**DETECTING AND MITIGATING APPLICATION LAYER DISTRIBUTED DENIAL OF SERVICE (DDoS) ATTACK USING APACHE HADOOP BIG DATA TECHNOLOGY.**

**CHAPTER ONE**

**INTRODUCTION**

**1.1 Background of the Study**

In todays’ world, attacks against large enterprises are growing at a very fast rate. One such attack is the Distributed Denial of Service (DDoS) attack whereby a victim’s network is loaded with large volume of infected attacking packets which are originated from large number of zombies or botnets. In the modern world of computer today, to maintain information over a network is quite a difficult task as some interruption on a network may occur on the local system (local attack) or network based systems (network attack). The attacker’s aim is to overload the victim’s network with infected packets and render it incapable of carrying out genuine tasks on the network (Sumathi and Koppula, 2015). In recent years, distributed denial-of-service (DDoS) attacks have brought about serious financial losses to many industries and various governments worldwide as reported in the information security reports (Anstee et al., 2017). These records matches with the increasing number of devices that are connected to the Internet, most especially driven by the familiarization of pervasive computing, materialized through the Internet of things (IoT) paradigm and distinguished by the concept of connecting anything, anywhere and anytime (Al-Fuqaha et al., 2015). In the majority of Internet schemes, devices relate with applications that run extrinsically on the network, which allows malicious agents to have control of devices. Hence, it makes it possible to have interruption of services or the use of such devices as a point of launching attacks on different domains such as DDoS attack (Zargar and Tipper, 2013). This has been combined for several reasons, such as; simplicity and facility of execution, that doesn’t require broad technological knowledge from the attacker’s side and a variety of platforms and applications for promoted attack set up. Most of these attacks have succeeded in causing damage to essential Internet services such as Domain Name System (DNS), thereby affecting millions of users the world over (Marzano et al., 2018), and commercial tenets such as the GitHub (Kottler, 2018) thereby causing serious financial losses to the organizations that rely on those services. One of the most dangerous malicious traffic on the Internet is the DDoS volumetric attack, which makes up for more than 65% of all such attacks (Cao et al; 2018). According to [Om Thoke](https://www.lifewire.com/om-thoke-3473482) (2019), there is a totally different kind of DDoS attack called the Application-Layer DDoS attack, which is also known as ‘Layer 7’ DDoS attack. These types of attacks are not easy to detect and are even harder to protect against. In fact, you might even fail to notice it until the time the website goes down which can also affect many back-end systems.

DDoS attacks have been increasing in volume, complexity and damage rate. In recent times, financial institutions and business organizations have experienced a very high share of DDoS attacks (Kitten, 2015). Attackers use botnets to launch these attacks. According to Dileep et al. (2018), a botnet is a network of machines or bots that have been compromised and are known as zombies or slaves. These botnets have command and control infrastructures that can be used to carry out malicious activities. Bots can automatically scan an entire network and infiltrate the network using known vulnerabilities and weak passwords on other machines and once any machine is compromised, a small program is installed for future activation by the attacker. The attacker can instruct the bots on the network to carryout actions at a given time such as sending large number of requests to a target server so that it would not be able to serve requests by legitimate users of the network which results in a DDoS attack (Dileep et al., 2018).

DDoS attack generally consists of four elements which are the attacker, controllers, zombies and the victim’s machine. The attacker is the person who launches the attack, the controllers are machines on the network that have been compromised by the attacker to run malicious code that will be used to control large set of zombies. The zombies are the compromised systems on the hijacked network that run the malware to generate large set of attack packets and send them to targeted victims, the victim’s machine is the machine targeted by the attacker (Dileep et al., 2018).

According to Patil et al., (2019) In the past few years, the traffic volume of legitimate traffic and attack traffic has increased tremendously up to Terabytes per second (Tbps) and as such, the processing of such a huge volume of traffic has made it very difficult to detect high rate attacks on time using conventional DDoS defense architectures. Patil et al., (2019) stated that presently, the majority of the DDoS defense systems are deployed predominantly at the victim-end domain but the victim-end defense systems themselves are equally vulnerable to Hit-and-run DDoS (HR-DDoS) attacks as the large volume of attack traffic is generated by such type of attacks.

Hoque et al., (2014) stated that in recent times, thousands of organizations such as domain name servers, payment gateways, banks, search engines, educational institutes, social websites such as What Sapp, Twitter, Facebook, stock trades, commercial cloud based service providers servers such as Amazon, Google, weather forecasting and so on are exponentially growing to provide Internet-based services and applications to various end users which has led to the increase in network traffic over the Internet. This increasing usage of Internet-based applications has further brought about a rise in Internet misuse. Out of the different types of malware present nowadays on the Internet, DDoS attacks pose a serious menace to the Internet and its architecture (Patil et al., 2019).

DDoS attacks have become increasingly common on campus network as such, higher education technology leaders should put into consideration investments in solutions that can safeguard their networks from these delibating attacks. There has been an increasingly rise in the number of colleges and universities that have experienced these attacks in recent times due to the fact that students are capable of executing these attacks on their own institutions either because of being upset with their grades or because they want a postponement on an examination they feel they are not ready for (Dennis, 2019). In today’s modern and global world that is Information Technology driven, the necessity and increase values provided by network infrastructures have shown its importance in government organizations, business enterprises and educational institutions. This has contributed greatly to the achievement of some important goals such as increased productivity, partnership, efficiency and acquiring of knowledge in frequent researches for educational purposes. Therefore, to increase the use of Information Technology in higher institutions of learning requires a robust technical infrastructure to provide for a secured and reliable network as a university campus network is a great necessity for knowledge sharing, easy communication and aids in collaborative research which are the essential ingredient to building a strong knowledge culture and efficiently support academic mission (Dominic et al., 2015).

Hameed and Ali (2018) stated that the exponential increase in the volume of internet traffic and the sophistication of DDoS attacks have presented serious challenges on analyzing of DDoS attacks in a scalable and accurate manner. The internet traffic doubles yearly and due to that, the monitoring of large amount of traffic in real-time anomaly detection with traditional intrusion detection system (IDS) has become a serious issue. This is due to the fact that existing solutions requires lots of resources and do not add any value to limited resource environment. Therefore, there is need to proffer new solutions based on big data technologies that can be able to detect DDoS attacks efficiently in near real time (Hameed and Ali, 2018).

**1.2 Statement of the Research Problem**

As stated by Sonal (2016), security is always an important issue especially in the case of computer network which is used to transfer personal or confidential information, ecommerce and media sharing. Data in computer networks are growing rapidly and the analysis of these large amounts of data to discover anomaly behaviour has to be done within a reasonable amount of time. Recently, threat of previously unknown cyber-attacks is increasing because existing security systems are not able to detect these attacks (Sonal, 2016). The goal of recent hacking attacks has changed from leaking information and destruction of services to attacking large-scale systems such as critical infrastructures and state agencies (Sonal, 2016). The analysis of malicious network packets and processing of network logs for the detection of threats has been a very hard task for many years now as attackers are continuously changing their methods and modes in the launching of DDoS attacks. Big data tools and technology have attracted network security community because of the simplicity, scalability and fault tolerance features it possesses (Tom, 2015). The volume of internet traffic has rapidly grown in recent times and current DDoS detection technologies have met new challenges that should efficiently deal with a large amount of traffic within the affordable response time (Yeonhee and Youngseok, 2014). Corero Network Security (2020) reported that during a DDoS attack, it is extremely important to have granular security event data so that in the event of an anomalous attack, network security managers can quickly adapt their DDoS mitigation efforts. DDoS hackers have become sophisticated in their attack strategies. That is, they now typically use multi-vector attacks, and they automate their attacks, changing vectors on the fly. Most often, volumetric attacks are used as a smoke screen to hide low and slow application layer attacks. Furthermore, cybercriminals sometimes leverage the power of one botnet to launch [pulse-wave, also known as burst attacks](https://www.corero.com/blog/875-bursts-waves-and-ddos-what-you-need-to-know.html?__hstc=753710.4d8169f700ea94045058d1c5fbbce0f8.1588600058566.1588600058566.1588600058566.1&__hssc=753710.1.1588600058566&__hsfp=733456091) that alternates between two or more targets. Corero Network Security (2020) stated that in these situations, visibility into the real-time attack definitely helps to effectively counter an attack appropriately with custom mitigations. Granular detection into traffic anomalies can provide guidance into possible configuration changes. Zuo (2020) explained that source attacks that target specific applications, such as Hypertext Transfer Protocol (HTTP) flood attacks against e-commerce websites and web games, require a TCP connection between the zombie host and servers targeted for attack and as such, to avoid detection, hackers usually reduce the attack traffic rate so that the attack footprint resembles that of a legitimate request. Application layer DDoS attacks are more difficult to detect and counter because they emulate the syntax and traffic characteristics of legitimate clients and access requests. DDoS attacks against the application layer involve carefully selected targets that threaten service availability through the use of slow, prolonged attacks, and dispersed attack sources. Although server addresses may remain reachable, ultimately these services will become unavailable (Zuo, 2020). To defend against the application layer DDoS attacks in a computer network, which are most times not been detected with existing detection algorithms; this study therefore seeks to propose a real-time framework based on big data technology using Apache Hadoop to detect DDoS attacks at the application layer of a computer network. Big data analysis technologies are able to extract information from variety of sources in order to detect DDoS attacks in a reasonable response time.

**1.3 Research Questions**

This research study therefore proposes the use of Apache Hadoop big data technology framework in dealing with the stated problems and to proffer solutions to these stated research questions:

1. Considering the fast rate of DDoS attack generation in computer networks, will the proposed big data technology detect and mitigate the application layer DDoS attacks in a reasonable response time?
2. With the difficulty in detecting application layer attacks such as the HTTPS GET attacks as they appear similar to legitimate requests from a legitimate user of a network, will the proposed big data technology give an edge over these type of attacks?
3. Will the proposed Hadoop framework solve the scalability issue of detecting multiple attacks from a huge volume of traffic by parallel data processing?

**1.4 Aim and Objectives of the Study**

This aim of this research work is to propose a real-time big data technology framework that is based on Apache Hadoop-based DDoS Detection and Mitigation (AHADM) framework to get a more efficient result in the detection and mitigation of application layer DDoS attacks by separating attackers from real users on a network in real time at a minimum possible response time. This would be achieved with the following objectives:

1. The generation of large amount of legitimate traffic and attack traffic on the network.
2. Implementation of a counter-based algorithm with set threshold within the Map Reduce paradigm of Hadoop to detect and mitigate the network against the DDoS attacks.
3. The deployment of a test bed to carry out experimentation for the evaluation performance of the proposed framework.

**1.5 Scope of the Study**

This research will be focused on the development of an Apache Hadoop-based DDoS Detection and Mitigation (AHADM) framework that will be able to detect and mitigate the application layer DDoS attacks on a computer network. The framework will be made up of four phases which will consist of the following:

1. The capturing of network traffic and generation of logs.
2. The transfer of the generated logs to the detection server.
3. The detection and mitigation of DDoS attacks by the detection server.
4. The result notification phase.

**1.6 Limitations of the Study**

The AHADM framework will be developed to capture logs from the network traffic and transfer the generated logs to the detection server by the capturing server, the detection server will automatically start the process of detecting the DDoS attacks on the logs that will be received based on the set threshold in the map reduce job function. This proposed framework will focus only on the detection of application layer DDoS attacks.

* 1. **Significance of the Study**

The proposed Hadoop big data technology framework is dependent on distributed architecture to detect and mitigate application layer DDoS attacks using low cost commodity hardware which can help government agencies, organizations and educational institutions to secure their network in-house at a minimum possible cost and response time. The results that will be obtained from this research work will determine the efficiency rate of the big data technology framework used in the detection and mitigation of application layer DDoS attacks on a network. It can help to ease the burden of high expenses on government agencies, organizations, educational institutions, small and medium enterprises and financial institutions to set up an in-house low cost defenses and most importantly, save the huge cost incurred when deploying a third party DDoS mitigation service provider.

**1.8 Definition of Terms**

Algorithm ‒ Is a process or a set of rules that is followed to solve a given problem. It is mostly used in processing data and other computer related operations.

Apache Hadoop – Is a collection of open-source utility software that helps to facilitate the use of a network of many computers to solve problems that involves large amounts of data and computation. It provides a software framework for distributed storage and processing of big data with the use of Map Reduce programming model.

Application Layer DDoS Attack ‒ This refers to a type of malicious attack behaviour that is designed to target the top layer in the OSI model where common internet request such as HTTP GET and HTTP POST occurs. The application layer attacks are particularly effective in the consumption of server resources and network resources.

Big Data ‒ Is a term that is used to describe large volume of data that is both structured and unstructured. It is field that deals with the analysis of large data sets that are too large or complex to be dealt with by traditional data processing application software.

Big Data Technology ‒ This is defined as a software-utility that is designed to analyze, process and extract information from an extremely complex and large data sets which the traditional data processing software cannot deal with.

Detection ‒ Is the action or process of identifying the presence of something that is concealed.

Denial of Service (DoS) Attack ‒ This is cyber-attack in which the perpetrator seeks to make a machine or network resources unavailable to the intended users by temporarily or indefinitely disrupting the services of a host connected to the internet.

Distributed Denial of Service (DDoS) Attack ‒ Is a DoS cyber-attack whereby the incoming traffic flooding the victim’s network originates from many different sources.

Domain Name Server (DNS) ‒ Is the hierarchical and decentralized naming system for computers, services or other resources that are connected to the internet or a private network. It associates various information with domain names assigned to each of the participating entities in the network.

Hypertext Transfer Protocol (HTTP) ‒ Is an application protocol for distributed, collaborative, hypermedia information systems. It is the foundation of data communication for the World Wide Web (WWW) where hypertext documents include hyperlinks to other resources that a user can access easily.

Information Technology ‒ Is the study or the use of systems especially computers and telecommunications for the storing, retrieving and sending of information.

Internet ‒ Is a massive network of networks. It is the global system of interconnected computer networks that uses the internet protocol suite (TCP/IP) to link devices worldwide.

Map Reduce ‒ Is a processing technique and a program model for distributed computing that is based on java programming language. The map reduce algorithm consist of two important tasks; the map and reduce function.

Mitigation ‒ Is the action of reducing the seriousness or severity of loss from the occurrence of an undesirable event.

Open Systems Interconnection Model (OSI Model) ‒ It is a model that was created to help standardize communication between computer systems. It divides communications into seven layers to include multiple hardware standards, protocols or other types of services.

**1.9 Breakdown of the Chapters**

This research work is being arranged into five chapters. Chapter one presents the introduction of the study, the background to the study, statement of the research problem, the research questions, the aim and objectives of the study, the scope and limitations of the study, the significance of the study and definition of key terms used in the study. Chapter two presents the historical overview of the study, the review of related literature and the theoretical frame work of the study. Chapter three contains the research methodology and design, the data analysis technique used and the instruments for data collection used. Chapter four contains the presentation and analysis of data and critical discussion of findings from the study. Chapter five presents the summary of the study, the conclusion of the study, recommendations from the study done and suggestion for further study.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.0 INTRODUCTION**

This chapter is focused on the detailed background of different network attacks, the Distributed Denial of Service (DDoS) attacks and the different big data technologies that are used for DDoS detection and mitigation. Related works on the use of big data technologies in the detection and prevention of DDoS attacks were also reviewed.

**2.1 Network Attacks**

Cynet (2020) define a network attack as an attempt to obtain an unauthorized access to an organization’s network with the sole objective of stealing data or carrying out other malicious activities on the network. There are two main types of network attacks which are the passive and active attack. The passive attack is a type of network attack whereby attackers gain access to a network and are able to monitor or lay hold of sensitive and vital information but without making any changes to the data whereas in an active attack, attackers after gaining unauthorized access to the data, also modifies the data either by deleting, encrypting or harming the data (Cynet, 2020).

Cynet (2020) stated that in a network attack, attackers are more focused on penetrating the corporate network perimeter and having access to internal systems and once inside, attackers combine other types of attacks such as compromising endpoints, spreading malwares and exploiting vulnerable systems within the network. The following are different types of network attacks that exist as stated by EC-council blog (2018).

1. **Virus:** A virus is not self-executable but requires the user’s interaction to be able to infect a computer and spread on the network. A good example is an email that has malicious link or malicious attachment that when opened, the malicious code becomes activated and circumvents the entire system’s security controls making it inoperable.
2. **Malware:** Malware attack is stated as one of the most severe cyber-attacks that is mainly designed to destroy a computer system or gain unauthorized access over a targeted computer system. Most malwares are self-replicating, that is; when it infects a particular system, it gains entry over the internet and from there infects all other system connected to the internet in the network. If an external endpoint device is also connected, it will also get infected. It works exceptionally faster than other types of malicious content.
3. **Worm:** Worms can gain access to a device without the help of the user. When a user runs a vulnerable network application, an attacker on the same internet connection can send malware to the application and the application may accept the malware from the internet and execute it hence creating a worm.
4. **Phishing:** Phishing is one of the most common type of network attack. It involves the sending of email purposefully as from a known resources or bankers and creating a sense of urgency to excite a user to act on it. The email may contain malicious link or attachment and may ask to share confidential information of the user.
5. **Botnet:** Botnets are network of private computers which are victims of malicious software. The attacker controls the computers on the network without the knowledge of the owner of the computer. Each of the computers on the network is considered as zombies as they serve the purpose of spreading and infecting large number of devices as guided by the attacker.
6. **Denial of Service (DoS) Attack:** A Denial of Service Attack is a serious attack that can destroy fully or partially, a victim’s network or an entire IT infrastructure to make it unavailable to the legitimate users of the network. Dos attacks can be grouped into the following three parts-

* **Connection Flooding:** In this type of DoS attack, the attacker bogs down the host by establishing a large number of TCP connections at the targeted host. These fake connections been established will then block the network making it unavailable to the legitimate users.
* **Vulnerability Attack:** In this type of DoS attack, the attacker sends a few well-crafted messages to the vulnerable operating system or applications running on a targeted host stops the service or make it worse to the extent of the host getting crashed.
* **Bandwidth Flooding:** In this type of DoS attack, the attacker prevents legitimate packets from getting to the server by sending a deluge of packets. The packets sent are large in number so much that the target’s link gets blocked for legitimate users to access.

1. **Distributed Denial of Service (DDoS) Attack:** the DDoS attack is a complex form of the DoS attack which is much harder to detect and defend when compared to the DoS attack. In this type of attack, the attacker uses multiple compromised systems to target a single DoS targeted system. The DDoS attack also makes use of botnets.
2. **Man-in-the-middle Attack:** A man-in-the-middle attack is someone who stands in between a user and the other person. By being in the middle, the attacker monitors, capture and controls the communication of the user effectively. For example, when information is sent from the lower layer, the computer in the layer may not be able to determine the recipient with which information is been exchanged.
3. **Packet Sniffer:** In this type of attack, a packet receiver placed in the territory of the wireless transmitter records a copy of every packet that is transmitted. These packets can contain confidential information, sensitive and crucial data, trade secrets and so on which when flown over a packet receiver will get through it. The packet receiver will then work as a packet sniffer sniffing all the transmitted packets entering the range. Cryptography is the best defense against this type of attack.
4. **Domain Name System (DNS) Spoofing:** DNS spoofing is all about compromising a computer by corrupting the domain name system data and then introducing it in the resolver’s cache. This causes the name server to return an incorrect IP address.
5. **Internet Protocol (IP) Spoofing:** IP spoofing is the process of injecting packets into the internet using a false source address. It is one of the ways to pose as another user. An end point authentication that ensures the certainty of a message originating from the place is been determined would help in defending from IP spoofing.
6. **Compromised Key:** In this type of network attack an attacker gains unauthorized access to a secured communication using a compromised key. A key refers to a secret number or code that is required to interprete secured information without any intimation to the sender receiver. When the key is obtained by the attacker, it is referred to as a compromised key which serves as a tool for retrieving information (EC-council, 2018).

**2.2 Distributed Denial of Service (DDoS) Attack**

Cloud flare (2020) defined a distributed denial of service (DDoS) attack as a malicious attempt to disrupt normal traffic of a targeted server, service or network by bombarding the target or its surrounding infrastructure with a flood of Internet traffic. Cloud flare (2020) stated that DDoS attacks achieve its effectiveness by making use of multiple compromised computer systems as sources of attack traffic. Exploited machines can include computers and other networked resources such as Internet of Things ([IoT) devices](https://www.cloudflare.com/learning/ddos/glossary/internet-of-things-iot/). From a high level, a DDoS attack is like a traffic jam clogging up with highway, preventing regular traffic from arriving at its desired destination (Cloud flare, 2020).

According to the emerging threat report by Norton (2020), they described a distributed denial of service (DDoS) attack as one of the most powerful weapons on the internet. Norton (2020) stated that when you hear about a website being “brought down by hackers” it generally means that the website has become a victim of a DDoS attack. That is, the hackers have attempted to make a website or computer unavailable by flooding or crashing the website with too much traffic.

Panda Security (2019) reported that Distributed Denial of Service (DDoS) attacks have been on the rise in recent times. They are usually implemented using botnets (a legion of zombie computers doing the bidding of a hacker or group of hackers), these attacks can result to some serious damage to a company’s operability. To combat this problem, many companies and organizations are beginning to rely on the power of Big Data Analytic. Generally speaking, DDoS attack vectors can be roughly classified into three categories, Application Layer Attacks, Protocol-Based Attacks, and Volumetric Attacks. When preparing to launch a targeted attack, the experienced adversary will take into account the defensive measures the target has, the software being used by the target, and the resources available to the adversary to execute the attack, such as botnet capacity. Taking this information into account helps the adversary decide which vector will have the most impact on the target. In more generic attacks where the adversary’s agenda may be little more than to gain attention, you may see this logic worked backward, first taking into account their resources before selecting a soft target susceptible to the attack (Panda Security, 2019).

**2.2.1 Motivation behind DDoS Attacks**

According to the report on DDoS Attacks by Imperva (2020), DDoS attacks are quickly becoming the most prevalent type of cyber threat, growing rapidly in the past years in both number and volume according to recent market research. The trend is towards shorter attack duration, but bigger packet-per-second attack volume.

Attackers are primarily motivated by the following reasons

1. **Ideology –** So called “hacktivists” use DDoS attacks as a means of targeting websites they disagree with ideologically.
2. **Business feuds –** Businesses can use DDoS attacks to strategically take down a competitor’s websites. For example, to keep them from participating in a significant event, such as Cyber Monday.
3. **Boredom –** Cyber vandals, popularly known as “script-kiddies” use prewritten scripts to launch DDoS attacks. The perpetrators of these attacks are typically bored and are usually hackers looking for an adrenaline rush.
4. **Extortion –** Perpetrators use DDoS attacks or the threat of DDoS attacks as a means of extorting money from their targets.
5. **Cyber warfare –** Government authorized DDoS attacks can be used to both cripple opposition websites and an enemy country’s infrastructure.

**2.2.2 How does a DDoS attack work?**

According to the report by the emerging threat Norton (2020), the theory behind a DDoS attack is simple, although attacks can range in their level of sophistication, the basic idea is that a DDoS attack is a cyber-attack on a server, service, website, or network being flooded with Internet traffic. If the traffic overwhelms the target, its server, service, website, or network, it is rendered inoperable.

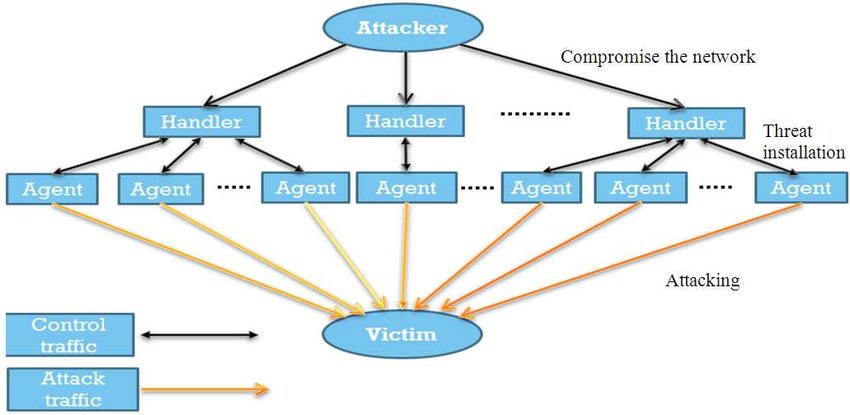
Network connections on the Internet consist of different layers of the Open Systems Interconnection (OSI) model. Different types of DDoS attacks focus on particular layers which are:

* **Layer 3:** The Network layer attacks which are known as Smurf Attacks, ICMP Floods, and IP/ICMP Fragmentation.
* **Layer 4:** The Transport layer attacks which include SYN Floods, UDP Floods, and TCP Connection Exhaustion.
* **Layer 7:** The Application layer attacks which are mainly HTTP-encrypted attacks.

Norton report (2020) stated that the primary way a DDoS attack is being accomplished is through a network of remotely controlled, hacked computers or bots. These are often referred to as “zombie computers.” They form what is known as a botnet or network of bots. These are used to flood targeted websites, servers, and networks with more data than they can handle. The botnets can send more connection requests than a server can handle or send overwhelming amounts of data that exceed the bandwidth capabilities of the targeted victim. Botnets can range from thousands to millions of computers controlled by cybercriminals. Cybercriminals use botnets for a variety of purposes, including sending spam and forms of malware such as ransom ware. Your computer may be a part of a botnet, without you knowing it.

Increasingly, the millions of devices that constitute the ever-expanding Internet of Things (IoT) are being hacked and used to become part of the botnets used to deliver DDoS attacks. The security of devices that make up the Internet of Things is generally not as advanced as the security software found in computers and laptops. That can leave the devices vulnerable for cybercriminals to exploit in creating more expansive botnets (Norton report, 2020).

In the report by Cloud flare (2020), they stated that a DDoS attack requires an attacker to gain the control of a network or online machines in order to carry out an attack. Computers and other machines such as IoT devices are infected with [malware](https://www.cloudflare.com/learning/ddos/glossary/malware/), turning each one into a [bot](https://www.cloudflare.com/learning/bots/what-is-a-bot/) (or zombie). The attacker then has remote control over the group of bots that has been acquired, which is known as a [botnet](https://www.cloudflare.com/learning/ddos/what-is-a-ddos-botnet/). Once a botnet has been established, the attacker is able to direct the machines by sending updated instructions to each bot through a method of remote control. When the [IP address](https://www.cloudflare.com/learning/ddos/glossary/tcp-ip/) of a victim is targeted by the botnet, each bot will respond by sending requests to the target, potentially causing the targeted server or network to overflow capacity hence, resulting in a [denial of service](https://www.cloudflare.com/learning/ddos/glossary/denial-of-service/) to normal traffic. Because each bot is a legitimate Internet device, separating the attack traffic from normal traffic could be difficult (Cloud flare, 2020). Figure 1 presents the architecture of a DDoS attack.



**Figure 1: The Architecture of a Distributed Denial of Service (DDoS) Attack (Wesam and Mehdi, 2014).**

**2.2.3 Common Distributed Denial of Service (DDoS) Attacks Types**

Some of the most commonly used DDoS attack types as stated by Imperva (2020) include:

* **User Datagram Protocol (**[**UDP) Flood**](https://www.imperva.com/learn/application-security/udp-flood/)

A UDP flood, by definition, is any DDoS attack that floods a target with User Datagram Protocol (UDP) packets. The goal of the attack is to flood random ports on a remote host. This causes the host to repeatedly check for the application listening at that port, and (when no application is found) reply with an ICMP ‘Destination Unreachable’ packet. This process saps host resources, which can ultimately lead to inaccessibility.

* **Internet Control Message Protocol (**[**ICMP) Ping Flood**](https://www.imperva.com/learn/application-security/ping-icmp-flood/)

Similar in principle to the UDP flood attack, an ICMP flood overwhelms the target resource with ICMP Echo Request (ping) packets, generally sending packets as fast as possible without waiting for replies. This type of attack can consume both outgoing and incoming bandwidth, since the victim’s servers will often attempt to respond with ICMP Echo Reply packets, resulting a significant overall system slowdown.

* [**SYN Flood**](https://www.imperva.com/learn/application-security/syn-flood/)

A SYN flood DDoS attack exploits a known weakness in the Transmission Control Protocol (TCP) connection sequence (the “three-way handshake”), wherein a SYN request to initiate a TCP connection with a host must be answered by a SYN-ACK response from that host, and then confirmed by an ACK response from the requester. In a SYN flood scenario, the requester sends multiple SYN requests, but either does not respond to the host’s SYN-ACK response, or sends the SYN requests from a spoofed IP address. Either way, the host system continues to wait for acknowledgement for each of the requests, binding resources until no new connections can be made, and ultimately resulting in [denial of service](https://www.imperva.com/learn/application-security/denial-of-service/).

* [**Ping of Death**](https://www.imperva.com/learn/application-security/ping-of-death/)

A ping of death (“POD”) attack involves the attacker sending multiple malformed or malicious pings to a computer. The maximum packet length of an IP packet (including header) is 65,535 bytes. However, the Data Link Layer usually poses limits to the maximum frame size – for example 1500 bytes over an Ethernet network. In this case, a large IP packet is split across multiple IP packets (known as fragments), and the recipient host reassembles the IP fragments into the complete packet. In a Ping of Death scenario, following malicious manipulation of fragment content, the recipient ends up with an IP packet which is larger than 65,535 bytes when reassembled. This can overflow memory buffers allocated for the packet, causing denial of service for legitimate packets.

* [**Slowloris**](https://www.imperva.com/learn/application-security/slowloris/)

Slowloris is a highly-targeted attack, enabling one web server to take down another server, without affecting other services or ports on the target network. Slowloris does this by holding as many connections to the target web server open for as long as possible. It accomplishes this by creating connections to the target server, but sending only a partial request. Slowloris constantly sends more HTTP headers, but never completes a request. The targeted server keeps each of these false connections open. This eventually overflows the maximum concurrent connection pool, and leads to denial of additional connections from legitimate clients.

* **Network Time Protocol (**[**NTP) Amplification**](https://www.imperva.com/learn/application-security/ntp-amplification/)

In NTP amplification attacks, the perpetrator exploits publically-accessible Network Time Protocol (NTP) servers to overwhelm a targeted server with UDP traffic. The attack is defined as an amplification assault because the query-to-response ratio in such scenarios is anywhere between 1:20 and 1:200 or more. This means that any attacker that obtains a list of open NTP servers (e.g., by a using tool like Metasploit or data from the Open NTP Project) can easily generate a devastating high-bandwidth, high-volume DDoS attack.

* **Hypertext Transfer Protocol (**[**HTTP) Flood**](https://www.imperva.com/learn/application-security/http-flood/)

In an HTTP flood DDoS attack, the attacker exploits seemingly-legitimate HTTP GET or POST requests to attack a web server or application. HTTP floods do not use malformed packets, spoofing or reflection techniques, and require less bandwidth than other attacks to bring down the targeted site or server. The attack is most effective when it forces the server or application to allocate the maximum resources possible in response to every single request.

* **Zero-day DDoS Attacks**

The “Zero-day” definition encompasses all unknown or new attacks, exploiting vulnerabilities for which no patch has yet been released. The term is well-known amongst the members of the hacker community, where the practice of trading zero-day vulnerabilities has become a popular activity (Imperva, 2020).

**2.3 Campus Network**

Rouse (2013) described a campus network as a proprietary local area network ([LAN](https://searchnetworking.techtarget.com/definition/local-area-network-LAN)) or set of interconnected LANs serving a corporation, government agency, university, or similar organization. Rouse (2013) explained that a typical campus encompasses a set of buildings in close proximity. The end users in a campus network may be dispersed more widely in a geographical sense than in a single LAN, but they are usually not as scattered as they would be in a wide area network ([WAN](https://searchnetworking.techtarget.com/definition/WAN-wide-area-network)). College and university campus networks interconnect administrative buildings, residence halls, academic halls, libraries, student centers, athletic facilities, and other buildings associated with the institution in a specific town or neighborhood. Corporate campus networks interconnect buildings that house key departments and staff members. The corporate campus network forms the user-facing aspect of the larger corporate network within a limited geographic area (Rouse, 2013).

As stated by Rouse (2013), Ideally all of the [nodes](https://searchnetworking.techtarget.com/definition/node) in a campus network are interconnected by means of [optical fiber](https://searchnetworking.techtarget.com/definition/fiber-optics-optical-fiber) media, taking advantage of  [Gigabit Ethernet](https://searchnetworking.techtarget.com/definition/Gigabit-Ethernet) or [10-Gigabit Ethernet](https://searchnetworking.techtarget.com/definition/10-Gigabit-Ethernet) technology. In some cases, [Wi-Fi](https://searchmobilecomputing.techtarget.com/definition/Wi-Fi) [hot spots](https://searchmobilecomputing.techtarget.com/definition/hot-spot) or even a [hot zone](https://searchmobilecomputing.techtarget.com/definition/hot-zone) make up the user end of the network, for example in university student centers or libraries where numerous people simultaneously use portable and mobile devices such as [notebook](https://searchmobilecomputing.techtarget.com/definition/notebook-computer) and [tablet](https://searchmobilecomputing.techtarget.com/definition/tablet-PC) computers to conduct research and carry on communications.

Occasionally the term ‘campus network’ is used in reference to geographically diverse Internet users with a common interest, such as the Roosevelt Institute Campus Network, a national student initiative, or the International Sustainable Campus Network (ISCN), a forum that supports colleges, universities, and corporations in their quest for sustainability in research and teaching (Rouse, 2013).

Building a Campus network is more than only interconnecting physical network infrastructure devices. The most challenging and important part of it is the planning and design phases where different technical variables and technologies need to be considered that could even effect the product selection and the design entirely. Also a good design is the key to the capability of a network to scale. A campus network is generally the portion of the network infrastructure that provides access to network communication services and resources to end users and also, devices that spread over a single geographic location. It might be a single floor, a building, or even a group of buildings spread over an extended geographic area (AL shawi, 2019).

AL shawi (2019) explained that Cisco’s hierarchical network design model breaks the complex problem of network design into smaller and more manageable phases. Each level or tier in the hierarchy is focused on specific set of roles. This helps the network designer and architect to optimize and select the right network hardware, software and features to perform specific roles for that network layer. A typical enterprise hierarchical campus network design includes the following three layers: The core layer that provides optimal transport between sites and high performance routing, the distribution layer that provides policy-based connectivity and control boundary between the access and core layers and the access layer that provides workgroup/user access to the network.

AL shawi (2019) stated that the two proven hierarchical design architectures for campus networks are the three-tier layer and the two-tier layer models. The three-tier layer model design can be used in large campus networks where multiple distribution layer and buildings need to be interconnected while the two-tier layer model can be used in small and medium campus network where core and distribution functions can be collapsed into one layer also known as a collapsed core or distribution model.

AL shawi (2019) explained that by using the hierarchical design model across multiple function blocks, the campus network will result in a more scalable and modular topology which is commonly known as ‘building blocks’ that allows the network to meet evolving business needs. The modular design makes the network more scalable and manageable by promoting deterministic traffic patterns. Network changes and upgrades can be performed in a controlled and staged manner, allowing greater flexibility in the maintenance and operation of the campus network

The core blocks are required for large networks only as it provides a very limited set of services and is designed to be highly available and operate in an always-on mode pattern. A separate core provides the ability to scale the size of the campus network in a structured fashion that minimizes overall complexity when the size of the network grows and the number of interconnections required to tie the campus together also grows (AL shawi, 2019).

**2.4 Big Data Analytics**

Sofiya and Soha (2015) defined the term ‘Big Data’ as the term used for large amount of complex and unprocessed data which are difficult and require a lot amount of time to process with the traditional processing methodologies. Big Data is characterized by volume, variety, velocity, variability veracity and complexity. Hence, to process this large data, big data tools are used to analyze the data and process according to its need. In today’s world, large amount of data are been produced every day. This huge amount of data is known as Big Data. (Sofiya and Soha, 2015). Intellects have predicted that this enormous amount of data may result in great waves of data which has made it a necessity to have tools that can handle these data in a systematic manner for proper applications in various fields (Sofiya and Soha, 2015).

Sowmya and Sravanthi (2017) described the term Big Data as being progressively used practically and extensively in the world both online and offline. Sowmya and Sravanthi (2017) stated that big data is not associated to computers only, but it is part of almost all other technologies, fields of studies and businesses. Big data is a blank term for the non-traditional strategies and technologies that are needed to gather, organize, process, and obtain insights from large datasets. Such a data is so large and complex that none of the traditional data management tools can store or process it efficiently. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years (Sowmya and Sravanthi, 2017). In this era of Big Data, data security is of paramount importance. With increase in the amount of data, the volume of attacks on networks also increases and one of the biggest threats nowadays is the Distributed Denial of Service (DDoS) attacks. (Paul et al., 2017).

According to Sriramoju, Namavaram and Ramesh (2017), Big Data has gained much attention in the last few years in the Information Technology (IT) industry as billions of people are connected to the internet worldwide, generating large amount of data at a very rapid rate. The generation of this large amount of data has given rise to various challenges along with huge benefits of big data to many organizations. In their paper, Sowmya and Sravanthi (2017) explained the features of Big Data also known as the 5Vs. They are: Volume, Value, Variety, Velocity and Veracity.

* Volume: In big data, volume is referred to as the enormous amount of data. It used to be employees created data, but nowadays that data is generated by machines, networks and human interaction on systems like social media, the volume of data to be analyzed is quite massive.
* Value: In terms of big data, value measures the usefulness of data in making decisions. A user can run certain queries against the data stored and thus can deduct vital results from the filtered data obtained and arrange it according to the dimensions they require.
* Variety: Variety in big data refers to the many sources and types of data which include both structured and unstructured data. We used to store data from sources like spreadsheets and databases but now, data comes in the form of emails, photos, videos, monitoring devices, Portable Document Files (PDFs), audio, etc. This variety of unstructured data creates problems for storage, mining and analyzing of data.
* Velocity: In big data, velocity is concerned with the pace at which data flows in from different sources like business processes, machines, networks and human interaction with things like social media sites, mobile devices, etc. The flow of data is large and continuous. This real-time data can assist researchers and businesses make valuable decisions that will provide strategic competitive advantages and Return on Investment (ROI) if the velocity of data is handled properly.
* Veracity: Big Data Veracity refers to the biases, noise and abnormality in data. Veracity in data analysis is the biggest challenge when compared to things like volume and velocity (Sowmya and Sravanthi, 2017).

The diagram of figure 2 explains the architecture of big data.

 **Figure 2: Big Data Architecture (Sowmya and Sravanthi, 2107)**

According to Sofiya and Soha (2015) Big Data analytics is the process of analyzing large datasets that is made up of different data types to uncover hidden patterns, unknown correlations, market trends, customers’ preferences and other useful information. Sofiya and Soha (2015) stated that the analytical findings can lead to a more improved operational efficiency, new opportunities, better services, fraud detection and other benefits.

Cardenas, Manadhata and Rajan (2013) describe big data analytics as a large scale analysis and processing of information. Cardenas et al., (2013) stated that big data analytics is actively in use in several fields and has attracted the interest of the security community in recent years for its promising ability to analyze and correlate data that are related to security efficiently and at an unprecedented volume.

As stated by Cardenas et al., (2013), data-driven information security can be dated back to fraud detection in banks and anomaly-based intrusion detection systems. Cardenas et al., (2013) said that a major problem in the information security community for decades now include analyzing logs, network flows and system events for forensics and intrusion detection and traditional technologies are not always enough to support long-term large scale analytics, performing analytics and complex queries on large unstructured datasets with incomplete and noisy features are inefficient but new big data applications are however starting to become part of security management software due to the fact that they can help clean, prepare and query data in heterogeneous, incomplete and noisy formats more efficiently.

Cardenas et al., (2013) stated that one of the most visible uses of big data analytics is in the detection of fraud and anomaly. Most credit card and phone companies have conducted large scale of fraud detection for years now but the custom built infrastructure that are needed to mine big data for the detection of fraud has not been economic enough to have wide scale adoption. The main impact of big data technologies however is that they facilitate a wide variety of industries to build infrastructures that are affordable for security monitoring. New big data technologies are enabling the analysis of large scale of heterogeneous datasets at an unequalled scales and speeds. These technologies are changing security analytics by facilitating the storage, maintenance and analysis of security information (Cardenas et al., 2013).

One of the promising features of big data analytics as explained by Cardenas et al., (2013) is the incorporation of unstructured data and multiple disparate datasets into a single analysis framework. Big data tools are particularly suited to become fundamental for advanced persistent threat detection and forensics because this type of threat operate in a low and slow mode and can occur over a long period of time while the victim remains unaware of the intrusion. Hence; to detect these attacks, large quantities of different data which include internal data sources and external shared intelligence data need to be collected and correlated to incorporate future information of an attack in the network’s history (Cardenas et al., 2013).

Panda Security (2019) stated that there has been a lot of talk on how big data can help companies better understands the profile of their customers by capitalizing on the large quantities of data it stores on them. Some companies are beginning to turn the transformed large volumes of data into useful information for businesses and using big data techniques to reinforce protection and also mitigate DDoS attacks; instead of trying unsuccessfully to use a server to track large volumes of traffic across the network, cloud-based big data systems are adopted to enable a more intelligent distribution. Tracking millions of IP addresses of all network traffic, monitoring abnormal traffic (analyzing the geolocalization of traffic, IP destination, and common attack ports) and applying learning algorithms to automatically detect relevant destination IPs are the types of systems that can [successfully detect these attacks in a more efficient way (Panda Security, 2019).](https://www.kentik.com/kentipedia/big-data-ddos-mitigation/)

Rani (2017) stated that the primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modellers and other analytics professionals to analyze large volumes of transaction data. Big data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as predictive analysis, data mining, text mining, and statistical analysis. Mainstream Business Intelligence (BI) software and data visualization tools can also play a role in the analysis process. But the semi-structured and unstructured data may not fit well in traditional data warehouses based on relational databases. Furthermore, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated frequently. Many organizations looking to collect, process and analyze big data have turned to a newer class of technologies that includes Hadoop and related tool such as MapReduce.

Cardenas et al., (2013) said the application of big data analytics to security problems has significant promise; therefore several challenges must be addressed to realize the true potentials of big data analytics.

**2.5 Big Data Tools and Techniques**

Joojay Huyn (2018) stated in his work that looking from the big data perspective, DDoS monitoring systems require modern big data technologies to process and analyze internet-scale network traﬃc in real time. Hence, the computational resources required to process traﬃc data of such large scale can easily overwhelm single machines, including large powerful ones, but modern big data technologies only require commodity hardware and software resources to process internet large-scale network traﬃc in a fault tolerant, distributed, horizontally scalable and parallelized manner. Consequently, many modern big data technologies interface seamlessly with many other useful systems such as data storage and warehousing systems, databases, data visualization dashboards, high-speed message buses, streaming platforms and they come packaged with advanced data mining tools that leverage and combine a wide variety of powerful artiﬁcial intelligence, machine learning, and statistical techniques to quickly assemble sophisticated algorithms for detecting, analyzing, and explaining anomalous and rare events hidden in large-scale data, such as DDoS attacks and other cyber-attacks.

Ullah and Babar (2018) in their paper listed some of the prominent big data tools to include Hadoop, Spark, Storm, Flume, Hbase, Hive, Kafka, Cassandra, and Mahout. Cardenas et al. (2013) proposed that big data tools and technologies would transform cyber security analytics by enabling organizations to be able to collect a large amount of security-related heterogeneous data from diverse sources such as networks, databases, and applications, perform deep security analytics in real-time and also provide a consolidated view of the security-related information. This combination of cyber security solutions and big data tools as stated by Cardenas et al., (2013) has given birth to the term “Big Data Cyber security Analytic”.

**2.5.1 Tools that are used in Big Data**

**Apache Hadoop:** Sowmya and Sravanthi (2107) explain that Apache Hadoop is aframework that can effectively store large amount of data in a cluster. This framework runs in parallel on a cluster and has the ability to allow one to process data across all nodes. Hadoop Distributed File System (HDFS) is the storage system of Hadoop which splits big data and distributes them across many nodes in a cluster. It also replicates data in a cluster hence, making it highly available. Apache Hadoop is a java based free software data.

**2. Microsoft HDInsight:** According to Sowmya and Sravanthi (2017) Microsoft HDInsight which is developed by Apache Hadoop is available as a ser**.** It is a Big Data solution from Microsoft powered vice in the cloud. HDInsight uses Windows Azure Blob storage as the default file system which also provides high availability with low cost.

1. **NoSQL:** NoSQL stands for Not Only SQL. As the traditional SQL can be effectively used to handle large amount of structured data, NoSQL is needed to handle unstructured data. NoSQL databases store unstructured data with no particular schema. Each row can have its own set of column values. NoSQL gives better performance in storing massive amount of data. There are many open-source NoSQL DBs that are available to analyze big Data (Sowmya and Sravanthi, 2017).
2. **Hive:** Hive is a distributed data management system for Hadoop. It supports SQL-like query option. Hive SQL (HSQL) is used to access big data. Hive is primarily used for Data mining purpose. It runs on top of Hadoop (Sowmya and Sravanthi, 2017).
3. **Sqoop:** Sqoop is a tool that connects Hadoop with various relational databases to transfer data. This can be effectively used to transfer structured data to Hadoop or Hive (Sowmya and Sravanthi, 2017).
4. **PolyBase:** PolyBase works on top of SQL Server 2012 Parallel Data Warehouse (PDW). It is used to access data stored in PDW. PDW is a data warehousing appliance built for processing any volume of relational data and provides integration with Hadoop allowing one to access non-relational data as well (Sowmya and Sravanthi, 2017).
5. **Big data in EXCEL**: Many people are comfortable with carrying out analysis in EXCEL which is a popular tool from Microsoft. So also can one connect data stored in Hadoop to EXCEL using EXCEL 2013. Horton works which is primarily working in providing Enterprise Apache Hadoop provides an option to access big data stored in their Hadoop platform using EXCEL 2013. The Power View features of EXCEL 2013 can also be used to easily summarize the data (Sowmya and Sravanthi, 2017).
6. **Presto:** According to Sowmya and Sravanthi (2017)Facebook developed recently an open-sourced Query engine (SQL-on-Hadoop) named presto which is built to handle petabytes of data. Unlike Hive, Presto does not depend on MapReduce technique and can quickly retrieve data.
   * 1. **Technologies used in handling Big Data**

Sowmya and Sravanthi (2017) in their paper explained that big data technologies are important in providing more accurate analysis which may lead to more concrete decision-making that will result in greater operational efficiencies, cost reductions, and reduced risks for the business and organizations. Harnessing the power of big data would require an infrastructure that can manage and process huge volumes of structured and unstructured data in real-time and can protect data privacy and security. There are various technologies in the market from different vendors including Amazon, IBM, Microsoft and many others to handle big data. The following two classes of Big Data technology are been explained.

1. **Operational Big Data**

Operational Big Data includes systems like MongoDB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored. NoSQL Big Data systems are also designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently. These features make operational big data workloads much easier to manage, cheaper, and faster to implement (Sowmya and Sravanthi, 2017).

1. **Analytical Big Data**

Analytical Big Data includes systems like Massively Parallel Processing (MPP) database systems and MapReduce that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data. MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL and a system based on MapReduce that can be scaled up from single servers to thousands of high and low-end machines. The Big Data handling techniques and tools include Hadoop, Map Reduce, and Big Table. Out of these, Hadoop is one of the most widely used technologies (Sowmya and Sravanthi, 2017).

* + 1. **Hadoop**

Hadoop is an Apache open source framework which is written in java. High volumes of data in any structure are processed by Hadoop. Hadoop allows distributed storage and distributed processing for very large data sets. The main components of Hadoop are: Hadoop distributed file system (HDFS) and MapReduce. The architecture of Hadoop is shown in figure 3. Hadoop has three layers. The two major layers are MapReduce and HDFS.

**Figure 3: The Architecture for Hadoop (Sowmya and Sravanthi, 2017).**

Sowmya and Sravanthi (2017) explained that Hadoop Distributed File System (HDFS) is a scalable and reliable distributed storage system that assembles the storage of every node in a Hadoop cluster into a single global file system. HDFS stores individual files in large blocks allowing it to efficiently store very large or numerous files across multiple machines and be able to access individual chunks of data in parallel without needing to read the entire file into a single computer’s memory. Reliability is achieved by replicating the data across multiple hosts with each block of data being stored by default on three separate computers. If an individual node fails, the data remains available and an additional copy of any blocks it holds may be made on new machines to protect against future failures. The diagram of figure 4 shows the architecture of Hadoop Distributed File System (HDFS).

****

**Figure 4: HDFS Architecture (Sowmya and Sravanthi, 2017).**

This process allows HDFS to dependably store massive amounts of data. For instance, in late 2012, the Apache Hadoop clusters at Yahoo had grown to hold over 350 petabytes (PB) of data across over forty thousand servers. Once data has been loaded into HDFS, we can begin to process it with MapReduce.

The concept of Map Reduce as explained by Sowmya and Sravanthi (2017)stated thatMap Reduce is the programming model that allows Hadoop to efficiently process large amounts of data. MapReduce breaks large data processing problems into multiple steps, namely a set of Maps and Reduces that can each be worked on at the same time (in parallel) on multiple computers. MapReduce is designed to work with HDFS. Apache Hadoop automatically optimizes the execution of MapReduce programs so that a given Map or Reduce step is run on the HDFS node that contains locally the blocks of data required to complete the step. Sowmya and Sravanthi (2017) stated that Map Reduce has proven itself in its ability to allow data processing problems that once required many hours to complete on very expensive computers to be written as programs that run in minutes on a handful of rather inexpensive machines. And while MapReduce can require a shift in thinking on the part of developers, many problems not traditionally solved using the method are easily expressed as Map Reduce programs. Figure 5 shows the architecture of Map Reduce.



**Figure 5: Map Reduce Architecture (Sowmya and Sravanthi, 2017).**

The following concept explains the components of Map Reduce (Thillaieswari, 2017).

* **Name Node**: The name node manages HDFS metadata. It does not deal directly with files.
* **Data Node**: The data node stores blocks of HDFS, defaults replication level for each block.
* **Job Tracker**: schedules, allocates and monitors job execution on slaves.
* **Task Tracker**: runs Map Reduce operations.

Hadoop MapReduceis a software framework for easily writing applications which process big amounts of data in parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. MapReduce actually refers to the following two steps of tasks that Hadoop program performs:

* **The Map Task:** It is the first task that takes input data and converts it into a set of data where individual elements are broken down into tuples (key/value pairs).
* **The Reduce Task:** This is the task that takes the output from a map task as input and combines those data tuples into a smaller set of tuples. The reduce task is always performed after the map task (Sowmya and Sravanthi, 2017).

Both the input and the output data are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executing any failed task. The MapReduce framework consists of a single master job trackerand one slave task trackerper cluster-node. The master is responsible for resource management, tracking resource consumption/availability and scheduling the jobs component tasks on the slaves, monitoring them and re-executing any failed task. The slaves’ task tracker executes the tasks as directed by the master and provide task status information to the master periodically. The job tracker is a single point of failure for the Hadoop MapReduce service which implies that if the job tracker goes down, all running jobs are halted.

* 1. **Review of Related Literature**

Mounir and Farah (2019) In their work, presented a cloud-based, fault tolerant, scalable and distributed intrusion detection system that used Apache Spark structured streaming and its Machine Learning library (MLlib) to detect intrusions in real-time. The efficacy and effectiveness of the system was demonstrated by implementing the proposed system within Microsoft Azure Cloud, as it provided them with both processing power and storage capabilities. They used the decision tree algorithm to predict the nature of incoming data using the MAWILab dataset as the data source to give better insights about the system capabilities against cyber-attacks. The results from their experiment showed about 99.95% accuracy and more than 55,175 events per second were being processed by the proposed system on a small cluster. Some of the issues in their approach that could be improved or worked on in the future are the improvement of the speed of data processing using the newly introduced feature of Spark structured streaming; continuous streaming which makes it possible to process data in milliseconds (1 ms) and not just seconds and the use of Fuzzy Decision Trees (FDTs) to classify data as against the traditional decision trees.

Francisco et al., (2019) in their paper presented a Machine Learning (ML) based DoS detection system which made inferences based on signatures that were previously extracted from samples of a given set of network traffic. Francisco et al., (2019) carried out their experiment using four modern benchmark datasets. Their software used the Random Forest Tree algorithm to classify network traffic based on samples taken by the sFlow protocol from network devices. They performed several experiments in order to calibrate and evaluate the performance of the system. The results obtained showed that the method they proposed was feasible and presented an improved performance compared with some other current and vital approaches available in the literature. Their proposed system was evaluated based on three intrusion detection benchmark datasets, namely, CIC-DoS, CICIDS2017, and CSE-CIC-IDS2018. They were able to classify different types of DoS/DDoS attacks; such as the Transmission Control Protocol (TCP) flood, User Datagram Protocol (UDP) flood, Hypertext Transfer Protocol (HTTP) flood, and HTTP slow attacks. Based on the experimental results, the Smart Detection approach delivered an improved detection rate (DR) of about 96% with a low false alarm rate (FAR) and a high Precision (PREC). Although the system achieved significant results in its scope, there is need for some improvements, such as a better hit rate among attack classes and an automatic parameter calibration mechanism that maximizes the detection rate of attacks.

Jing et al. (2019) in their study presented a DDoS attack detection method that is based on network abnormal behaviour in a big data environment. Based on the characteristics of flood attack encountered, the method would filter the network flows and leave only the “many-to-one” network flows so as to reduce the interference from normal network flows and improve the detection accuracy. The network abnormal feature value (NAFV) was defined to reflect the state changes of the old and new IP address of “many-to-one” network flows. The DDoS attack detection method which was based on the NAFV real-time series was built to identify the abnormal network flow states. The results from the experiments showed that the method had higher detection rate, lower false alarm rate and missing rate when compared with similar methods.

Patil et al., (2019) in their paper proposed a distributed and collaborative architecture which they called E-Had. It was capable of efficiently processing large amount of data by distributing it among a number of mappers and reducers in a Hadoop based cluster. Their proposed E-Had system was comprehensively validated using different publicly available benchmarked datasets and real datasets also which were generated in HA-DDoS test bed considering various detection system evaluation metrics. The results from their experiment showed that the proposed detection system was capable of early detection of different scenarios of DDoS attacks and equally able to differentiate them from flash crowds as the result produced 100% detection accuracy, 89% precision, 88% True Negative Rate (TNR), 100% Negative Predictive Value (NPV), 94% F-Measure, and 94% classification rate at False Predictive Rate (FPR) of 13%. As part of their future work, Patil et al want to explore the use of more advanced entropy and other information theory based divergence measures in detecting different types of DDoS attacks and to analyze high volume network traffic using Apache Spark and Apache Kafka application tools on Hadoop.

Ziyad et al., (2019) in their paper discussed the HTTP flooding attacks against the web servers on cloud computing and presented an organized survey that concerns the detection of HTTP GET flooding DDoS attack using Map Reduce in cloud computing based on the work of Lee et al (2011). In their paper, they focused on the DDoS attacks and especially on HTTP GET flooding attack. The DDoS attacks were discussed based on their history to have a proper insight into DDoS attacks. They also discussed the detection and prevention techniques for distributed denial of service attacks using the latest technique of Hadoop with Map Reduce. Ziyad et al., (2019) stated that regardless of the work that has already been done in this particular area, there was also a need to focus on the challenges that are still affecting the cloud services stating that a method should be developed that could help in detecting the application bug level as well as infrastructure level attacks, efficient research should be performed for developing such detection solutions and existing techniques should be modified in relation to the recent attacks and techniques or detection mechanism should be able to provide a good solution by detecting new attack patterns within the possible time to minimize its effect.

In their paper, Hameed and Ali (2018) used HADEC, a Hadoop-based live detection framework to tackle the efficient analysis of flooding attacks on a designed network by joining Map Reduce and HDFS. They implemented a counter-based DDoS detection algorithm on four major flooding attacks (HTTP, UDP, ICMP and TCP-SYN flooding attacks) in Map Reduce consisting of the map and reduce functions. A test bed was deployed to evaluate the performance of the HADEC framework for live DDoS detection on low-end commodity hardware. The result from their experiment showed HADEC’s capability of processing and detecting DDoS attacks in near to real-time. Their future work is to build similar systems to detect other types of DDoS attacks other than flooding attacks.

In their study on leveraging big data analytics for real-time DDoS attacks detection in Software Define Network (SDN), Dileep et al., (2018) proposed an architecture that combined big data analytic and SDN. The Support Vector Machine (SVM) was adopted and two other novel methods (Fuzzy Deterministic Clustering (FDC) and Threshold based Map Reduce (TMR)) were equally proposed for the detection of DDoS attack for a real-time DDoS detection in SDN environment. The result of comparing the methods showed that the proposed TMR detection method provided a high detection rate of 100% as compared to FDC and SVM methods.

In the study of Ting-Yuan and Chang-Jung (2018), they aimed to provide a fast and effective way to detect DDoS attacks and as such, integrated Apache Spark, a big-data computing framework, with the back-propagation artificial neural network for their attack detection. They used the Dataset from the Center for Applied Internet Data Analysis (CAIDA) DDoS Attack 2007 to simulate attack-associated network flows, and Dataset from the 1999 Defense Advanced Research Projects Agency (DARPA) Intrusion Detection Evaluation to simulate normal network flows so as to evaluate the approach of their study in experimental settings. By using the back-propagation artificial neural network for this study, they obtained a detection rate as high as 99.80% in their first stage experiments and true positive, true negative, false positive, and false negative rates were 99.81, 91.06, 8.94 and 0.19%, respectively. The experiments carried out in the second stage were conducted by the recombination of the parameters from the first stage into a one minute long data set of network flows and making use of a real-time detection system. The best accuracy they obtained was 87.18%, whereas the true positive, true negative, false positive and false negative rates were 87.61, 72.86, 27.14, and 12.39%, respectively. The results obtained from the second stage showed that they were inferior to those from first stage hence, it was clear that there was still chance for the improvement of the artificial neural network model in terms of replicability and the timeout value setting of the detection system As seen from the results of the two stages, most of the false positive rates were higher than the false negative rates. Although, the high false positive rates would not open the system to attacks, but they can prevent the normal users thereby causing them inconveniences. Similarly, the false negative results implied that the attacks were likely to scale through the detection system, and even though the false negative rates were relatively low; their consequences should not be overlooked. For accuracy purpose, these two types of error detection can be improved upon by adjusting the parameters of the artificial neural network and increasing the data volume for training. It was concluded that the method was effective in the detection process and the Apache Spark big data framework was fast in processing.

Lakshmanan and Karyhikeyan (2018) in their paper, studied the characteristics of DDoS attack, the various models involved in these attacks and provided a timeline of defense mechanism with their improvement to limit DDoS attacks. They proposed a novel scheme to detect DDoS attack efficiently by imploring the Map Reduce programming model using the K-means clustering algorithm and assigning only one task to each data node. Their result proved that the method was effective when only one task is assigned to a data node.

In their paper, Saraladevi et al. (2017) explored how machine learning with big data provided for higher prediction accuracy of malicious node in a network. A scalable framework for peer-to-peer botnet detection was designed using Hadoop and H2O machine learning platform. The parallel processing power of H2O was adopted and compared with Random Forest and Deep learning implementation results. The accuracy of finding malicious nodes attained a maximum level when multi-layer feed forward neural network was used.

Paul et al. (2017) in their survey on the detection of application layer DDoS attack using big data technologies, discovered that big data technologies could be used to detect application layer DDoS attacks whereby the time taken to analyze and process the data in real time is reduced and a better insight into the attack is given.

In their paper, Parakasha et al., (2016) performed a DoS attack on a client machine and tried to present a layered based approach to prevent the attack. Their prevention algorithm basically came in three layers and any user making any request has to scale through all three layers of protection in order to have access to the request. Parakasha et al., (2016) presented the three layers as: Puzzle, Mac filtration and Cryptography based authentication .The results obtained from the proposed algorithm outweighs other existing algorithms in terms of retrieval time taken to identify legitimate users and intruders abstained from the server. The proposed algorithm has a time complexity of O (M) as it takes a constant time in retrieving the MAC address from the data structure which makes the proposed algorithm better than all the existing algorithms.

Sumathi and Koppula (2015) in their paper presented various significant areas where data mining techniques appear to be strong candidates for the detection and prevention of DDoS attack. They proposed a new methodology based on big data map reduce concept that could be implemented for the detection and prevention of DDoS attack at the stage of verifying the network Internet Protocol (IP) addresses. Their method proved to be effective and could prevent loss of network resources, delay work and interruption of communication among various legal network users.

In their paper, Navale et al. (2014) proposed the use of counter-based and access pattern algorithm with Map Reduce in Hadoop framework. Flume tool was used for the generation of log data from a set of application servers. The counter-based algorithm using Map Reduce was adopted to detect DDoS attacks with Unified Resource Language (URL) counting. To lower the false positive rate alarm, response rate was adopted as against page request using time interval, threshold and unbalance ratio as parameters for the algorithm. The output of the counter-based algorithm served as the input to the access pattern algorithm in other to increase the efficiency of their system to be able differentiates normal clients from infected clients. The regression algorithm was implemented to predict the future activity of any detected attacker. The result obtained proved that the algorithms were effective and efficient.

Yeonhee and Youngseok (2014) in their paper proposed a novel DDoS detection method based on Hadoop that implemented a HTTP GET flooding detection algorithm in Map Reduce on the distributed computing platform. They implemented a counter-based detection method that counted the total traffic volume or number of requested web page with URL counting illustrated with the Map Reduce algorithm that used time interval, threshold and unbalanced ratio as input parameters. The result from their experiment showed that the counter-based detection method could easily be implemented in Hadoop and its performance showed the gain of using multiple nodes in parallel.

In their paper titled ‘Exploring Netflow Data Using Hadoop’, Zhou et al. (2014) presented enlightenment on the effective uses of Hadoop framework by identifying basic tasks that made up any exploratory analysis process of Net Flow dataset using Map Reduce to implement several basic fundamental analysis. They described their realization in Hadoop framework and evaluated the efficiency of different data formats. They concluded that shared state aggregation was sensitive to Hadoop set up and it was important to identify correct number of reducers for any particular job and system resources. They also evaluated the effects of combiners and came to the conclusion that combiners can also help scale up aggregation.

Wesam and Mehdi (2014) presented a hybrid approach known as centroid-based approach to detect and prevent a real-world DDoS attacks collected from “CAIDA UCSD DDoS Attack 2007 Dataset” and normal traffic traces from “CAIDA Anonymized Internet Traces 2008 Dataset” using unsupervised k-means data mining clustering techniques with proactive rules method. The results from their experiments showed that the centroid-based rules method performed better than the centroid-based method in terms of accuracy and detection rate. Their proposed solution obtained a very low false positive rate in the training process and testing phases. The accuracy results were more than 99 % in the training and testing phases respectively. It was concluded that the centroid-based rules method could be used in a real-time monitoring as DDoS defense system.

Prathibha and Dileesh (2013) in their paper presented a hybrid detection system using Snort and Hadoop. In their work, Prathibha and Dileesh (2013) used grid computing framework Hadoop to analyze the packets captured by Snort which is used for Big Data analysis. For a more user friendlier analysis they presented a data warehouse system for Hadoop and Hive was also used. For IP addresses that generated large number of packets, Snort rules were generated so that when the number of packets from a particular source exceeds a set number, the node will generate an alert to other nodes for possibility of an attack. The work was implemented with Hadoop cdh4.1.2, Snort 2.9.1, Hive 0.9.0. Hadoop cluster was implemented as pseudo distributed mode. Hping3 was used to generate the attack packets. Packets were sent in such a way that there were hundreds of packets sent in every second to simulate the flood attack. The packets were logged by running Snort in packet logger mode for five minutes. The log files where then converted into suitable readable format and each of them had an average of 400Mb size. The various packets were in different size and in different format. The source IP, destination IP, protocol and count were extracted and then, Map Reduce job was done on Hadoop cluster with various file sizes of 419.9Mb, 843.5 Mb, 1.7 Gb, and 2.5 Gb. The experiments were conducted with various block size of 32 Mb, 64 Mb, 128 Mb with the packet file size of 421.8 Mb. It was seen that the performance was better when block size was 128Mb, but the default size of 64 Mb also worked well. The results from their experiments showed that the analysis did not take much time to produce results and it was concluded that if the number of nodes increased, the performance can be made better. The results also showed that Snort rules generated after analysis was very much efficient in detecting many attacks.

Junho et al. (2013) in their study proposed an integration method between HTTP GET flooding attack and Map Reduce processing in other to obtain fast attack detection in cloud computing environment. In their experiment, authorized traffics among user request traffics were used. The authorized traffics were sent to master node in Hadoop cluster. The transmission traffics performed the preprocessing after classification using packet. The signature transmitted the preprocessed packets to data node in parallel. The results obtained in all the sections were similar which showed that the proposed Map Reduce method was better than the snort detection method because the processing time shorter with increasing congestion.

Jerome et al. (2011) in their paper proposed a novel approach for the efficient and accurate detection of botnets for large service providers and large campus network monitoring faced with high speed network connections. Their approach was based on processing Net Flow- related information in order to build a host dependency model that captured information on which host communicated with other host. Linkage analysis on this host analysis together with a clustering algorithm was used to build clusters of similar behaving nodes. They were able to show that bot-infected systems tend to belong to the same cluster. They were able to show the viability of their method on large dataset.

In their paper, Fang and I-Long (2010) proposed an agent based intrusion detection system (AIDS) that is a distributed detection architecture to detect both DoS and DDoS attacks. The method detected these attacks by using the goodness of fit test of chi-square test. It analyzed amount and variation of source IP addresses that sent packets to the network and statistics of IP address distribution. When an attacker employs attack tools to generate huge amount of packets of random source IP addresses, they checked to confirm if their chi-square value exceeded the set threshold. The experimental result showed that the method could effectively detect DoS/ DDoS attacks.

Most of the literature that has been reviewed in this research work reveals that many works has been done in the detection of DDoS attacks on different types of DDoS attacks using different big data technology tools. The unique feature of this research work is that the Apache Hadoop big data technology framework will be used to detect and mitigate the application layer DDoS attacks.

**2.7 Research Gap**

Based on the literature reviewed, the work of Hameed and Ali (2018) is very close to the proposed research but the difference is that they considered only flooding attacks in their work while this research work will be considering the detection and mitigation of application layer DDoS attack types. No work has yet been carried out using Apache Hadoop to detect and mitigate application layer DDoS attacks.

**CHAPTER THREE**

**METHODOLOGY**

**3.0 INTRODUCTION**

This chapter presents the working model for our proposed framework which will be tailored towards the research questions and the objectives of the study. The procedures and steps that will be followed to achieve the desired results include:

1. The capturing of network traffic to include both the legitimate and attack traffic through a web interface.
2. The generation of logs from captured traffic through the traffic handler.
3. The transfer of generated logs from the capturing server to the detection server.
4. The automatic process of detection is started by implementing the counter-based algorithm with set threshold to detect and mitigate any attack packet.
5. The notification of result obtained once an attack is detected.

The AHADM framework which comprises of five major phases will be applied to detect and mitigate application layer DDoS attack in close to real time. Each of the phases will be implemented as a distinct component but they will communicate with each other to carry out their assigned task. The capturing of network traffic, generation of logs and the transfer of logs to the detection server will all take place at the capturing server while the detection and mitigation of DDoS attack and the notification of results process will be done by detection server. The transfer of logs from the capturing server to the detection server will be carried out through the web using tshark.

**3.1 Work Flow Chart**

Capture Network Traffic

Generate Log Files

Detection and Mitigation of Attack Packets from Logs

Transfer Log Files to Detection Server

Result Notification

Detection Server

Capturing Server

Figure 3.1: The Work Flowchart of the Proposed AHADM Framework.

**References**

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials. 17* (4), 2347–2376.

AL shawi, M. (2019). *Campus Network Design Guideline.* [https://community.cisco.com/t5/networking-documents/campus-network-design- guideline/ta- p/3140160](https://community.cisco.com/t5/networking-documents/campus-network-design-%09guideline/ta-%09p/3140160)

Anstee, D., Chui, C. F., Bowen, P., & Sockrider, G. (2017). *Worldwide Infrastructure Security* *Report*, Arbor Networks Inc., Westford, MA, USA. <https://www.arbornetworks.com>

Cao, Y., Gao, Y., Tan, R., Han, Q., & Liu, Z. (2018). Understanding internet DDoS mitigation from academic and industrial perspectives. *IEEE Access Journal.* *6*, 66641–66648. [https://doi.org: 10.1109/ACCESS.2018.2877710](https://doi.org:%2010.1109/ACCESS.2018.2877710)

Cardenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big Data Analytics for Security. *IEEE Security and Privacy 11 (6). 74-76.*

Cardenas, A.A., Manadhata, P.K., & Rajan, S. P. (2013). Big Data Analytics for Security Intelligence. <https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Analytics_for_Se> [curity\_Intelligence.pdf.](curity_intelligence.pdf.) *University of Texas at Dallas@ Cloud Security Alliance. 5, 1-22.*

Dennis, P. (2017). *Safe guarding Campus Networks from DDoS Attacks.* <https://www.att.com/edu.>

Dileep, G. K., Jawaharial, N., & Guru, C. R. (2018). Leveraging Big Data Analytics for Real- Time DDoS Attacks Detection in Software Defined Network (SDN). I*nternational Journal for Research in engineering Application and Management. 4,* 677-684.

Distributed Denial of Service (DDoS) Attack (2020). <https://www.imperva.com/learn/application-security/denial-of-service/>

Dominic, B., Inyiama, H. C., Ahmed, A., Abdullahi, M. B., & Olaniyi, O. M. (2015). A Packet Sampling Threshold Technique for Mitigating Distributed Denial of Service (DDoS) Attacks in a University Campus Network*. International Engineering Conference (IEC, September, 2015)*. At the PTDT Lecture Theatre, Chemical Engineering Complex, Federal University of Technology, Minna, Niger State. 1, 399-406. [*https://www.seetconf.futminna.edu.ng*](https://www.seetconf.futminna.edu.ng/)

Emerging Threats (2020). [https:www.us.norton.com/internetsecurity-emerging-threats-what-is-a- ddos-attack-30sectech-by-norton.html](https://www.us.norton.com/internetsecurity-emerging-threats-what-is-a-ddos-attack-30sectech-by-norton.html)

Fang, Y.L., & I-Long, L. (2010). A DoS/ DDoS Attack Detection System Using Chi-Square Statistic Approach. *Journal of Systemic, Cybernetics and Informatics. 8,* 41-51.

Francisco, S. F., Frederico, A. F. S., Agostinho, M. B., Genoveva, V.S., & Luiz, F. S. (2019). Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning. *Hindawi Security and Communication Networks.* *2019* (1574749), 1-16. <https://doi.org/10.1155/2019/1574749>

Jerome, F., Shaonan, W., Radu, S., & Thomas, E. (2011). Bot Track: Tracking Botnets Using Netflow and Page Rank. <https://www.doi:10.1007/978-3-642-20757-0_1.>

Jing, C., Xiangyan, T., Jieren, C., Fengkai, W., & Ruomeng, X. (2019). DDoS Detection Method Based on Network Abnormal Behaviour in Big Data Environment. *ArXiv Computer* *Science Journal, Corpus ID: 85543777*, 1-11. <https://www.researchgate.net/publication>

Junho, C., Chang, C., Byeongkyu, K., & Pankoo, K. (2013). Detecting Web Based DDoS Attack Using Map Reduce Operations in Cloud Computing Environment. *Journal of Internet Services and Information Security. 3*, 28-37.

Hameed, S., & Ali, U. (2018). HADEC: Hadoop-Based Live DDoS Detection Framework. *EURASIP Journal of Information Security.11,* 1-19. [https://doi.org/10.1186/s13635-018- 0081-z](https://doi.org/10.1186/s13635-018-%090081-z%20%20%20)

Hoque, N., Bhuyan, M.H., Baishya, R.C., Bhattacharyya, D., & Kalita, J.K. (2014). Network attacks: taxonomy, tools and systems. *Journal of Network Computing and Application.* *40*, 307–324.

Joojay, H. (2018). Scalable Real-Time DDoS Traffic Monitoring and Characterization. UCLA Electronic Theses and Dissertations, 1-49. <https://escholarship.org/uc/item/Odr8k7td>

Kitten, T. (2015). *DDoS Attacks against Banks Increasing, blog, Bank Info Security.* <https://www.bankinfosecurity.com/ddos-a-8497>.

Kottler S. (2018). *February 28th DDoS incident report*. [https://github.blog/2018-03-01-ddos- incident-report/.](https://github.blog/2018-03-01-ddos-%09incident-report/.%20)

Lakshmanan, S.M., & Karthikeyan, P. (2018). MRI- A map Reduce Interpretation Framework for Mitigating DDoS Attacks in Big Data. *International Research Journal of Engineering and Technology. 5,* 4769-4773.

Marzano, A., Alexander, D., Fonseca, O., Fazzion, E.C., Hoepers, C., Jessen, K.S., Chaves, M. J., Cunha, I., Guedes, D.O., & Meira, W. (2018). The Evolution of Bashlite and Mirai IoT botnets. *IEEE Symposium on Computers and Communications (ISCC), Corpus ID: 53718419,* pp 00813-00818*.* <https://doi:10.11o9/ISCC.2018.8538636>

Mounir, H., & Farah, J. (2019). Comparative Study between Big Data Analysis Techniques in Intrusion Detection. *Journal of Big Data and Cognitive Computing. 3* (1), 1-13. <https://www.mdpi.com/2504-2289/3/1/1>

Navale, G. S., Vivek, K., Vijay, G., & Sravanthi, B. (2014). Detecting and Analyzing DDoS Attack Using Map Reduce in Hadoop*. International Journal of Industrial Electronics and Electrical Engineering. 2* (2),56-58.

Network Attacks and Network Security Threats (2020). [https://www.cynet.com/cyber- attacks/network-attacks-and-network-security-threats/](%20%20https://www.cynet.com/cyber-%09attacks/network-attacks-and-network-security-threats/)

Om, T (2019). *Understanding Application-Layer DDoS Attacks* <https://file:///C:/Users/user/Desktop/Research%20Papers%20downloaded/Understanding> <%20Application-Layer%20DDoS%20Attacks.html>

Panda Security 2019 report on *Businesses Are Using Big Data Security to Combat DDoS* *Attacks.* [https:www.panda.com/mediacenter/security/big-data-security-ddos-attacks/#top](https://www.panda.com/mediacenter/security/big-data-security-ddos-attacks/#top)

Patil, N. V., Rama, C. K., Krishan, K., & Sunny, B. (2019). E-Had: A distributed and collaborative detection framework for early detection of DDoS attacks, *Journal of King Saud University – Computer and Information Sciences.* 1-15. <https://doi.org/10.1016/j.jksuci.2019.06.016>.

Paul, A., Jyoti, M. G., & Moin, K. (2017). Detecting Application Layer DDoS Using Big Data Technologies. *Journal of Emerging Technologies and Innovative Research. 4* (6), 25- 31.

Prakasha, A., Satisha, M., Sri, T., Bhargava, S., & Bhalajia, N. (2016). Detection and Mitigation of Denial of Service Attacks Using Stratified Architecture. *4th International Conference on Recent Trends in Computer Science & Engineering. Procedia Computer Science 87* (2016), 275-280. <https://doi.org/10.1016/j.procs.2016.05.161>

Prathibha.P.G & Dileesh.E.D (2013).Design of a Hybrid Intrusion Detection System using Snort and Hadoop. *International Journal of Computer Applications. 73* (10), 5-10.

Rani, T.J. (2017). Intrusion Detection Using Big Data Analysis. *International Education and Research Journal of Computer Science. 3* (4), 27-28.

Rouse, M. (2013) *Campus Network.* [https://searchnetworking.techtarget.com/definition/campus- network](https://searchnetworking.techtarget.com/definition/campus-%09network)

Saraladevi, D., Sathiyamurthy, K., & Vijayaprabakaran, K. (2017). Big Data Framework for Peer-to-Peer Botnet Detection Using random Forest and Deep Learning. *International Journal of Computer Science and Information Security. 15* (11), 269-277.

Sofiya, M., & Soha, K. (2015). Big Data: Tools and Applications. *International Journal of Computer Applications. 115* (23), 7-11.

Sonal, A. H. (2016). Detection of Network Attacks Using Big Data Analysis. *International Journal on Recent and Innovation Trends in Computing and Communication. 4* (5), 86– 88. http://www.ijritcc.org

Sowmya, M., & Sravanthi, N. (2017). Big Data: An Overview of Features, Tools, Techniques and Applications*. International Journal of Engineering Science and Computing. 7 (6), 13644-13647.* <http://ijesc.org/>

Sriramoju, A. B., Namavaram, V., & Ramesh, G. (2017). An Overview of Big Data Challenges, Tools and Techniques. *International Journal of Research and Applications 4 (16),* *596-601.*<https://www.globalsciencepg.org>

Sumathi, R. M., & Koppula, S.R. (2015). Effective Detection and Prevention of DDoS Based on Big Data-Map Reduce. *Global Journal of Computer Science & Technology: C software & Data Engineering. 15* (6), *20-26.*

Thillaieswari, B. (2017).Comparative Study on Tools and Techniques of Big Data Analysis. *International Journal of Advanced Networking & Applications (IJANA) 8 (5), 61-66.*

Ting-Yuan, C., & Chang-Jung, H. (2018). Detection and Analysis of Distributed Denial-of- service in Internet of Things-Employing Artificial Neural Network and Apache Spark. *Journal of Sensors and Materials. 30* (4), 857–867. <http://dx.doi.org/10.18494/SAM.2018.1789>

Tom, W. (2015)*. Hadoop: The Definitive Guide.* O’Reilly Media, Inc., pp. 150-275.

Types of Network Security (2018). <https://blog.eccouncil.org/types-of-network-security-attacks>

Ullah, F., & Babar, M, A. (2018). ‘Architectural Tactics for Big Data Cyber Security Analytic Systems: A Review’ *Computer Science > Cryptography and Security, arXiv: 1802.03178 [cs.CR]. 1-48.*

Using Splunk for Big Data and DDoS Analytics. Blog post by Corero Network Security (30 April, 2020). [https://www.corero.com/blog/875-bursts-waves-and-ddos-what-you-need- to-know.html](https://www.corero.com/blog/875-bursts-waves-and-ddos-what-you-need-%09to-know.html)

Wesam, B., & Mehdi, E.M. (2014). Review Clustering Mechanisms of Distributed Denial of Service Attacks. *Journal of Computer Science. 10* (10), 2037-2046. <https://doi:10.3844/jcssp.2014.2037.2046>

What is a DDoS Attack (2020). [https://www.cloudflare.com/learning/ddos/what-is-a-ddos- attack/](https://www.cloudflare.com/learning/ddos/what-is-a-ddos-%09attack/)

Yeonhee, L., & Youngseok, L. (2014). Detecting DDoS Attacks with Hadoop. 1-3. [http://conferences2.sigcomm.org/conext/2011/workshops/StudentWorkshop/papers/1569 500805](http://conferences2.sigcomm.org/conext/2011/workshops/StudentWorkshop/papers/1569%09500805)

Zargar, S. T., Joshi, J., and Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (DDOS) flooding attacks. *IEEE Communications Surveys* *and Tutorials*. 15 (4), 2046–2069.

Ziyad, R. Al. A., Mohammed, A., Manmeet, M. S., Kamal, A., & Wajih, I. A. G. (2019). Detection of HTTP Flooding DDoS Attack using Hadoop with Map Reduce: A Survey. *International Journal of Trends in Computer Science and Engineering. 8* (1), 71- 77. [https://doi.org/10.30534/ijatcse/2019/12812019](https://doi.org/10.30534/ijatcse/2019/12812019%20)

Zhou, X., Petrovic, M., Eskridge, T.C., & Tao, X. (2014). Exploring Netflow Data Using Hadoop.  Proceedings of the Second ASE International Conference on Big Data Science and Computing. pp 1-10. [https://pdfs.semanticsscholar.org/29cc/6f582ba9d01f9f4e4d21d0650d2.pdf](https://pdfs.semanticsscholar.org/29cc/6f582ba9d01f9f4e4d21d0650d2.pdf%20)

Zuo, W. (2020). White Paper on DDoS Defense and Big Data. [https://e.huawei.com/de/publications/global/ict\_insights/hw\_331605/industry%20focus/h w\_3 27000](https://e.huawei.com/de/publications/global/ict_insights/hw_331605/industry%20focus/h%09w_3%0927000)