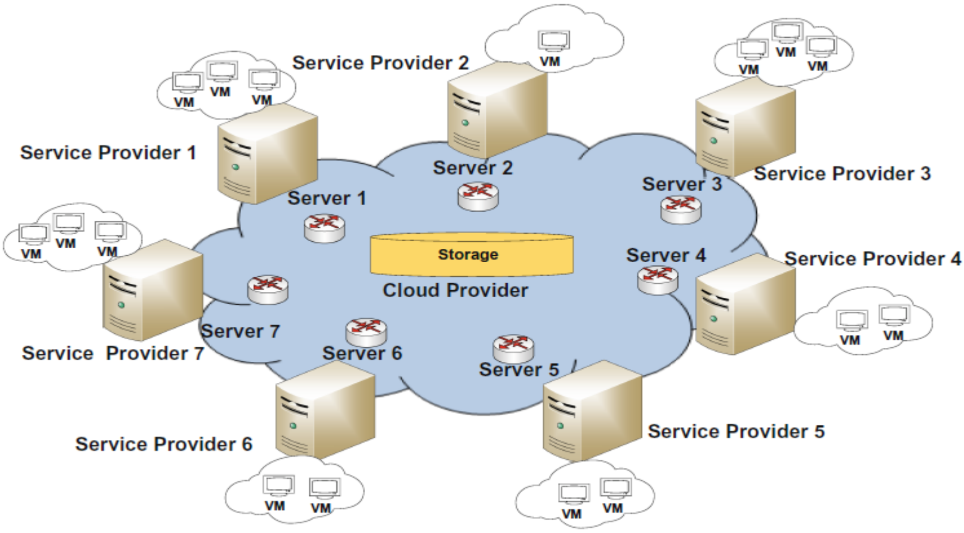

Figure 4: Cloud Computing System (Alarqan et al. 2020).

2.1.5 OwnCloud Platform

Owncloud is open-source software that allows you to build a personal cloud file storage service. It was first designed in 2010. It provides capabilities that are equivalent to Dropbox and other cloud storage providers. The OwnCloud server software is available for free on Linux, and the client software is available for Windows OS and Linux PCs

(Rinkewitz 2022). Table 1 shows the advantage and disadvantages of the owncloud platform.

Table 1: Advantages and disadvantages of the Owncloud platforms.

Advantages	Disadvantages
➢ It consists of readily made resources which can be used for the implementation of this research without wasting time building new resources.	➢ It requires technical skills to configure owncloud with its full functionality.
➢ Navigation to the platform is easy	➢ It has compatibility issues with some applications like PHP 8.0
➢ Applications can be easily deployed in any region around the world.	
➢ There is total ownership of the cloud platform with its security	.
➢ OwnCloud has a free virtual box OVA file which can be used in an isolated environment.	
➢ It contains an application package centre used to build up required resources for the implementation process.	
➢ It does not pose any risk when using it to implement any attack.	

2.1.6 DDoS Threats on Cloud System

Cloud computing services are frequently offered over the HTTP protocol, which provides access while lowering costs, but these services are subject to HTTP DDoS attacks. DDoS is a subset of DoS where a hostile user launches a DDoS strike on purpose to interrupt and degrade the legitimate user's operation and resources. Several negotiating computers as shown in Figure 5 are used in this form of attack to target system resources and services in the cloud, causing a flood of messages, corrupted frames, and connection requests, and a denial of service to the genuine user (Wani et al. 2019).

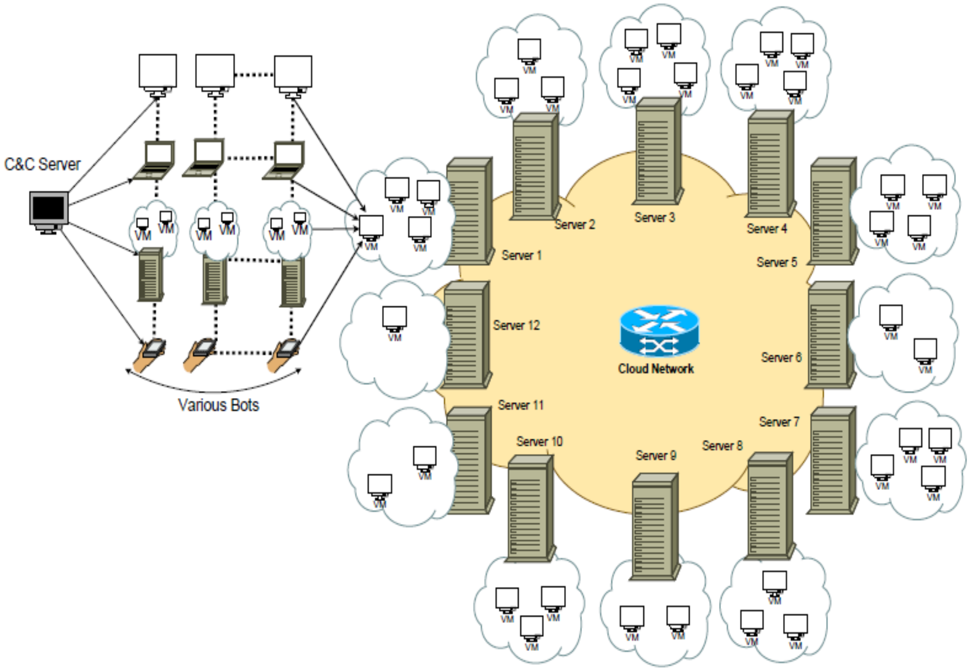

Figure 5: Cloud System DDoS attack Concept (Alarqan et al. 2020).

2.1.7 Methods of DDoS attacks

Different methods of performing DDoS attacks on systems include;

2.1.7.1 Slow request/response attack

The victim's resources are slowly depleted in this kind of attack. Examples of this kind of attack include Slowlories, HTTP fragmentation attacks, Slowpost or RUDY (R.U. Dead Yet) attacks, and Slowreading attacks. In Slowlories, the goal is to crash the victim's computer using just one device (Silva et al. 2020).

2.1.7.2 Malformed packet attack

An attack using a deformed packet that could confuse the target and cause a system crash is a primary concept underlying a malformed packet attack. Examples are ping of death, teardrop, and land attacks (Mahjabin et al. 2017).

2.1.7.3 Protocol exploited attack

Attacks that take advantage of protocol vulnerabilities may employ a network layer protocol like Internet Control Message Protocol (ICMP) or a transport layer protocol like User Datagram Protocol (UDP) (ICMP) like UDP, ICMP, HTTP flood attacks and TCP SYN attack (Mahjabin et al. 2017).

2.1.7.4 Amplification

The fundamental idea behind these kinds of attacks is to produce a massive answer for a tiny request and send those responses to the victim, gradually using up all the network capacity of the victim. Example NTP amplification attack (Lohachab and Karambir 2018).

2.1.8 DDoS Attack Tools

January et al. (2022) suggested that DDoS attacks can be carried out in a variety of tools and ways including remotely or manually. Most attacks are carried out manually since they are directly dependent on the attacker's intention and capability. Botnets are one of the most common manual approaches used in cloud environments. Scripts and devices are being utilized as attack tools to generate more powerful DDoS attacks. However, some technologies are still used to scan devices and launch various forms of DDoS attacks. Due to the complex technology involve Attackers are getting more intelligent and stealthier, carrying out more severe strikes (Zhou et al. 2017). The tools include Low Orbit Ion Cannon, High Orbit Ion Cannon, Slowloris, Tor hammer tool, R U dead yet? (RUDY) and others.

Among all the tools mentioned, this research intends to use the slowloris DDoS attack tool.

2.1.8.1 Slowloris

A technique for creating slow-rate DoS attacks is called Slowloris. The tool exploits a flaw in the HTTP protocol where a valid TCP connection to the target is established. Once the attacker has established a connection, it overwhelms the host by flooding it with incomplete HTTP connections (Fredrick et al.2022). As shown in Figure 6 the attacker continuously sends small quantities of data while keeping the partial HTTP connection active, draining the target's resources. Since no packets are being delivered that are faulty, it is impossible to identify the attack (Shorey et al. 2018). The tool's thread-based nature makes it particularly effective against Apache servers (Papadie and Apostol 2017).

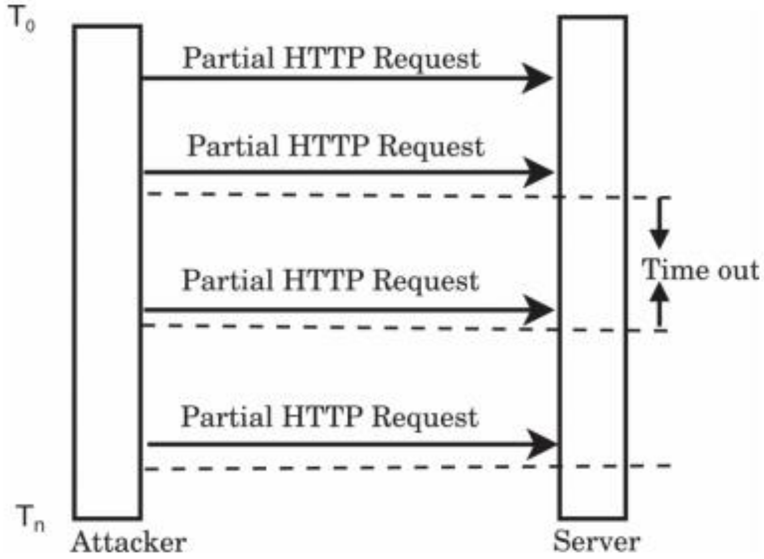

Figure 6: Slowloris DDoS Attack (Fredrick et al.2022).

2.1.8.2 Strengths and Weaknesses of Slowloris Tool

Among all the attack methods, January et al. (2022) suggested that the slow rate DDoS attack method is one of the greatest methods for conducting DDOS attacks. This technique uses the Slowloris tool. The strengths and weaknesses of this tool are pointed out in table 2.

Table 2: Strengths and weaknesses of slowloris DDoS tool

Tools	Strength	Weakness
Slowloris	➢ It is referred to as the most efficient instrument accessible. ➢ The server becomes overloaded with requests due to the limited bandwidth. ➢ Using this tool gives attackers the ability to connect to the victim's server and maintain those connections as long as they are required. ➢ It is lightweight and easier to perform on a system ➢ It uses low bandwidth ➢ A DDoS attack could be performed using a system without compromising any other system	➢ It runs very slow to build more attacks. ➢ It can send requests in chunks as slowly as possible to force the server to wait longer to utilize all the server's resources (Zhou et al. 2017).

2.1.9 Programming tools

Detection of DDoS attacks using ML techniques requires the use of programming methods. This implies using codes or scripts and automated applications to implement ML models that can be used to detect benign and malicious network traffic. Programming tools used for ML techniques are Python language, R language (R Studio), and weka tool.

2.1.10 Machine Learning

One of the effective methods of detecting DDoS attack traffic is Machine learning (Sambangi et al. 2022**).** Machine learning is a kind of artificial intelligence that aims to solve issues using historical or precedent data. Machine learning entails discovering hidden patterns in data (data mining) and then using those patterns to classify or forecast a problem-related event (Dridi 2022). This research used a supervised machine learning technique to build models for the detection of DDoS attacks in the cloud.

2.1.10.1 Supervised Machine Learning

Supervised ML uses specified output attributes. The algorithms try to predict and classify the predefined attribute, their accuracy and misclassifications, as well as other performance measures, based on the counts of the predetermined attribute that are correctly predicted or categorized. It's also worth noting that the learning process ends when the algorithm reaches a satisfactory level of performance (Alloghani et al. 2020). The supervised ML classifiers that will be used in this study for the detection of DDoS are Support vector machines (SVM), Naive Bayes (NB), K-nearest-neighbour (K-NN), and Decision tree (DT), and Random Forest (RF).

In this research, five distinct classification techniques—K nearest neighbour, SVM, naive Bayes, decision tree, and random forest were tested and trained. These algorithms were chosen based on the following standards which include;

I. using a combination of parametric and nonparametric techniques

II. featuring a variety of algorithms from many categories and

III. To employ algorithms that are frequently found in existing work

Parametric and Non-parametric algorithms

A pre-defined property of the data can be used to broadly describe a parameter. An algorithm with parameters has a set number of parameters. A parametric algorithm produces more predictions about the dataset yet is computationally more efficient. If the presumptions are accurate, this would be perfect. But with false assumptions, parametric algorithms perform badly (Kiourkoulis 2020). The parametric algorithm used in the study is SVM and Naive Bayes.

Non-parametric algorithms, on the other hand, are more adaptable, in nonparametric cases, as the model learns, the number of parameters increases. Although this kind of algorithm runs calculations more slowly, it makes many fewer assumptions about the dataset (Kiourkoulis 2020). In this work, k-nearest neighbour, decision trees, and random forests were used as nonparametric algorithms.

2.1.10.2 Models employed in this research

Random Forest

The scalable and user-friendly Random Forest machine learning technique typically yields excellent results. It is popular due to its ease of use and the fact that it can be applied to both classification and regression.

It works by creating a large number of decision trees throughout the training phase, with the output taking the form of the classification of specific trees (Wani et al.2019).

Decision Tree

The term "decision tree" refers to one of the supervised machine learning methods that can be applied to both classification and regression applications. This approach, which divides data into two subsets depending on predetermined criteria known as "decision rules," can be thought of as a binary tree structure in binary classification. The tree is growing top-down, and each node has a characteristic that is chosen to classify training local data. The dividing process is carried out repeatedly until the best classification is reached or all attributes are used (Seraf 2020).

Support Vector Machine

The machines that plot training vectors in feature space and label each vector with its class are called support vector machines (SVM). Like a quadratic optimization problem is this classification issue. They employ a method that can get around the curse of dimensionality. The main characteristic of SVMs is their ability to categorise datasets by identifying the set (collection) of support vectors produced by the set of training inputs and then creating a hyper-plane in a high-dimensional feature space (Rahman 2020).

K Nearest Neighbor

The KNN classification method divides test data observations into groups according to how near their nearest class neighbours they are. The nearest neighbours are found using KNN, which is utilised as a semi-supervised learning method. The unlabeled point is then assigned to the surrounding class K after the distance between various locations on the input vector is calculated (Alduailij et al. 2022).

Naïve Bayes

One of the most well-known probabilistic models, Naive Bayes determines the probabilities for each class, establishes classification, and learns to predict the values of the new class. An example of a classification problem is represented by the vector x= x1…….xn representing n independent variables that are allocated to occurrence probabilities p(Ck/x1……..,xn).

2.1.10.3 Confusion Matrix

A table called a confusion matrix is used to describe how well a classification system performs. The output of a classification algorithm can be seen and summarised in a confusion matrix. Table 3 explains the components of the confusion matrix (Bhandari 2020).

Table 3: explanation of confusion matrix components

Confusion matrix component	Description
True Positive (TP)	The model's prediction and the actual value align perfectly. Also, means that the model anticipated a positive value, and the actual value was positive.
True Negative (TN)	The projected value coincides with the observed value. This means that the model predicted a negative result, and the actual value was negative.
False Positive (FP)	The result was incorrectly predicted which means that the model predicted a positive result, but the true value was negative.
False Negative (FN)	The predicted number was incorrectly forecasted.

	The model projected a negative result, while the actual value was positive.

2.2 Experimental Background

This section describes the experimental context and examines how researchers have used these tools in their works. The project investigation report contains a more complete and in-depth review of this part.

2.2.1 Method

The papers reviewed in this report were obtained from educational repositories like IEEE Xplore digital library, Elsevier search website, google scholar search engine, Google search and others. The search was performed using some search strings like “DDoS attack detection on the cloud, DDoS attack detection using machine learning, DDoS attack detection on the network”. On google search, after each string journals were added to get journal papers from different search databases. Out of 60 papers reviewed, 22 were used in the literature review.

Detection of DDoS attacks using especially machine learning is becoming incredibly valuable. The researcher explores how other scholars use machine learning and a few other techniques to resolve the problem of DDoS attack detection. The main insights from this section are to understand how different algorithms are used to solve the problem, which algorithms are used, and the results obtained from applying such techniques/methodologies.

For instance, Wani et al. (2019) conducted a DDoS detection on the cloud by exploring their dataset from the OwnCloud platform and trained it on three machine learning algorithms: RF, NB, and SVM. Similarly, Shabaan et al. (2019) performed their study on a large sample benchmark dataset using five classification algorithms; Neural Networks (NN), Convolutional Neural Network (CNN), (SVM), DT, and K-NN to detect DDoS attacks with a sample dataset. More so, Sarraf (2020) used a dataset with 200,000 DDoS and benign classifications samples consisting of 83 dataset features. The dataset was trained with ML models namely NB, SVM, RF, and DT for the detection of DDoS attacks on network systems. Similarly, Abdulrahman and Ibrahem (2019) proposed an evaluation that used an open-source dataset to perform DDoS attack detection based on classification algorithms in machine learning. Moreover, Hosseini and Azizi (2019) proposed a hybrid methodology to detect DDoS attacks using the machine learning technique NB, RF, DT and Multi-Linear Regression (MLP). Another study by Pei et al. (2019) proposed a novel DDOS attack detection approach based on machine learning using random forest and SVM algorithms. Feature extraction and format conversion were done on the three protocols namely TCP, UDP, and ICMP attack packets of the DDOS data. Consequently, Tuan et al. (2020) in their study used numerous classifiers SVM, ANN, NB, DT, and UML (K- and X-means) to detect Botnet DDOS attacks with open datasets. Furthermore, Rehman et al. (2021) proposed a method for detecting and identifying DDoS attacks on three protocols (UDP, DNS, and LDAP). ML models (GRU), (RNN), Naïve Bayes (NB), and Sequential Minimal Optimization (SMO) were used to train their dataset. Lima Filho et al. (2019) presented a study on The Smart Detection system, an online approach to DoS/DDoS attack detection using the RF tree method, a software that classifies network traffic which is based on samples taken directly from network devices through the sFlow protocol. Their proposed system was tested using three intrusion detection benchmark datasets with a sample rate (SR) of 20% of network traffic. Comparatively, Kim (2019) used three separate network traffic sets, including DDoS attacks, which are subjected to supervised learning algorithms. TensorFlow was used to extract features using BNN and LSTM RNN. Similarly, Alzahrani and Alzahrani (2021) performed a security analysis of DDoS attacks in-Network traffic using a machine learning model: SVM, K-NN, DT, NB, RF, and LR. Aytaç et al. (2020) presented a different view by examining the success rate of an intrusion detection system using a variety of machine learning techniques. Several ML algorithms, including the ANN, SVM, Gaussian NB, Multinomial NB, Bernoulli NB, Logistic Regression, KNN, DT, and RF methods, to train a dataset. Also, (Saghezchi et al. 2022) used machine learning to identify DDoS attacks on 5g networks in Industry 4.0 CPPSs. To detect anomalies in network traffic flows, they exported network traffic traces (PCAP files) from a large-scale semiconductor fabrication factory in the real world and used 11 different semi-supervised, unsupervised, and supervised ML methods. A study conducted by Alghoson and Abbass (2021) proposed detecting distributed denial of service attacks using machine learning. They used an open dataset to examine the efficiency of DDoS prediction techniques by using accurate metrics. Kushwah and Ranga (2021) proposed a hybrid machine learning-based approach by combining Extreme Learning Machine (ELM) algorithm and the black-hole optimization technique to detect DDoS attacks. To evaluate the performance of the proposed hybrid machine learning system, the scholars carried out experiments using different open datasets. In addition, Alzahrani and Alzahrani (2021) performed their study on security analysis on K-NN, SVM, NB, decision DT, RF, and logistic regression are six different types of ML methods that were used in this study. In contrast, Elsayed et al. (2020) proposed SDN systems, which is an Intrusion Detection System against DDoS attacks (DDoSNet). They used Deep Learning (DL) technique, which combines autoencoders with a recurrent neural network (RNN) on their dataset used to test the created system. A study conducted by Bolodurina et al. (2020) used a dataset, which contains significant information regarding reflection-based and exploitation-based attacks. The Researchers used the dataset to evaluate the influence of the data balancing technique in the network traffic classification problem on several forms of DDoS attacks. Conversely, Cil et al. (2021) used a deep learning technology called the Deep Neural Network (DNN) to detect DDoS attacks on a sample of packets acquired from network traffic because it includes feature extraction and classification algorithms. The scholars used a dataset, which covers numerous DDoS attack types that were produced in 2019.

On the other hand, other studies were conducted using machine learning and other methods. For instance, Alanazi et al. (2019) conducted reviews for detecting DDoS attacks in the cloud on signature-based, Anomaly-based, and hybrid-based detection. Similarly, Mittal et al. (2022) conducted a systematic review of different research papers on deep learning approaches for detecting DDoS attacks. They concluded that using large sample data records with multiple ML algorithms helps to achieve high accuracy.

The use of tools to conduct DDoS attacks in the cloud, for instance, Fredrik et al. (2022) proposed a service that cloud providers could use to improve defences against slow-rate DDoS attacks. Similarly, Tripathi and Hubballi (2022) conducted an organised and thorough analysis of the current application layer DDoS attacks using the slowloris tool and a countermeasures mechanism. They categorise existing attacks and defences into various groups, explain how they function and evaluate them using relevant criteria like proactive, Implementation Complexity, scalability, and holistic defence.

2.2.2 Research Gaps

The research gaps from the existing work reviewed in this study were explained in-depth in the project investigation report (CMM 512). The summary is displayed in Figure 7.

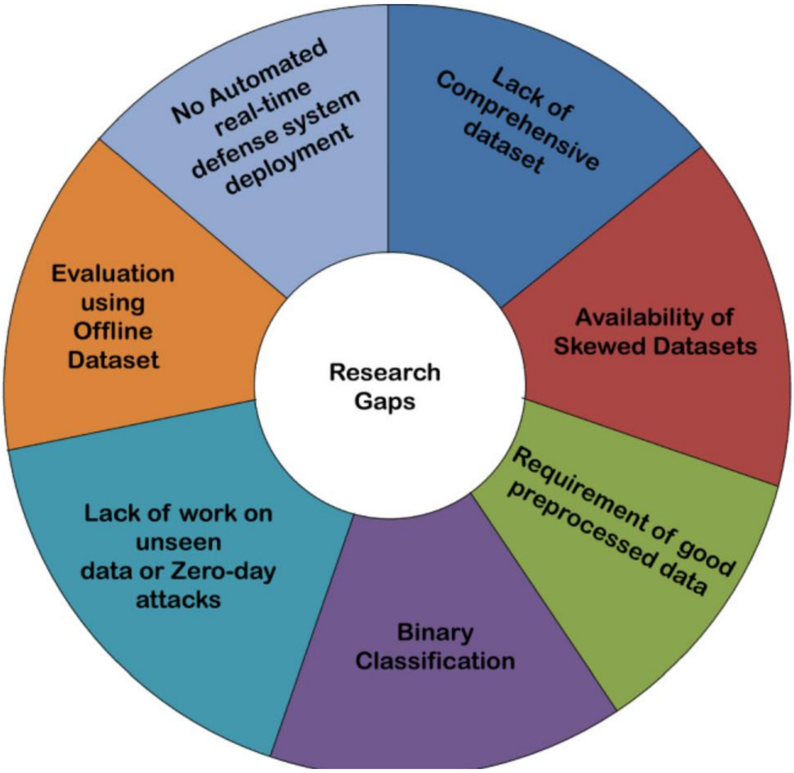
Research Gaps
Lack of Comprehensive dataset
Availability of Skewed Datasets
Requirement of good preprocessed data
Binary Classification
Lack of work on unseen data or Zero-day attacks
Evaluation using Offline Dataset
No Automated real-time defense system deployment

Figure 7: Research gaps from existing work.

2.2.3 Investigation Report Summary

Before proceeding to the actual design and implementation of this study, let's review and list all of the investigation report's findings as its phase in this project has been summarised in Sections 1.0 and 2.0;

- Recently, the trend of software applications and data stored in the Cloud Computing environment has become more challenging because of numerous threats, such as DDoS attacks. Although the detail has been explained in-depth in the project investigation (CMM512), Infiltration is one of the approaches that has been used to monitor and detect threats and attacks against a cloud or virtualized server (See Chapter 1).

- Owncloud was chosen because it has a VirtualBox which can be implemented in an isolated environment or closed platform to avoid disruption of other resources and systems during the attack phase. Using Owncloud VirtualBox does not pose any risk to other systems (Section 2.1.5).

- Slowloris DDoS attack tool was selected because it is easily accessible to use. It is lightweight and easier to perform on a system. It uses low bandwidth. A DDoS attack could be performed using a system without compromising any other system (Section 2.1.8.1).

- Support vector machines (SVM), Naive Bayes (NB), K-nearest-neighbour (K-NN), Decision tree (DT), and Random Forest (RF) ML models were selected because they perform very in training intrusion detection datasets. Also, they are usually used by other researchers (Section 2.1.10.2).

- Most literature on the emergence of a complete definition of a cyber-attack has proposed and applied different machine learning classifiers with model optimization and good results were obtained as explained in (Section 2.2) Despite all the methods of the application of using machine learning to improve the performance of the attack in the Cloud Computing environment, there is still ongoing research in this field.

- The implementation of the experiment will be performed using python 3.9 programming language with a specialized ML environment anaconda- Jupyter notebook. Many researchers used supervised ML Models for DDoS dataset training.