DDOS ATTACK DETECTION BY USING PACKET FEATURES

by

Ammar Yasir Korkusuz

B.S, in Telecommunication Engineering, Istanbul Technical University, 2011

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical Electronics Engineering
Boğaziçi University
2016

DDOS ATTACK DETECTION BY USING PACKET FEATURES

APPROVED BY:

Prof. Emin Anarım ....
(Thesis Supervisor)

Prof. Fatih Alagöz ....

Assoc. Prof. Ali Emre Pusane ....

DATE OF APPROVAL: 18.12.2015

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Professor Emin Anarim for his patience and contributions during this thesis. He has always been sophisticated, kind and ready to help.

Secondly, i want to thank Professor Fatih Alagoz for his precious time to approve this thesis and Assoc. Prof. Ali Emre Pusane, who motivated me to keep studying in Bogazici University 4 years ago.

Derya Erhan has great contribution in this thesis. I appreciate her and my other team-mates, Ramin Fouladi and Seray Özdemir, for their friendship and assistance. Thank you for being always helpful and supportive.

Finally, i would like to thank my father, Professor Mehmet Refik Korkusuz and my mother Reyhan Korkusuz, for encouraging me with their best wishes.

Ammar Yasir Korkusuz, Istanbul, 2015

ABSTRACT

DDOS ATTACK DETECTION BY USING PACKET FEATURES

DDoS attacks have been in internet life for a long time and most of hosts are still vulnerable for DDoS attacks. Complete detection and prevention of DDoS attacks is almost impossible, since their working method. Especially, if you are observing a network, not only one host, detecting DDoS attack can be much harder. To detect DDoS attacks existence, we used 11 features. We first used only threshold value of each features to detect DDoS attacks. Then, we used RMS (Root Mean Square) to improve our detection rates. We found different features are the best for Syn flood attack detection and UDP Flood attack detection.

The hardest issue for working on DDoS attacks is lack of publicly available datasets. We used UCLA dataset (University of California, Los Angeles), NUST datasets (National University of Sciences and Technology) and we composed 2 more datasets in Bogazici University to work on. In total, we applied our methods on 5 different datasets from 3 different institutes. Then, we compared our results with other similar studies. Our analysis showed that the best feature to detect TCP Syn flood attack is ”SYN/ACK ratio” and the best feature to detect UDP flood is ”flow generating rate”.

ÖZET

PAKET KARAKTERLERİ İLE DDOS ATAK TESPİTİ

Ddos saldırıları uzun bir zamandır sanal alemde görülmesine rağmen çoğu kullanıcı bu saldırılara karşı hâlâ savunmasızdır. Çalışma prensiplerinden ötürü, bu saldırıların önceden tespiti ve önlenmesi çok zordur. Özellikle tek bir servis değil de bir ağ denetleniyorsa, bu saldırıların tespiti çok daha zorlaşabilir. Bu çalışmada, DDoS atak tespiti için 11 tane parametre kullandık. İlk olarak, bu parametrelerin eşik seviyesine göre atak tespiti yaptık. Daha sonra; saldırı tespit oranını artırabilmek amaçlı karekök ortalama yöntemini kullandık. Syn Flood saldırıları ve UDP Flood saldırılarının tespiti için farklı parametrelerin daha iyi sonuç verdiğini gördük.

Bunun yanı sıra, DDoS saldırılarının tespiti noktasında karşılaşılan en büyük zorluk, kamuya açık saldırı verikümelerinin eksikliğidir. UCLA, NUST verikümelerini kullandık ve kendimiz Boğaziçi Üniversitesi’nde 2 tane daha verikümesi oluşturduk. Sonuç olarak, kullandığımız metotları 3 farklı kurumdan 5 farklı dataset üzerinde uygulama fırsatı bulduk. Daha sonra sonuçlarımızı bu alandaki diğer çalışmaların sonuçları ile karşılaştırdık. Analizlerimiz sonunda TCP Syn atağının tespiti için en iyi parametrenin ”SYN/ACK oranı” olduğunu, UDP atağı için en iyi parametrenin ”akış üretme hızı” olduğunu gördük.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

D_{KL}	Kullback-Leibler divergence
$H_2(X)$	Shannon entropy
JSD	Jensen Shannon divergence

LIST OF ACRONYMS/ABBREVIATIONS

ANP	Average Number of Packets per Flow
APL	Average Packet Length
DDoS	Distributed Denial of Service attack
DNS	Domain Name Server
DoS	Denial of Service attack
FGR	Flow Generating Rate
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FTP	File Transfer Protocol
HTTP	Hyper-Text Transfer Protocol
IP	Internet Protocol
ODGS	One Direction Generating Speed
PCF	Percentage of Correlative Flow
RMS	Root Mean Square
TCP	Transmission Control Protocol
TN	True Negative
TPR	True Positive Rate
TP	True Positive
UDP	User Datagram Protocol

1. INTRODUCTION

In 1994, the number of internet users was around 25 million people, which was around 0.4 percent of the world's population in that time. By 2015, 3 billion and 250 million people are using internet, which is around 40 percent of the world's population [1].

In the beginning, internet was invented for functionality, not security and it was indeed successful in reaching its aim. It provides users fast, easy and cheap communication. Moreover, it offers reliability and certain level of quality of service. The internet is regulated by distributed way, so no common policy can be deployed to users. Protocols were open and policies were based on mutual respect. This kind of design has security problems and it has some issues which would offer occasions for distributed denial of service attack. Attacks which have an aim to obstacle the availability of computer systems or services are mainly called DDoS attacks [2].

The first time DDoS attacks became aware was at year 2000. 15 year old Canadian teenager started series of DDoS attacks to the biggest websites of its time; Yahoo!, Fifa.com, Amacon.com, Dell, eBay and CNN. He succeeded to shut these websites down for a while. Yahoo! was the biggest search engine in that time and it was not reachable for a few hours. Attacks caused 1.2 billion US dollars in global economic damages [3]. DNS root servers was under attack at year 2002 and 9 over 13 root servers were affected by this attack. In 2007 Estonia and in 2008 Georgia experienced DDoS attack, the source of both attacks was claimed as Russia. In 2010, after WikiLeaks released some confidential information about world politics, it was hit by DDoS attack as well. Then, WikiLeaks supporters attacked to MasterCard, Visa and PayPal payment systems, to protest the blockage of donation to WikiLeaks by these companies [4].

In 2015, DDoS attacks were increased aroun 130 %, comparing with the same period of 2014. The longest attack duration was more than 64 days. Moreover, 20 % of the all attacks lasted more than 5 days [5].

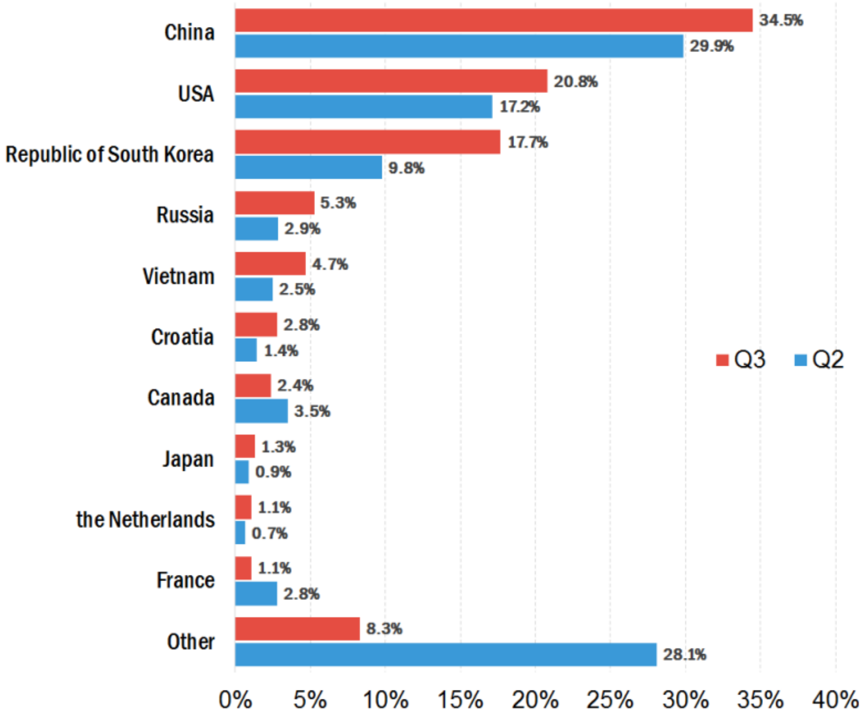


Figure 1.1. Target of DDoS Attack in 2015 q3 [6].

Kaspersky anti-virus company stated that almost 92 % of all attack's targets were 10 countries, in the third quarter of 2015. China is the first target of all DDoS attacks worldwide, and then USA comes second. South Korea has significant increase comparing with second quarter's result. The entire results can be seen in figure 1.1 [6].

If we look at the source of the attacks, we see similar results. China is the biggest source of DDoS attacks and USA follows it as second. Then North Korea comes with a big increase. The other details can be seen in Figure 1.2 [6].

There are different kinds of DDoS attacks, such as TCP Syn flood attacks, UDP flood attacks, ICMP flood attacks. According to general statistics, TCP Syn Flood attacks cover the biggest proportion of DDoS attacks as shown in Table 1.1. It is easy to launch and the attacker do not need to send big size of packets. Second most common DDoS attacks is UDP Flood attacks. UDP is very common protocol and UDP packets can pass easily from routers and firewalls [7].

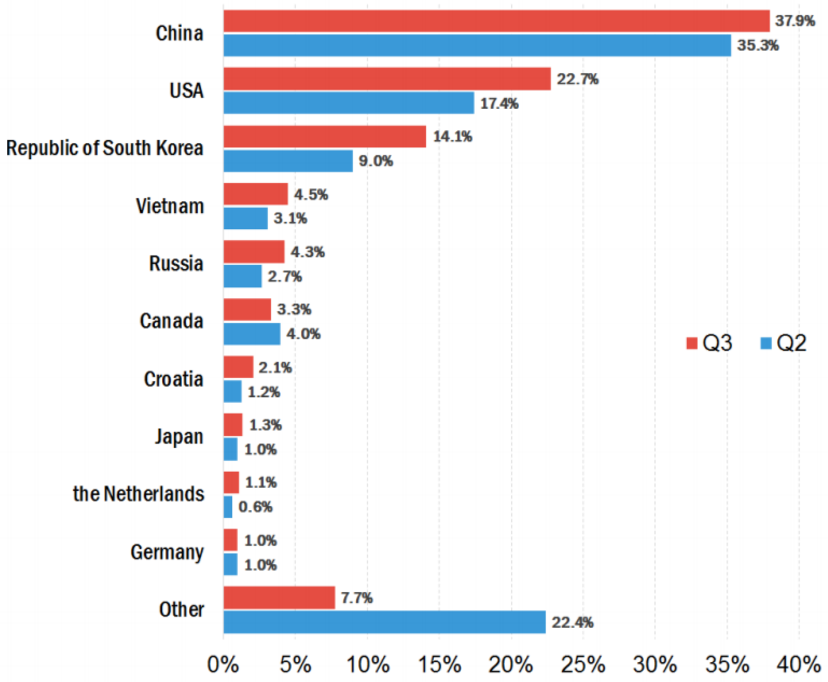


Figure 1.2. Sources of DDoS Attack in 2015 q3 [6].

Table 1.1. Proportion of The Various DDoS Attack On Different Protocols [7].

Protocols	**The proportion in total DDoS Attacks**
TCP	90 % - 94 %
UDP	2.4 % - 5 %
ICMP	2.1 % - 2.6 %
Others	2 % - 2.9 %

Analysing general statistics, we decided to work on TCP Syn flood attacks and UDP Flood attacks. Our aim was to find the best features to detect these 2 types of DDoS attacks by using most accurate and less complicated way. We wanted to find best detection scheme for TCP Syn flood and UDP flood attacks individually.

In the beginning the hardest work was to find a dataset to work on. We found 3 public datasets;

- UDP flood dataset UCLA (University of California, Los Angeles)
- TCP Syn flood dataset of NUST (National University of Sciences and Technology)
- UDP flood dataset of NUST (National University of Sciences and Technology)

Then we composed 2 more datases in Bogazici University Campus. At the end we had 5 different datasets from 3 different institutes.

For detection of attacks, we used 11 easy and flexible features those we can apply on datasets. Some of them gave good results and some of them were poor in detecting attacks. After the first results, we used root mean square (RMS) to improve our results.

The organization of this thesis is as follows. In the next section, DDoS attack's technical details are given. In the third chapter, related studies are discussed. In the fourth chapter, features and methods those used to detect DDoS attack are explained. In the fifth section, datasets and their results are introduced. After that, comparative studies those similar to this work are discussed and the results are compared. Last chapter is conclusion which ends this thesis and gives future work recommendations.

2. DDOS ATTACKS

2.1. Victims and Aims of DDoS Attacks

2.1.1. Victims of DDoS Attacks

3 most common target of DDoS can be listed as below [8];

(i) E-commerce: Internet is the key element of many businesses. Millions of customers are purchasing goods/services every day via their computer or mobile devices. According to JP Morgan, 2011 e-commerce revenue is 680 billion dollars. Users like shopping online because e-commerce websites are secure, responsive and always available. This big market attracted criminals to see the big opportunities. Online shopping companies are victimized mass attacks. Hackers commit data theft, extortion, identity theft and fraud. DDoS attacks can cause millions of dollars loose, if service is gone or slowed. Even though you can get your services up in a short time, your customers think that your website is not secure. According to survey by the Ponemon Institute, the average total cost of single data breach was more than 7.2 million dollars in 2011.

(ii) Financial Services: Nowadays, a person can use online banking systems, global money transfer or payment processes over internet. People can make these activities in anywhere, anytime and any devices. They expect their information to be secure and the service is reliable and quick. DDoS attacks can have very big loose of money, if these services got slow or interrupted. Especially the attacks whose target's transaction processes or big trade systems can cause catastrophic affects. Many banks and stock exchanges, including Bank of America, New York stock exchanges, reported that they have been under DDoS attack.

(iii) Online Gaming: It is very big area with millions of player who can play so many games such as gambling and video games. People can play these games on different platforms; PCs, Xbox, PlayStation. Performance and availability are the key elements of online gaming. The activist hacker group Anonymous directed DDoS attack to Sony PlayStation network in 2010.

2.1.2. Aims of DDoS Attacker's

DDoS attackers are motivated by different reasons. We can classify these incidents in 5 categories depending on the aim of attackers [9].

(i) Economical gain: These are the most dangerous and effective attacks. Attackers' target is usually big corporations so attackers usually have enough technical knowledge and experience.

(ii) Revenge: Disappointed people those experienced personal injustice behave are in this category. They usually have low technical knowledge.

(iii) Ideological opinion: Attackers who have ideological dispute with others may launch an attack to their target. For example; political problems cause DDoS attack to Estonia (2007) and WikiLeaks (2010).

(iv) Intellectual Challenge: In this category, motivation of attackers is to increase their experience and to learn how to do DDoS attack. They are usually young hacking volunteers. There exist so many easy to use DDoS attack tools.

(v) Cyberwarfare: These attackers are who belong to the military or terrorist organizations. They usually try to destroy civilian departments, financial and telecommunication organizations, energy and water infrastructures.

2.2. DoS and DDoS Attacks

2.2.1. Dos Attacks

A Denial of Service attack is an attack that one or more hosts attack to the victim and try to obstacle its useful work. The target of this attack can be server, router, an entire network, an Internet Service Provider (ISP) or country. DoS attack makes these targets unavailable by intended users mainly by exhausting target devices CPU and memory resources. This may cause the device to be very slow or shut down [10].

A DoS attack illustration is shown in Figure 2.1.

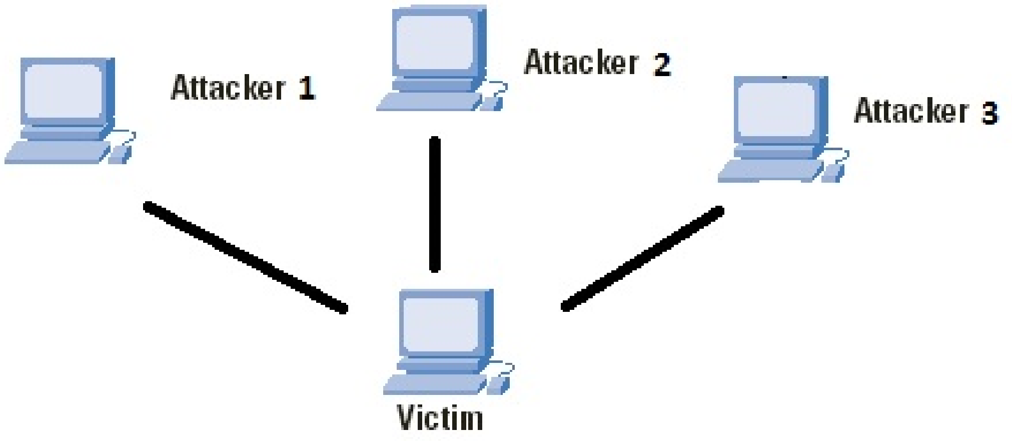


Figure 2.1. DoS Attack Illustration.

2.2.2. DDos Attacks

A distributed denial of service attack is a DoS attack with a multiple sources of attack. The distributed means many to one dimension. Presence of DDoS attacks is a big problem of reliability of the internet. In a DDoS attack, an attacker takes over the control of compromised hosts over the internet. Compromised host means that a host, who doesn't have secure configuration, can be controlled by attackers. DDoS attacks can be classified into 2 types; direct attacks and reflector attack. In a case of direct attack, an attacker sends a huge volume of attack packets directly to victim. The packets can be UDP, TCP or ICMP. Reflector attack means that, compromised machines send attack packets to intermediary machines and these intermediary devices send attack packets to victim. This system hides compromised machines IP address. Typical DDoS attack consists of network with 3 different elements [10]:

- A host device which is commanded by an attacker.
- A number of masters those are controlled by the host device.
- A number of slaves those are controlled by masters and cause DDoS attack.

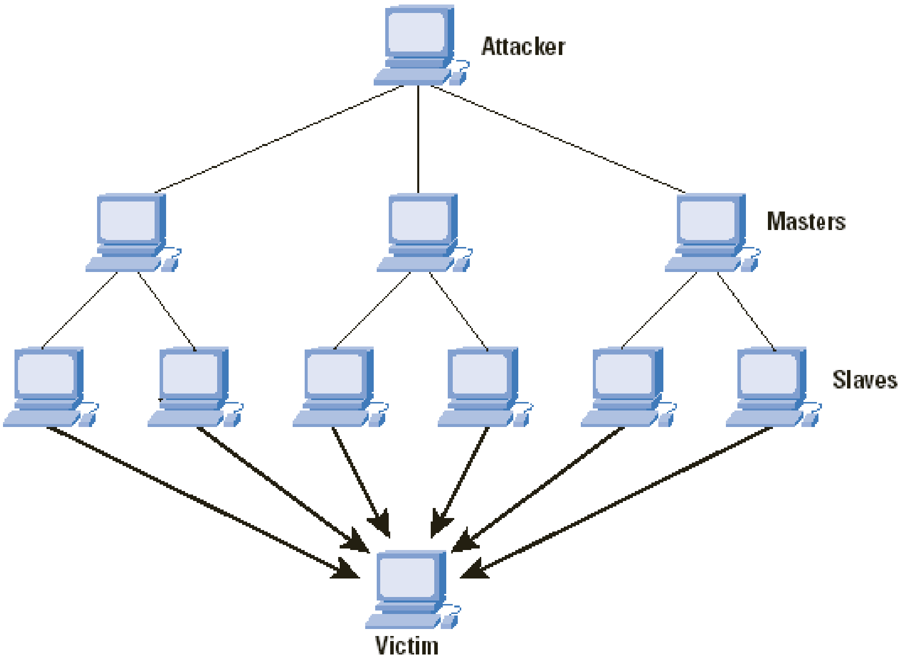


Figure 2.2. DDoS Attack Illustration.

A DDoS attack illustration is shown in Figure 2.2. There are many DDoS attack tools on the internet, anybody who does not have expertise on hacking can launch a serious attack. Examples of DDoS attack tools are Trinoo, Tribe Flood Network (TFN), Shaft, TFN2K.

2.2.3. Challenges of DDoS Attack Detection

Challenges of DDoS attack detection can be seen below; [2, 11, 12];

- DDoS attacks send very similar packets like legitimate packets. Pattern of the traffic can be very similar to normal traffic. The attacker needs only volume, the content is not needed. Moreover, size of packets can be very small so it can be ignored by protection systems.
- DDoS packets are generally created by similar tools but they have different characteristics. So, statistical analysis are needed for detecting DDoS attack.
- Internet is highly independent. There is no possibility to deploy security policies

for global internet.

- Connection resources are not limitless. Each host or connection can have limited resources.
- Limited knowledge to work on DDoS attacks exist. There is no common characteristics of packets, they seem like legitimate packets.
- There is no standardized dataset or testing approaches for DDoS attacks.

2.2.4. Tools to Launch DDoS Attack

There are a lot of different DDoS attack tools on the internet and they do not need any specific knowledge to use. Some of them have special properties like disable or uninstall itself, when certain conditions exist. Here is the list of DDoS attack tools and descriptions [13] ;

(i) Floodnet: It is a Java application which uses TCP flood to send packets to target with non-existent source IP addresses. This tool makes victim waste its bandwidth and processing capability.

(ii) Trin00: One of the earliest DDoS attack tool. It implements UDP flood attack via master/slave mechanism. Attacker sets the starting time and victim IP of the attack. Source IP addresses of this tool is not spoofed, since it uses slaves for attack.

(iii) Tribal Flow Network (TFN): TFN uses master/slave mechanism as well. It can launch different kinds of DDoS attacks such as TCP Syn flood, UDP flood, ICMP flood.

(iv) Stacheldraht: It is combination of TFN and Trin00. It can make IP spoofing and it can make different kind of attacks. It uses encrypted TCP connection between attacker and handler.

(v) TFN2K: This is improved version of TFN attack tool. Detection of attacker is much harder since connections are encrypted. It can work on Linux systems and Windows systems.

(vi) Shaft: Shaft uses UDP between master and agent connection. Telnet is used for remote connection. Shaft can launch TCP Syn flood, UDP flood or ICMP flood attacks. Also, it can launch all of them simultaneously, if it is desired. Connection is encrypted between agents and master.

(vii) Mstream: It launches TCP Syn flood by sending packets with ACK bit set as 1. Source IP addresses and ports are random.

(viii) Trinity: In this tool, the attacker does not need any handler device to connect agents. Trinity can launch almost all kinds of DDoS attacks.

(ix) Hping: This is very simple and effective DDoS attack tool. It can launch TCP Syn flood, UDP flood and ICMP flood attacks. Source IP addresses can be spoofed and some subnets can be chosen as source IP address pool. This can allow attacker to send victim packets which has an IP address of devices those are in victim's network. It allows attacker to send as many packet as wanted and it allows attacker to send packets to any port number wanted. Moreover, only one strong machine is enough to shut down a server, since it can send a lot of packets in the same time. To compose a dataset in Bogazici University, we used this DDoS attack tool. TCP Syn flood command of Hping is;

```
hping 193.140.199.86 -S -p 80 spoof 192.168.1.150 -i u1000
```

Victim IP: 193.140.199.86

Packet Type: Syn (-S)

Destination Port: 80 (-p)

Source IP: 192.168.1.150 (spoof)

Packet number: 1 packet in every 1000 miliseconds / 1000 packet in every 1 second (-i u)

2.3. Transport Control Protocol (TCP)

Transport Control Protocol (TCP) offers a communication via port numbers. It is connection oriented protocol. Connection oriented means that, when host A wants to communicate with host B, a connection has to be set up between hosts A and B. This set up lets both hosts to send their data as a stream of bytes and receive the data as stream of bytes. TCP is transport-layer protocol which is one level higher than IP. TCP can send and receive very different length of data streams. A transport unit in TCP is named as segment. Features of TCP can be summarized as below [10];

- Numbering: Header of the segment has sequence and acknowledgment numbers. This sequence number is appointed by TCP; it shows the first byte of the stream and each byte's number. The sequence number can start from any number; it should be unique for the each direction of connection.
- Flow control: To get over the problem of receiving too much data traffic, the receiver checks the amount of traffic that was sent by transmitter. The control was done via byte check. Sliding windows is used for flow control.
- Error control: In case of any corrupted, lost or duplicated segments, TCP can recognize it. Acknowledgment, retransmission and checksums are used for error control.
- Congestion control: The traffic amount is not controlled only by receiver, but also network, to be sure that network does not have huge amount of traffic.

2.3.1. TCP Packet Field Descriptions

TCP connection has 3 parts. In the first part, the connection set up via 3 way hand-shake method. Data transfer starts in the second part. In the last part, the connection is terminated and all allocated resources became idle for this connection.

The following descriptions summarize the TCP packet fields [14].

- Source Port and Destination Port: Identifies points at which upper-layer source

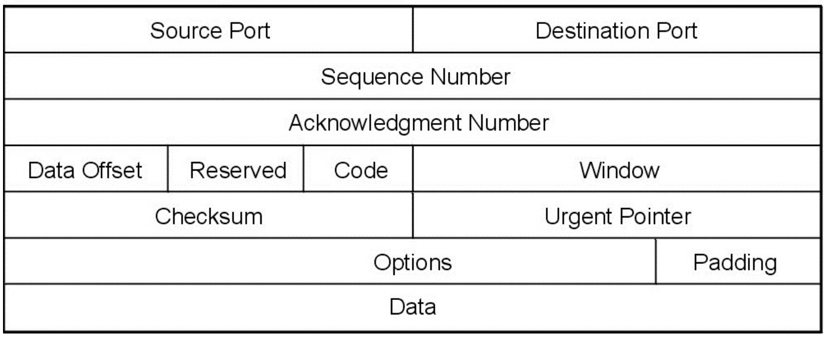


Figure 2.3. TCP Packet Fields.

and destination processes receive TCP services.

- Sequence Number: Usually specifies the number assigned to the first byte of data in the current message. In the connection-establishment phase, this field also can be used to identify an initial sequence number to be used in an upcoming transmission.
- Acknowledgment Number: Contains the sequence number of the next byte of data the sender of the packet expects to receive.
- Data Offset: Indicates the number of 32-bit words in the TCP header.
- Reserved: Remains reserved for future use.
- Flags: Carries a variety of control information, including the SYN and ACK bits used for connection establishment, and the FIN bit used for connection termination.
- Window: Specifies the size of the sender's receive window (that is, the buffer space available for incoming data).
- Checksum: Indicates whether the header was damaged in transit.
- Urgent Pointer: Points to the first urgent data byte in the packet.
- Options: Specifies various TCP options.
- Data: Contains upper-layer information.

2.3.2. 3-Way Handshake Rule

3-way handshake is the part where connection set up, before data transmission starts between host and server. Server should allocate a TCP port for host to connect. Steps of 3-way handshake rule shown in figure 2.4. [10];

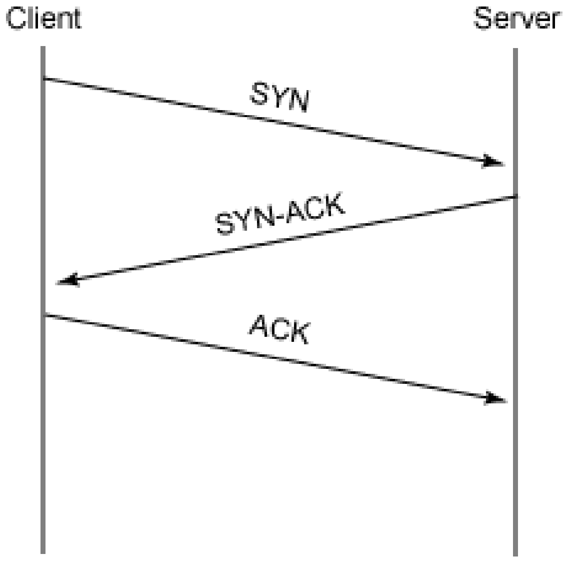


Figure 2.4. 3-way Handshake Mechanism.

(i) A host sends TCP packet with Syn flag set to request a connection. Its number of segment sequence is a random number A.

(ii) The server responds this request with a SYN+ACK flag set. Acknowledgment number is one more than received sequence number (A+1). At this time, server sets a random number B as sequence number. This packet informs client that the previous SYN packet was successfully received and server is ready for connection.

(iii) In this part, the server receives another TCP packet with ACK flag set from the client. Sequence number of this packet is one more than received packet's sequence number (B+1).

2.3.3. TCP Syn Flood Attack

Syn flood attack is the most common DDoS attack in the world. Its main target is to prevent server from establishing new connections with other hosts. As explained above, clients send Syn packet to server, then server sends back SYN/ACK packet. At this point server allocates necessary resources for client to connect. Basic concept is to make server not able to accept new connections, by allocating all resources.

In TCP Syn flood attacks, the attacker sends a large number of TCP Syn connection requests to the victim server. Victim server replies these packets with SYN/ACK packets, then wait for ACK packet to return. However, the attacker never sends Ack. In this case, the server waits for ACK packet for a certain amount of time. Waiting time can be 1 minute up to 4 minutes. This connection is called "half open" connection [15].

An illustration of TCP Syn flood attack can be seen in figure 2.5.

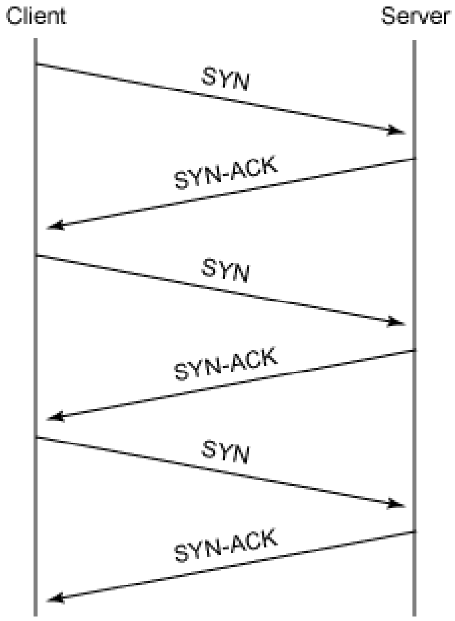


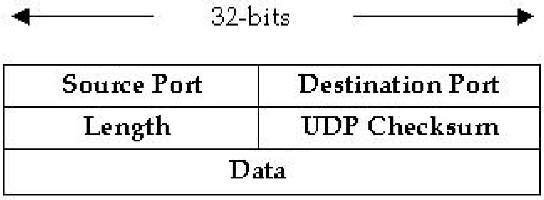
Figure 2.5. TCP Syn Flood Attack.

Attackers usually use IP spoofing techniques to increase the attack's severity. Also, professional attackers use fake IP addresses. If attackers send thousands of TCP

Syn packets to the server with real IP addresses, the server will send SYN/ACK packet to real devices. After that, those real devices will respond to server with TCP Rst (reset) packet to close the connection. To prevent this mechanism, professional attackers use fake IP addresses and make sure that the victim server is not going to receive TCP Rst packet and the server will wait until the time out. After a certain amount of received connection, the waiting queue of the server will be filled completely. Then, the server will start rejecting new connection requests. At the end, the server will not be reachable by other hosts and its resources will be unavailable [15].

2.4. User Datagram Protocol (UDP)

UDP is a widely used protocol on the internet, for applications such as DNS resolution, online games, instant messages, and real time audio communications. UDP provides port numbers to distinguish different requests for different users. Also, it optionally provides checksum capability to check if arrived data is untouched. TCP is a dominant protocol over internet. It checks and resends all lost packets and needs acknowledgments for successful transmissions. However, these additional services cause delays and it reduces speed of transmission. UDP sends packets without any additional services and any verification. Packets can be lost or arrive out of order. But, it provides faster and real time communication. That's why UDP is a good protocol to use in real time application which has no tolerance for latency. It is called connectionless protocol, since it does not allocate any resources before connection. Figure 2.6 shows UDP packet format [16].



Source Port	Destination Port
Length	UDP Checksum
Data	

Figure 2.6. UDP Packet Format.

2.4.1. UDP Flood Attacks

UDP flood attacks abuse UDP's connectionless mechanism and its limitations. The attacker sends a large number of UDP data to victim server. Victim server's queue becomes completely filled and it will not be able to respond other legitimate user's requests. IP spoofing is used in UDP Flood attacks, as well as TCP Syn flood attacks. Attackers use mostly fake IP addresses to make attack more severe. Normally, victim server receives the packet and checks if there is any application waiting for this port. If not, victim server sends back ICMP data packet to close the connection. When an attacker sends thousands of UDP packets to the victim server, the waiting queue will be full and victim server will refuse new requests [17]

Figure 2.7 illustrates UDP flood attack.

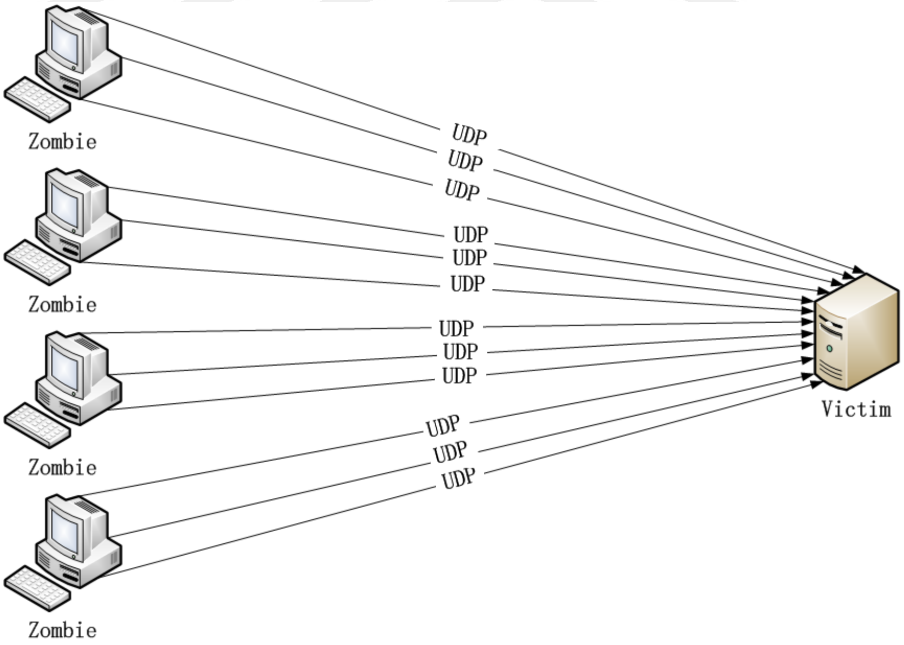


Figure 2.7. UDP Flood Illustration.

2.5. Performance Analysis of Detection Mechanism

To evaluate performance of DDoS attack detection mechanisms, true positive rate (TPR) and false positive rate (FPR) are used.

$$\mathbf{TPR} = \frac{TP}{FN + TP} \tag{2.1}$$

TP is the number of samples those correctly defined as attack. FN is the number of normal traffic those defined as attacks.

$$\mathbf{FPR} = \frac{FP}{TN + FP} \tag{2.2}$$

FP is the number of attack samples those detected as attacks. TN is the number of samples those correctly defined as legitimate traffic.

3. RELATED WORK

DDoS attacks have been very hot topic in last 10 years for researchers. Each method has its own advantages and disadvantages. Moreover, each research group compose their own dataset and they apply their method on their dataset. In that case, there is no possibility to work on those datasets for other working groups to compare their results.

Sun et al. proposed a method to count number of SYN-FIN packets difference. Under normal conditions, each TCP connection which completes the three-way handshake starts with a SYN packet, and finishes with a FIN or RST packet. In this case, there should be equal number of SYN and FIN packets in a long run. If there is inequality of these packets, like SYN packets are more than FIN packets, it means that there is attack going on [18].

Nashat et al stated that the difference between SYN packets and SYN/ACK packets can be good way to detect an attack. During normal 3-way handshake process, SYN packets arrive and SYN/ACK packets transmit. In this case, there should be similar number of SYN packets and SYN/ACK packets in a long run. However, during an attack, there will be so many SYN packets and victim will not be able to send SYN/ACK packets back [19].

DDoS attacks are usually done by packets with spoofed IP addresses. In this case, entropy of source IP addresses seem to be logical feature to detect DDoS attacks [20]. Also, the target of attacks is usually only one victim and it is expected that entropy of destination IP addresses would decrease, since attack packets' target will be the only victim [21].

A group of packets those have the same source IP, destination IP, source port, destination port are called ”flow”. During SYN flood attack, there will be so many packets with different IP addresses and ports and in this case, there will be less packets in a flow, comparing with normal time [22].

Correlative flow means a group of packets those have opposite source IP - destination IP and source port - destination port couples, such as packet 1: source IP is A destination IP B source port is K destination port is L, packet 2: source IP is B destination IP A source port is L destination port is K. It is expected that, there will be so many incoming traffic but no outgoing traffic during a DDoS attack. So, percentage of correlative flows will reduce in case of an attack [23].

Rahmani et al. used a ”connection size distribution” feature to detect attacks. Connection size distribution depends on the service which customer use, such as FTP, HTTP, DNS. Normally, it is assumed that this distribution is generally stable in a short term. Also, new IP addresses are suspicious in the system. If there are new IP addresses, their connection size distribution is checked and decided whether it is attack or not [24].

In the other study, source IP address distribution and packet number in a certain time interval are features are used to detect attacks. Lee et al. used hash functions to map source IP of incoming traffic. And they also checked the number of total packets. So, it is expected that, source IP of incoming traffic and total number of packets should increase during DDoS attack [25].

Sharma et al. investigated the difference of some entropy based features. During an attack, uncertainty will increase because of nature of DDoS attack. So, they used source IP, destination IP, source port, destination port, TCP flag set, protocol and length distributions. Then, they calculated these features’ entropies to detect attack presence [26].

4. DESCRIPTIONS of FEATURES and METHODS

4.1. Methods

4.1.1. Root Mean Square

The Root Mean Square is the square root of the arithmetic mean of the squares of the values. It can be defined as:

$$RMS = \sqrt{\frac{\sum_{i=1}^{N}(Y_c - Y_0)^2}{N}} \tag{4.1}$$

RMS makes arrays more smooth and it makes single increases less visible. In our work, we used RMS to increase our detection rate.

4.1.2. Entropy

Entropy is well-known and valuable concept in information theory. It is measure of uncertainty which was introduced by Claude E. Shannon in 1948. Entropy is commonly used in DDoS studies. Some of our features are entropy based. Using entropy has 2 main advantages. The first advantage is that, calculated value of entropy does not depend variable itself, the value depends only on each variable's distribution, for example source IP. This prevents entropy to increase during new legitimate source IP connections. Second advantage of using entropy is the number of IP addresses can be large in normal traffic. However, entropy looks only on unique IP addresses and its distribution. Entropy helps to realize DDoS attack, even though the total number of packets are less than normal time [24].

In information theory, bigger entropy values are expected when the selected variable is more random. As opposite, when the uncertainty of the selected variable is less random, then the entropy value is small.

Let us consider a discrete probability distribution;

$$P = p_1, p_2, p_3.....p_n \tag{4.2}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{4.3}$$

In this case, Shannon entropy is defined as [21];

$$H_2(X) = -log_2 \sum_{i=1}^{n} p_i^2 = 1 \tag{4.4}$$

We used entropy of source IP, destination IP, source port, destination port and TCP flag type as feature for attack detection.

4.1.3. Kullback-Leibler Divergence

Kullback-Leibler divergence is another method used in DDoS studies. It is non-symmetric distance between two probability distributions. For normal flow P and abnormal flows Q, the formula of Kullback-Leibler divergence is below [27] :

$$D_{KL}(P||Q) = \sum_{i=1}^{n} p_i log_2 \frac{p_i}{q_i} \tag{4.5}$$

4.1.4. Jensen Shannon Divergence

Jensen Shannon Divergence is a smoothed and symmetric version of Kullback-Leibler divergence. For 2 discrete probability distribution;

$$P = p_1, p_2, p_3.....p_n \tag{4.6}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{4.7}$$

$$Q = q_1, q_2, q_3.....q_n \tag{4.8}$$

$$\sum_{i=1}^{n} q_i = 1 \tag{4.9}$$

Jensen Shannon Divergence is defined as

$$JSD(P, Q) = \frac{1}{2}KL(P, M) + \frac{1}{2}KL(Q, M) \tag{4.10}$$

where KL is Kullback-Leibler distance between P and Q and M is the mean distribution of P and Q;

$$M = \frac{P + Q}{2} \tag{4.11}$$

$$m_i = \frac{p_i + q_i}{2} \tag{4.12}$$

Jensen Shannon divergence is average of Kullback-Leibler distances to average distribution M. It can be expressed also in this way [28] ;

$$JSD(P,Q) = \frac{1}{2}[\sum_{i=1}^{n} p_i log_2 \frac{p_i}{m_i} + \sum_{i=1}^{n} q_i log_2 \frac{q_i}{m_i}] \quad (4.13)$$

- If P and Q are identical (pi = qi), JSD = 0.
- If P $\neq Q$, then $JSD > 0$
- If p_i and q_i are orthogonal distributions ($p_i.q_i = 0$), it is a bounded metric ($0 \leq$ JSD $\leq$ log(2)).

JSD is a divergence between 2 distributions to detect anomalies. Comparing 2 legitimate traffic distribution should give small value and comparing a normal and an attack traffic distribution should give high value [28].

4.2. Features

We selected 11 different features for attack detection.

4.2.1. Entropy, KL Distance, JS Distance Of Source Ip Addresses

Entropy is a measure of uncertainty in given distribution. During the DDoS attack period, there will be so many packets with different source IP addresses since the traffic comes from different sources. It is assumed that The uncertainty of source IP addresses will be increased [20]. Figure 4.1 shows entropy of source IP address difference between attack and normal packets.

After seeing results of source IP entropy, we wanted to see the Kullback-Leibler and Jensen-Shannon divergences between attack and normal dataset. We assumed that, the divergence values would keep small between 2 normal datasets but it should increase between normal and attack datasets.

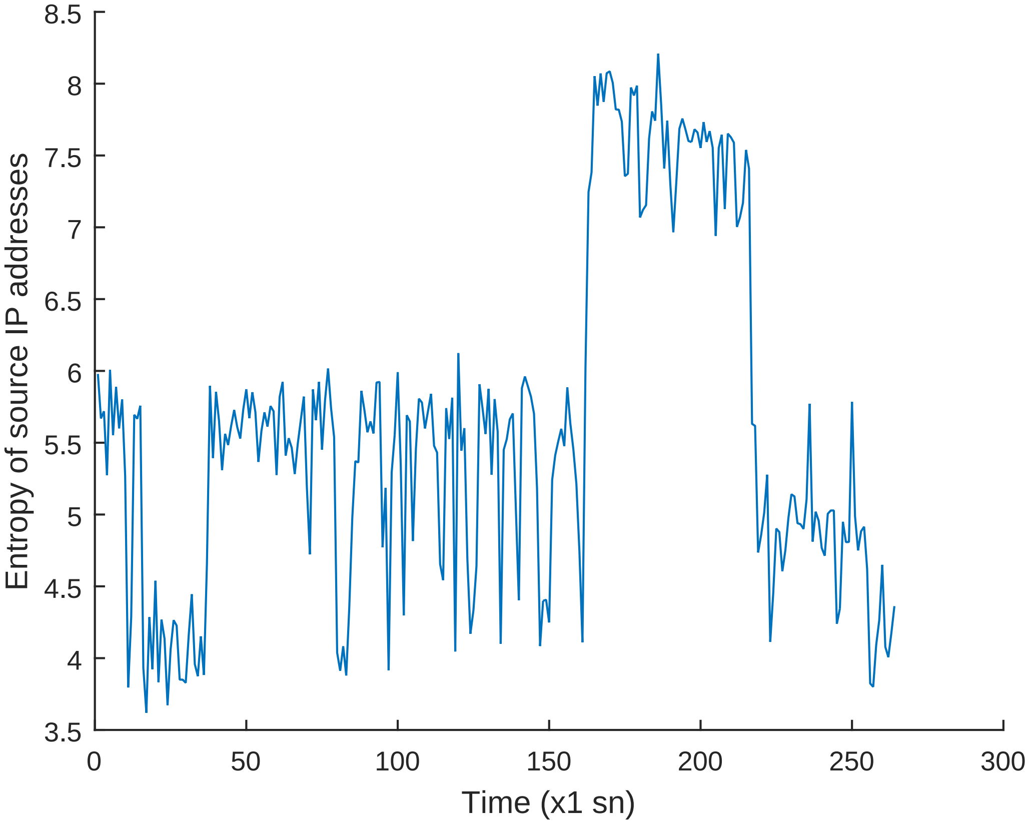


Figure 4.1. Entropy of source IP addresses for the attack maintains between 161-215 seconds.

4.2.2. Entropy Of Destination Ip Addresses

During DDoS attack, there will be so many packets with the same destination IP addresses. Because the target of attack is probably one or a few servers and hosts. In this case, unique destination IP addresses converges to a small value. Hence, uncertainty of destination IP addresses is decreased and entropy is decreased as well since it is measure of uncertainty [21]. Figure 4.2 shows increase for entropy of destination IP addresses during an attack.

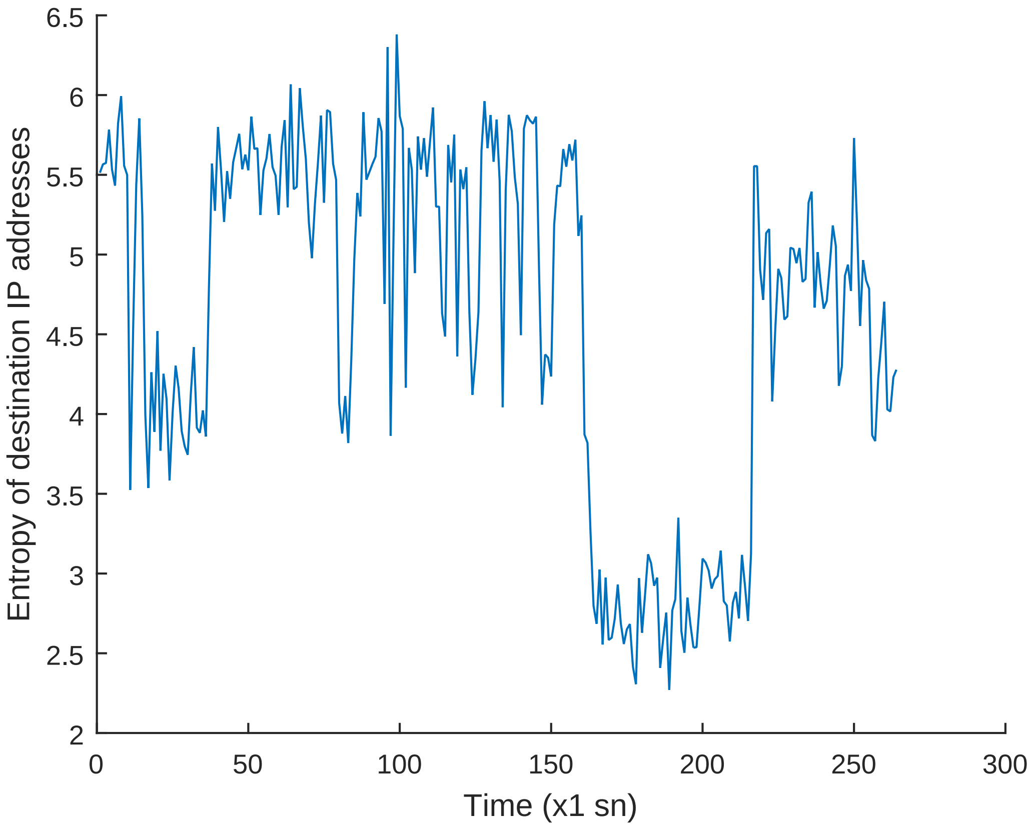


Figure 4.2. Entropy of destination IP addresses for the attack maintains between 161-215 seconds.

4.2.3. Entropy Of TCP Flag Type

The entropy value of TCP flag type is worth observing because DDoS attacks use specific flag type such as SYN flags, during SYN flood attack. If the entropy of packet type converges to a small value, it is could be symptom of DDoS attack [26]. Figure 4.3 shows difference of TCP flag entropy.

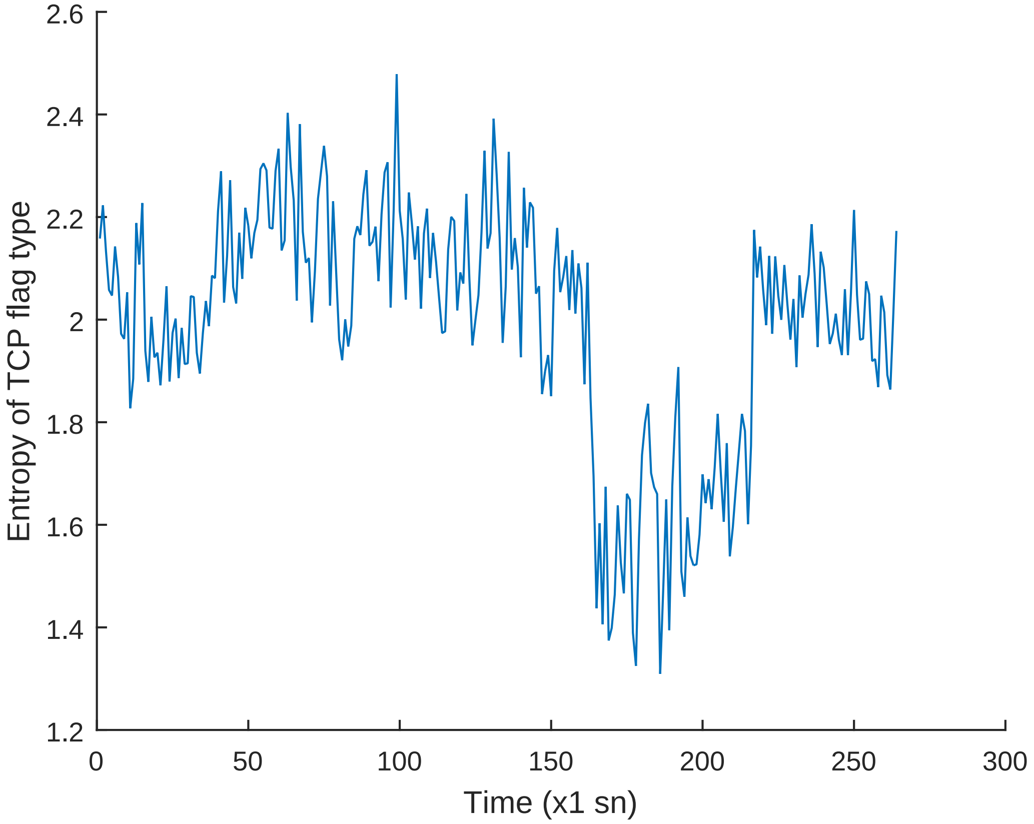


Figure 4.3. Entropy of TCP flag type for the attack maintains between 161-215 seconds.

4.2.4. Entropy Of Source Port Numbers

The entropy value of source ports is also good parameter to measure. DDoS attacks usually use variable number of ports. Normally, incoming traffic is distributed on some certain source ports. However, in attack case, there will be so many used port numbers, its certainty and entropy will increase [26]. Entropy of source port numbers can be seen in figure 4.4.

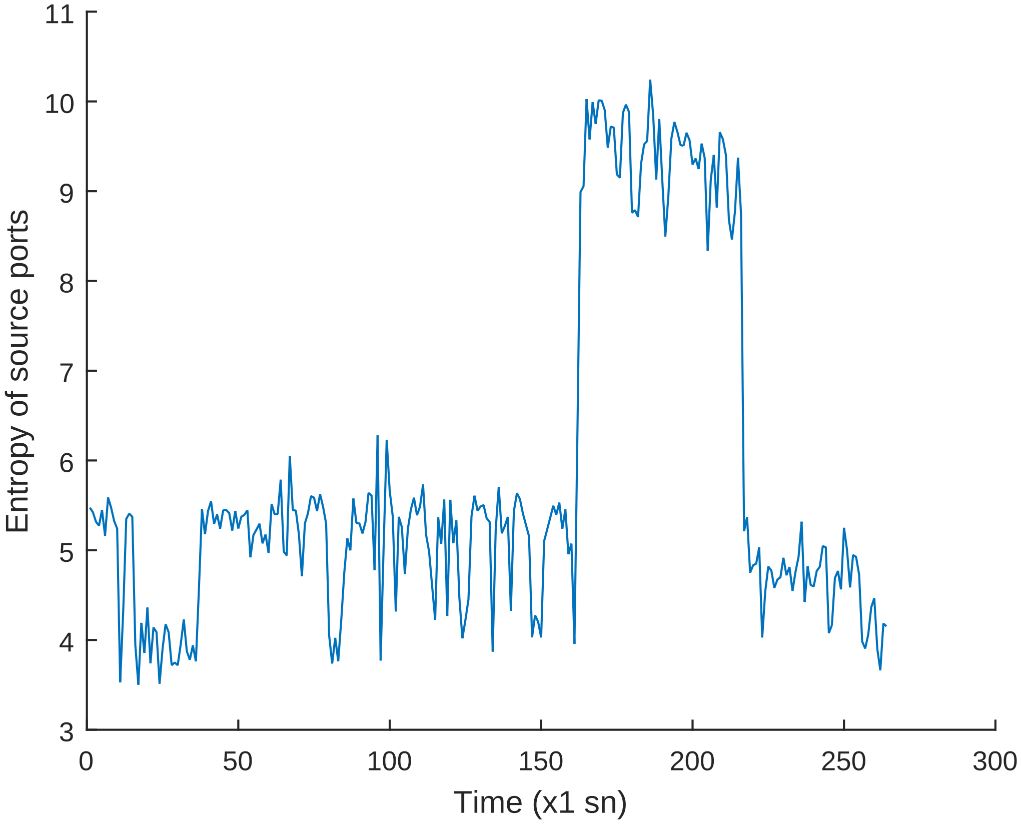


Figure 4.4. Entropy of source ports for the attack maintains between 161-215 seconds.

4.2.5. Entropy Of Destination Port Numbers

In normal networks, servers serve only in a few ports. If the port is closed, server sends back ”unreachable port” warning packet. However, DDoS attackers send too many packets to too many different ports to make server busier. At the end entropy and uncertainty of destination port numbers will be more than normal traffic [26]. Entropy of destination port numbers are visible in figure 4.5.

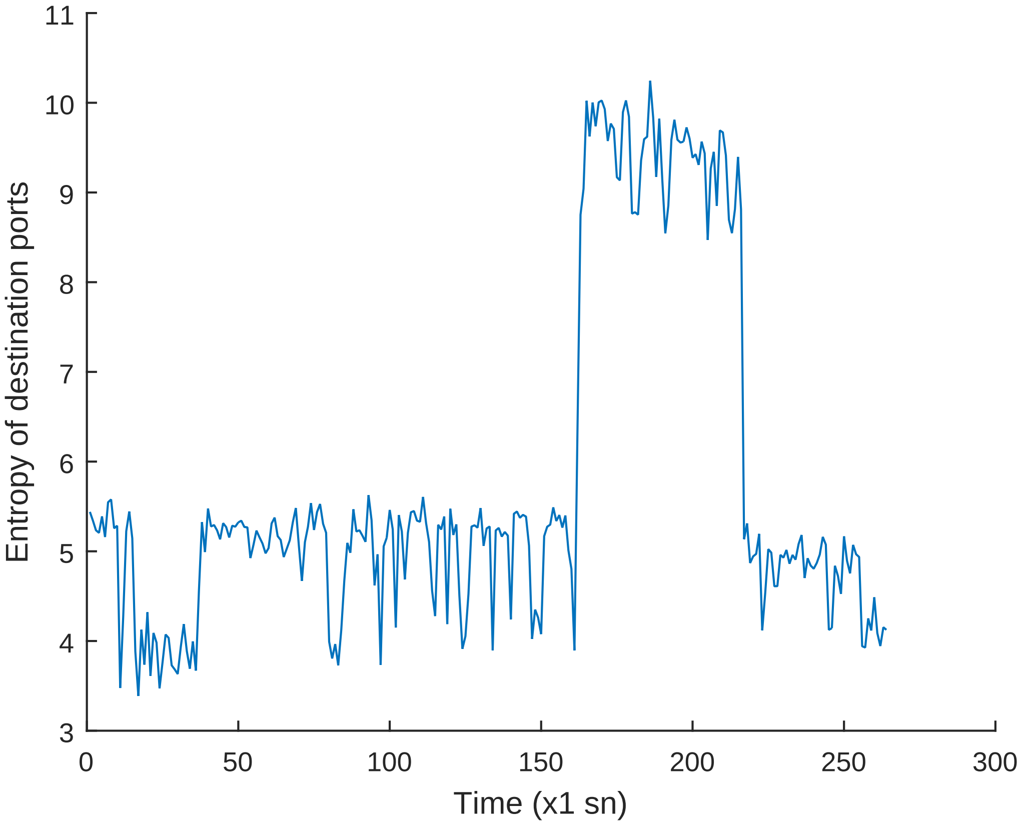


Figure 4.5. Entropy of destination ports for the attack maintains between 161-215 seconds.

4.2.6. Average Packet Length

Content of attacking packets have usually similar sizes. So, average packet length converges to that value. APL represents this characteristic of DDoS attack by calculating the average packet length in an interval [29] . Figure 4.6 is the comparison of

APL of normal traffic and abnormal traffic. As we can see, there is a significant change of APL when DDoS happens.

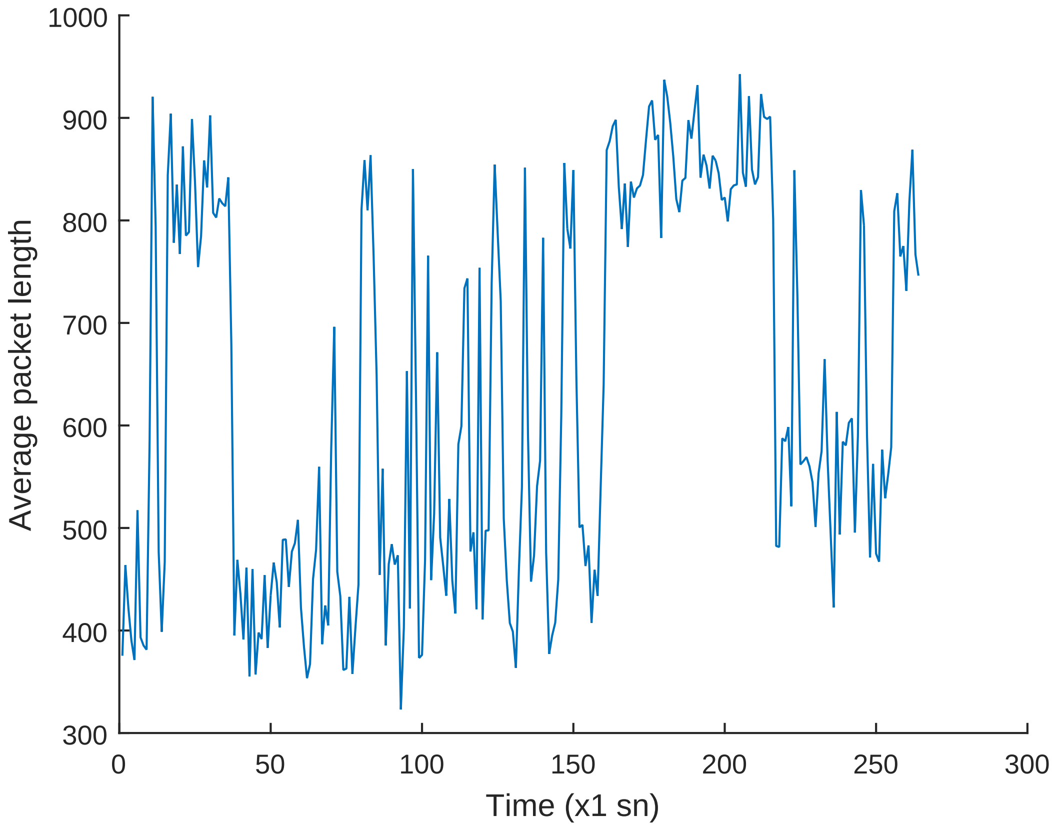


Figure 4.6. Average packet length for the attack maintains between 161-215 seconds.

4.2.7. Average Number Per Flow

Average number of packet in a flow is the first feature which differs significantly when an attack is going on. During an attack, the number of flows in an interval becomes close to the number of packets. In another words, flows with same keys would consist of just packet. This is mostly because by the means of the IP spoofing. The source IP addresses change so frequent and there would be fewer packets with similar IP addresses [22]. Average number per flow definition is;

ANP = Number of Total Packets/Number of Flows

An experimental comparison of ANP (Average Number of Packets in each flow) between normal traffic and DDoS traffic is depicted in Figure 4.7 different intervals of time in normal and attack traffic.

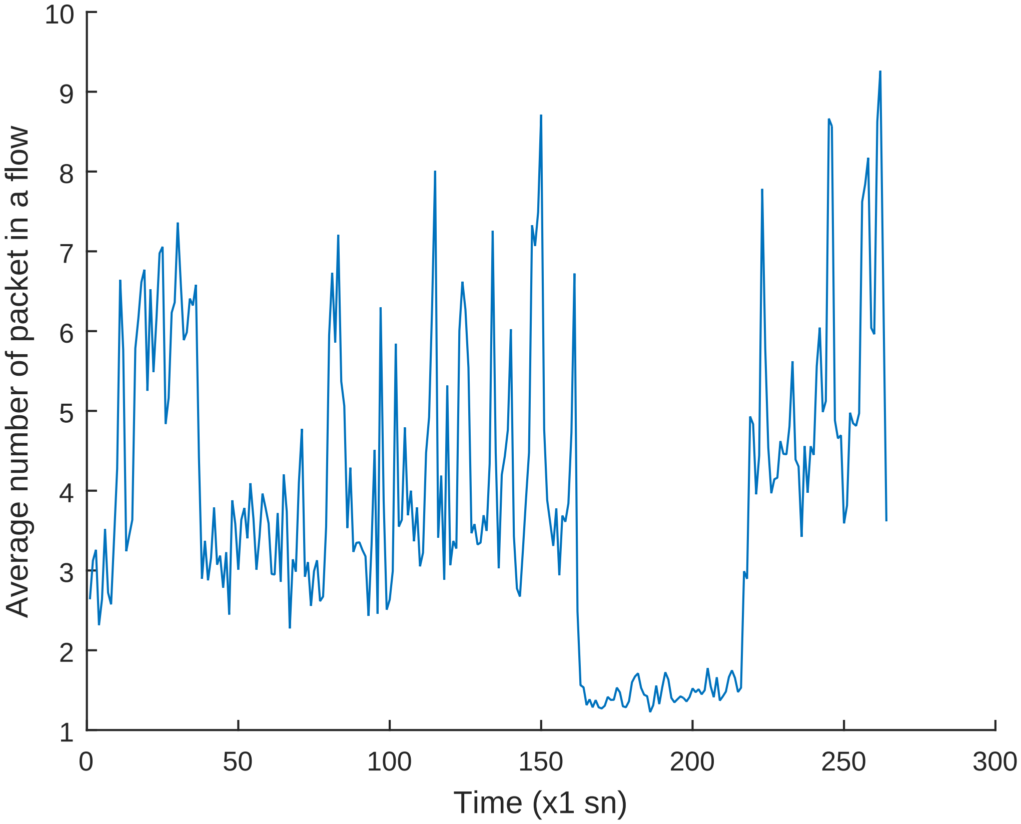


Figure 4.7. ANP for the attack maintains between 161-215 seconds.

4.2.8. Percentage Of Correlative Flow

Proportion of number of correlative packet flows (CFN) to number of flows (FN) in each interval of time is another feature which is useful to detect the attack. This is an index for correlation of the source IPs and Destination IPs and the ability of the victim to reply to receiving packets. During attack, since the source IP addresses are spoofed, even though the victim is still capable of answering attacking packet's requests, the replying packets cannot reach the attacking machines. To extract this feature, we need to find those flow numbers whom are answered by the victim [23].

Proportion of CFN to FN is defined below;

PCF = Number of Correlative flows/ Number of Flows

When an attack is going on, the change in number of correlative flows is negligible while the number of flows rises dramatically. Consequently, PCF in attack times decreases which can be used as an alarm for attack detection. Figure 4.8 is the result of PCF in each interval of time in both attack and normal traffic.

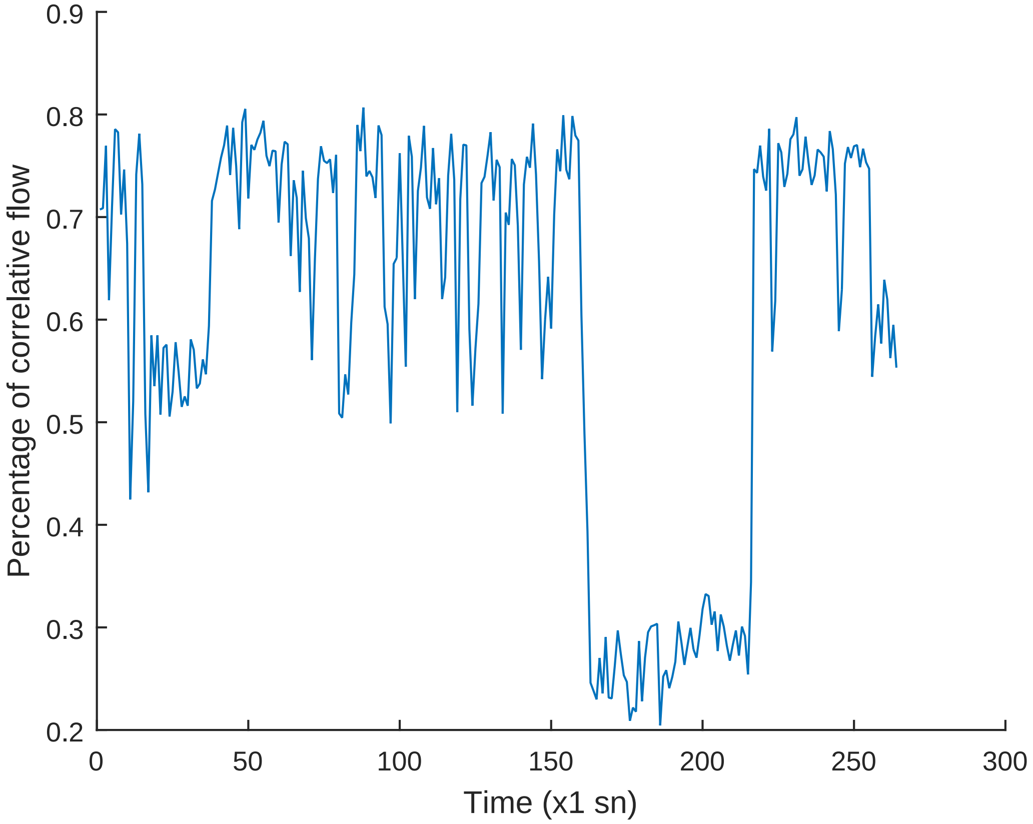


Figure 4.8. PCF result for the attack maintains between 161-215 seconds.

4.2.9. One Direction Flow Generation

One direction flow generating speed (ODGS) is another flow based feature which is defined as;

ODGS = (Number of correlative Flow-Number of Flow)/Time

We use ODGS because it is an indicator of number of one-directional flows generated in an interval of time [30]. As it is obvious in Figure 4.9, this value increases during a DDoS attack.

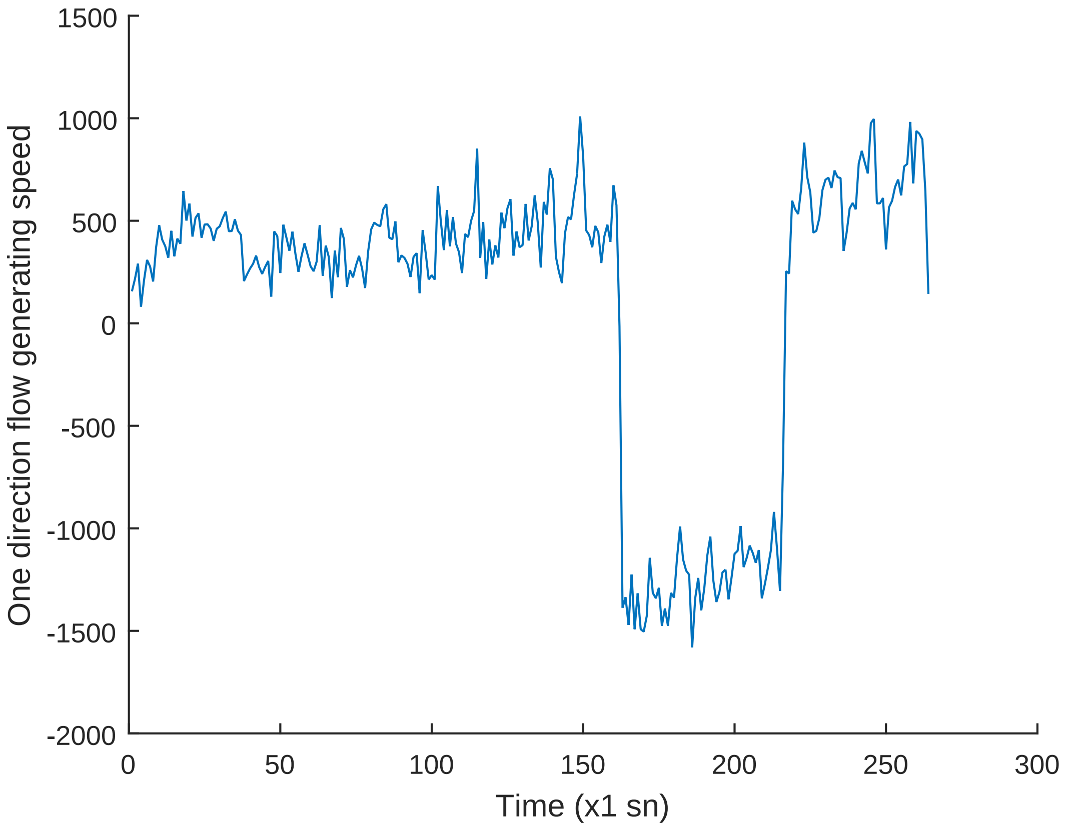


Figure 4.9. ODGS for the attack maintains between 161-215 seconds.

4.2.10. Flow Generating Rate

Flow numbers are another good indicator for DDoS attack. During the attack, it is expected that flow number in an interval will increase, since the attack packets have different and fake IP addresses [31] . Figure 4.10 shows the FGR feature during attack and normal time.

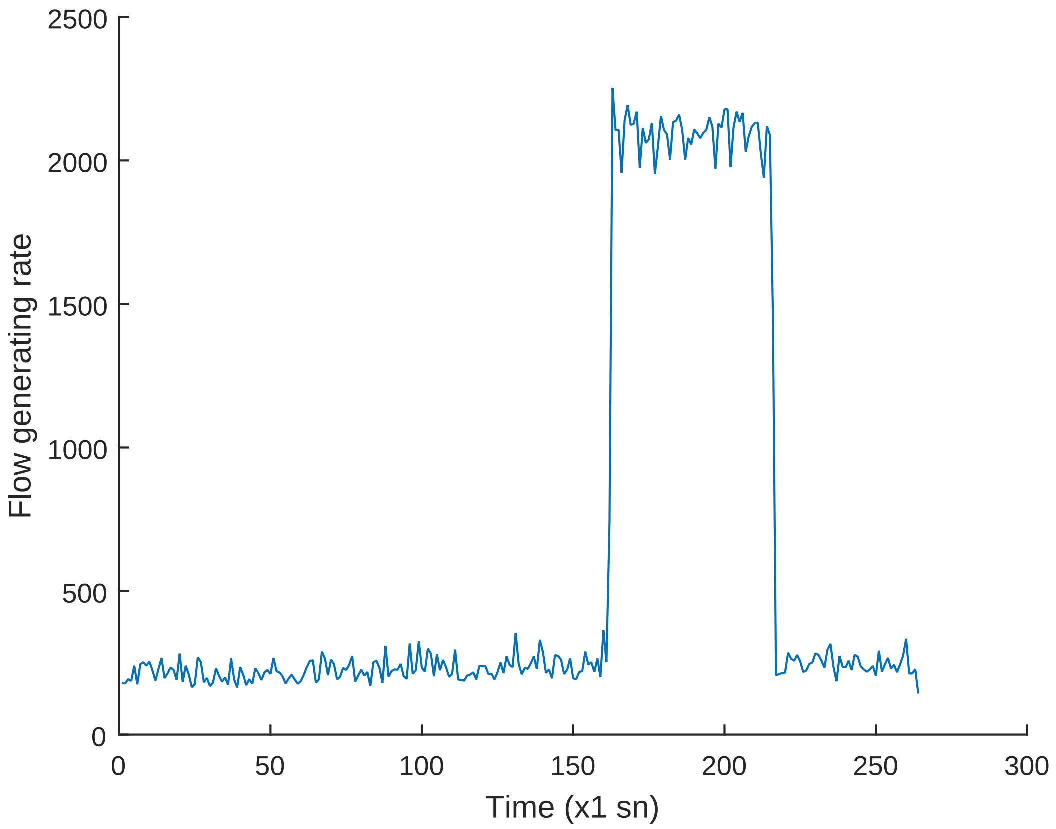


Figure 4.10. Flow generating rate for the attack maintains between 161-215 seconds.

4.2.11. SYN packets/ACK packets ratio

This feature tries to detect incomplete handshakes by monitoring the first and third rounds of each handshake. If either round is missing, it is regarded as an incomplete handshake. We use the first and third rounds because both of them belong to the outgoing traffic. Therefore, this feature which requires only one-way traffic monitoring has the advantage of being flexible and hence can easily be deployed at the source side, the intermediate network or the victim side [32].

5. DESCRIPTIONS of DATASETS and RESULTS

There are so many works in litherature on DDoS attacks. They basically mix legitimate and attack traffics, then they try to detect the differences. This studies are good to see the differences of attack and normal traffics'. However, they are not realistic.

Stealthy DDoS attack sophisticatedly contains low rate of attack packets to pass from detection mechanisms. These attacks are difficult to detect because they send small amount of suspicious packets, comparing with legitimate traffic [33].

We wanted to work on network-based datasets, not victim-based datasets. It means that, we wanted to detect existence of attack by observing the entire network. If only one server is observed, it would be easy to detect the attack, since there will be sharp increase in traffic. If the entire network is observed, there will be only small amount of change which will look like a random moves. Tajer notes that it is very challenging to detect an attack whose packet number has no difference than normal period [34].

Rahmani et al. states that it is really difficult to find a real DDoS dataset. In that work, a dataset which contains only attack traffic was used and it was synthetically injected into legitimate traffic. It is stated that working on traces, those contain only attack, is not useful for complete study of DDoS problem [24].

One of the most challenging issues in DDoS research was obtaining a dataset. Most researchers use synthetic datasets or their own dataset and they don't make the dataset public, due to the privacy reasons. In that case, it is not possible to compare their results with new studies.

In the beginning, we composed a dataset in deterlab simulation tool. The attack was synthetic and detection can be done by using only packet numbers. Then, we obtain a dataset from a private bank, where DDoS attack simulated. However, packet rate was significantly increasing during an attack, so detection was easy. We wanted to work on stealthy DDoS attacks where the packet rate is not increasing too much. This is a case of network-based DDoS detection.

After a long search, we found 3 different public dataset from 2 different institutes and we composed 2 datasets in Bogazici University.

In the results section, instead of giving the each ROC curves, we decided to set threshold false positive rate as 10 % and give the corresponded true positive rate.

5.1. UDP flood dataset UCLA (University of California, Los Angeles)

UCLA Dataset contains only UDP Flood attack. There are legitimate traffic from the department and mix traffic (legitimate and attack) was given. Attack data contains different intensity of attacks. The dataset size is small and it is easy to work with this data. It is in plain text shape, no need any further application to use dataset [35]. For normal traffic, UCLA Computer Science Department traces were used and for attack traffic, constant rate UDP attack was used.

5.1.1. Results for UCLA UDP Flood

UCLA Dataset results were already given above where the features were presented. Almost all features can successfully detect attacks. Only “average packet length” has high false positive rate and “flag type entropy” has small false positive rate. All others can detect attack without any false positive rate. This is synthetic and small dataset. We used this dataset to see how each feature works and change during attacks.

5.2. TCP Syn flood and UDP flood datasets of NUST (National University of Sciences and Technology)

It is a dataset of National University of Science and Technology in Pakistan. There are different types of DDoS attack datasets; TCP Flood, UDP Flood, ICMP Flood and more. Datasets are in pcap form. We used TCP Flood and UDP Flood datasets [36].

For normal traffic, home network dataset was collected in an actual residential setting. The data was collected around 21 hours. Eight different hosts were active by various applications including file transfer, web browsing, instant messaging, real time video streaming etc. Data was collected from each host individually and merged by mergepcap tool. For attack traffic, three end host in the research lab was used. 5 different attack rates launched. 3 low rate attacks (0.1, 1, 10 packets/sec) and 2 high rate attacks (100, 1000 packets/sec) were applied. Average packet number per second was around 500 packet per second for normal situation. Each attack lasted 2 minutes. All attack packets are labeled by setting the reserved bit in the IP header. TCP SYN Flood consists of attacks on two remote servers at ports 143, 22, 138, 137 and 21. Similarly, UDP flood also attacks two remote servers at ports 22, 80, 135, 143. After attack dataset was composed, it was merged with home network dataset by using mergecap tool [37].

TCP dataset contains around 8 million packets and UDP dataset contains around 7 million packets. Both datasets are as plane text in text files. Figure 5.1 shows packet number distribution of NUST TCYP Syn flood and figure 5.2 shows packet number distribution of NUST UDP flood. As discussed above, there are 2 intensive attacks and these are obviously visible in number of attack packet distribution. However, other attacks are very small rate that can not be visible in attack packet distribution.

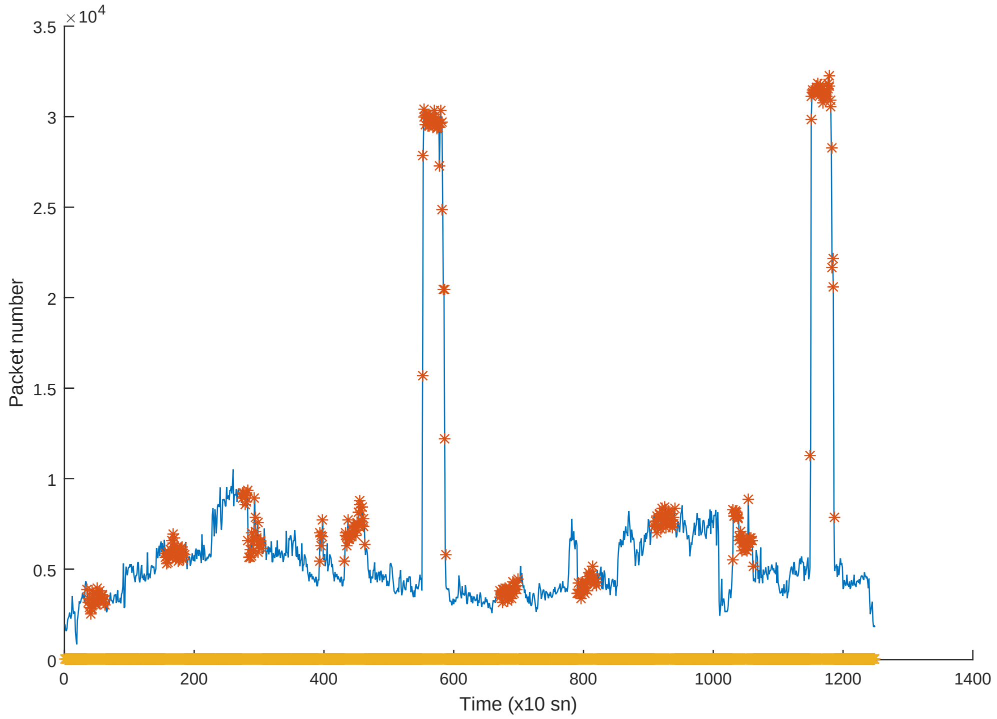


Figure 5.1. NUST TCP Syn flood dataset packet number (stars indicate attack exists in that time interval).

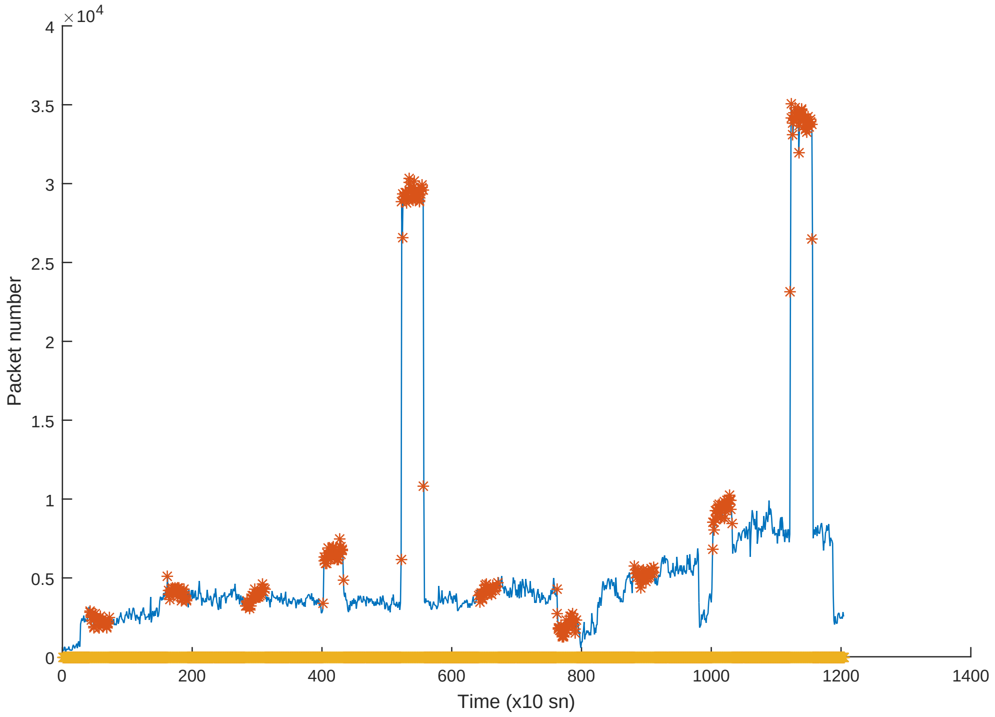


Figure 5.2. NUST UDP flood dataset packet number (stars indicate attack exists in that time interval).

5.2.1. Results for NUST Datasets

This is very challenging dataset. Detection rate is around 50 % for most of the features. In the "Revisiting Traffic Anomaly Detection using Software Defined Networking" article, NUST researchers (the owners of this dataset) present a ROC curve and they have 40 % true positive rates (with respected to 10 % false positive) with 2 features and 90 % detection rate (respected with 10 % false positive) with one feature, which is called "rate limiting" feature [37].

Table 5.1. Detection rates of features for TCP and UDP floods for NUST dataset.

Features	TCP syn flood	UDP flood
Entropy of Source IP	59 %	35 %
KL distance of Source IP	61 %	37 %
JS distance of of Source IP	64 %	38 %
Entropy of destination IP	43 %	44 %
Entropy of TCP flag type	60 %	51 %
Entropy of source port numbers	56 %	42 %
Entropy of destination port numbers	54 %	44 %
Average packet length	52 %	33 %
Average number per flow	48 %	59 %
Percentage of correlative flow	61 %	60 %
One direction flow generation	36 %	48 %
Flow generating rate	55 %	61 %
Syn packets/ACK packets ratio	76 %	12 %

Rate limiting feature is not so reliable because it is assumed that recently connected clients are not dangerous and the system allows them pass without any control. The new clients are dangerous and should be checked. Their dataset is suitable to support this claim, because their attackers are new clients, so that they can detect easily. However, this assumption can cause mistakes, since IP spoofing is very easy nowadays. The attacker can change the packet's source IP as recently connected client and the system would not recognize it as attack.

In our features, there are no such assumptions, they can work under IP spoofing as well. Our results were as shown in table 5.1.

The best detection rate for TCP syn flood attack is SYN packet number/ACK packet number. Because SYN packets and ACK packets are normally not common in a network. Hosts use these packets only for starting a communication. Although there are very few SYN attack packets in this dataset, SYN-ACK ratio can detect 76 % of attacks because of missing ACK numbers.

The best feature to detect UDP flood attack is and flow generating rate (FGR). During UDP attack, the traffic is mostly one way. There is coming in but no going out. Most of the packets have different IP addresses. So, FGR is the best detector for UDP flood attacks.

5.3. TCP Syn flood and UDP flood datasets of Bogazici University

After analyzing online available attacks, we decided to compose our own datasets. This dataset is from Bogazici University where we maintained attack via “Hping” DDoS tool by attacking one victim server inside the campus. The dataset was not recorded in victim side. It was recorded in the one main switch of the campus, so it contains very different types of network traffic in addition to attack traffic. We have SYN Flood and UDP Flood attack datasets.

The attack launched from electrical-electronic engineering department in North Campus. The victim server was located in Information Technologies office in South Campus. And the data trace was recorded between LAN switch and backbone switch, which has over 400 host connections. Average packet per second was around 14000. The topology of flood attack was given in figure 5.3.

We composed 3 different attack scenarios. IP spoofing was used and the destina-

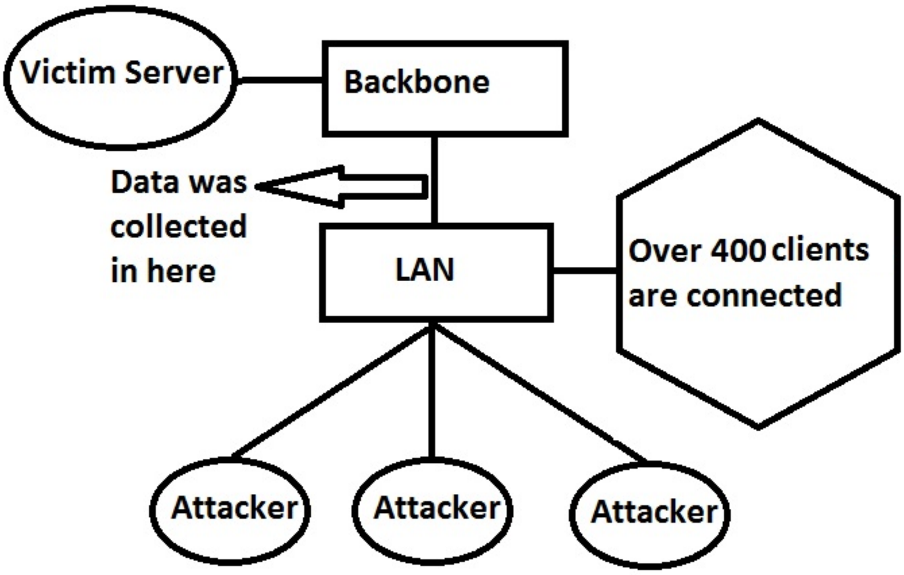


Figure 5.3. Bogazici University DDoS attack dataset topology.

tion port was 80. All of datasets lasted 8 minutes. In each of them, 80 seconds waiting period, then 20 seconds attack period is applied. We attacked by using different packet rates in each attack period. The time schedule can be seen below;

- 80 seconds normal traffic
- 20 seconds attack launched
- 80 seconds normal traffic
- 20 seconds attack launched
- 80 seconds normal traffic
- 20 seconds attack launched
- 80 seconds normal traffic
- 20 seconds attack launched
- 80 seconds normal traffic
- Total: 480 seconds (8 minutes)

First attack was TCP SYN Flood attacks. We attacked by packet rates of 1000, 1500, 2000 and 2500 packets/second. Size of TCP Syn flood attack dataset is around

8 gb.

Then we made 2 UDP Flood attacks for low rate and high rate scenarios. In first UDP Flood scenario, we attacked by using packet rates of 1000, 1500, 2000 and 2500 packets/second, same as TCP SYN Flood attack. However, in the second scenario, we attacked by using packet rates of 2000, 3000, 4000 and 5000 packets/second. Size of first UDP flood dataset is around 6 gb and second dataset is around 8 gb. We used only the first UDP flood attack dataset, since it was low rate attack and hard to detect.

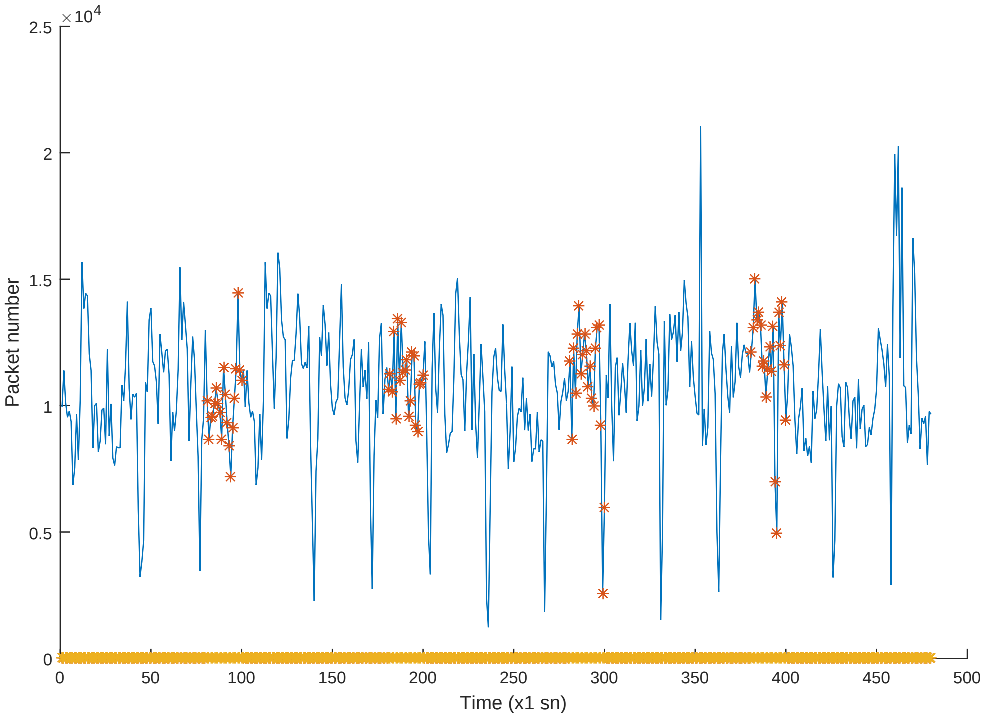


Figure 5.4. BOUN TCP Syn flood dataset packet number (stars indicate attack exists in that time interval).

Figure 5.4 shows the distribution of packet number in BOUN TCP Syn flood attack and figure 5.5 shows the distribution of packet number in BOUN UDP flood.

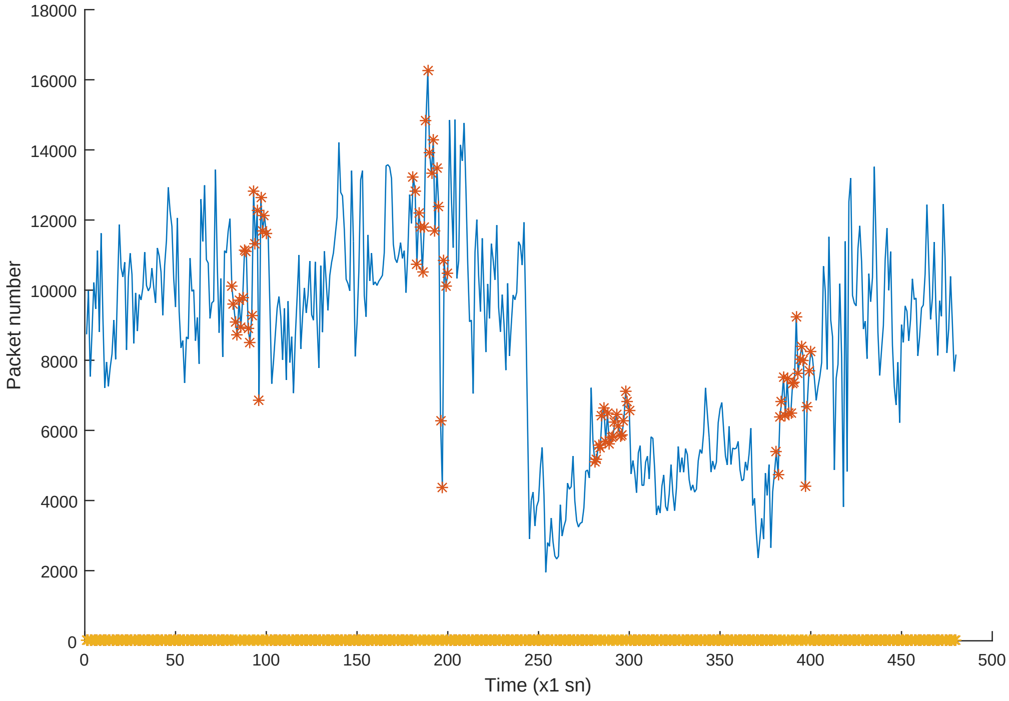


Figure 5.5. BOUN UDP flood dataset packet number (stars indicate attack exists in that time interval).

5.3.1. Results for Bogazici University Datasets

This is very realistic dataset, as described above. Some features have very good results but some of them have poor results on this dataset. We will make this dataset publicly available, to allow other researchers work on it.

Table 5.2 shows the detection rates of features for BOUN TCP syn flood attack and table 5.3 shows detection rates of features for BOUN UDP flood attack.

Table 5.2. Detection rates of features for BOUN TCP syn flood attack.

Features	True Positive	False Positive
Entropy of Source IP	98 %	6 %
KL distance of Source IP	98 %	5 %
JS distance of of Source IP	99 %	4 %
Entropy of destination IP	73 %	10 %
Entropy of TCP flag type	98 %	4 %
Entropy of source port numbers	100 %	8 %
Entropy of destination port numbers	73 %	10 %
Average packet length	92 %	10 %
Average number per flow	100 %	4 %
Percentage of correlative flow	99 %	10 %
One direction flow generation	91 %	10 %
Flow generating rate	98 %	4 %
SYN packets/ACK packets ratio	100 %	3 %

Table 5.3. Detection rates of features for BOUN UDP flood attack.

Features	**True Positive**	**False Positive**
Entropy of Source IP	96 %	8 %
KL distance of Source IP	98 %	6 %
JS distance of of Source IP	99 %	6 %
Entropy of destination IP	28 %	10 %
Entropy of TCP flag type	18 %	10 %
Entropy of source port numbers	93 %	10 %
Entropy of destination port numbers	50 %	10 %
Average packet length	55 %	10 %
Average number per flow	100 %	5 %
Percentage of correlative flow	95 %	10 %
One direction flow generation	49 %	10 %
Flow generating rate	100 %	3 %
SYN packets/ACK packets ratio	12 %	10 %

6. COMPARATIVE STUDIES

The first study to compare with our study is NUST researchers' investigation about DDoS attacks. They composed the dataset and made it public. The dataset is very challenging and their detection rate seems very high with one feature they used. However, a smart attacker can get over this feature easily as explained in section 5.2.1. Their features' names are netad, maxent and ratelimiting. We compared them with our best features for each dataset in table 6.1. [37].

Table 6.1. Compare of NUST dataset detection rates.

Features	**TCP SYN flood**	**UDP flood**
Netad	38 %	40 %
Maxent	35 %	25 %
Rate limiting	92 %	90 %
SYN packets/ACK packets	76 %	12 %
Flow generating rate	55 %	61 %

It should be noted that, there is no common DDoS attack dataset for all researchers to work on. So, comparison of different studies is difficult. In this study, we used attack ratio to make comparison between our studies and other studies. Attack ratio is the ratio between number of attack packets and total number of packets.

Some studies investigate only changes during DDoS attack, they do not contain attack/normal packet ratios or detection rates. So, our second comparison was with other studies those clearly give attack, normal packet numbers and detection rates. Rahmani et al. has 90 % detection rate in 34 % of attack intensity. This is the smallest attack percentage among other 3 studies. In study of Lee et al. [25], their attack ratio is 50 % and their detection ratio is 97 %. Lastly Sharma et al. used a dataset with 50 % attack ratio and their detection rate was 100 % [26].

In table 6.2, number of attack packets, number of legitimate packets, ratio of attack packets over total packets and detection rates can be seen for each studies individually.

Table 6.2. Compare of attack ratios and detection rates.

Author	**Rahmani et al.**	**Lee et al.**	**Sharma et al.**	**This Thesis**
Attack Pck/Sec	50	250	1000	14000
Legitimate Pck/Sec	100	250	1000	1800
Attack/Total Pck	34 %	50 %	50 %	13 %
Detection Rate	90 %	97 %	100 %	95 %

7. CONCLUSION

In this thesis, we firstly focused on finding and composing DDoS attack dataset. Because there is lack of common dataset and each research group use their own dataset to test their system. Then, we described 11 features and we tested each of them in 5 different dataset. Our aim was to find best feature to detect TCP Syn flood attacks and UDP flood attacks. The best feature is "SYN/ACK ratio" for Syn flood attacks in both datasets. And, the best feature to detect UDP flood is "Flow Generating Rate". These features gave the best detection rates in both NUST datasets and BOUN datasets.

Syn flood attacks and UDP floods have different characteristic properties. Therefore, the best feature to detect each of them is different. During Syn flood attack; there will be too many Syn packets but lack of ACK packets because attacker will not send any ACK packet. However, during UDP attack, there is no such a case, since UDP flood does not use any handshake mechanism.

At the end, we compared our detection rates with other studies. It is harder to detect attacks, as the attack ratio goes down. Our attack ratio was smaller than all other studies but our detection rate was still around 95 %, which is very good detection rate for DDoS detection.

Satrya et al. states that, IPv6 solves only lack of IPv4 address problem and we will not need to use NAT (Network Address Translation). However, it still has security issues [38]. Our features are able to use in IPv6 with some modifications in the code. However, we could not obtain any IPv6 DDoS dataset to test or features.

For future work;

- DDoS attacks in IPv6 should be analyzed with detail.
- Researchers should compose new datasets and make it publicly available.
- Datasets with less attack ratio should be investigated.

REFERENCES

1. “Internet Users”, *internetlivestats.com/internet-users*, 2015.

2. Mirkovic, J., J. Martin and P. Reiher., “A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms”, *ACM SIGCOMM Computer Communication Review*, 2004.

3. “Mafia Boy”, *wikipedia*, n.d., `https://en.wikipedia.org/wiki/MafiaBoy`.

4. Liu., X., “Mitigating Denial-of-Service Flooding Attacks with Source Authentication”, *PhD Thesis in Department of Computer Science in the Graduate School of Duke University*, 2012.

5. Meek, A., “DDoS attacks are getting much more powerful and the Pentagon is scrambling for solutions”, *BGR*, 2015.

6. Kaspersky, “Kaspersky DDoS Intelligence Report Q3 2015”, *Kaspersky Lab*, 2015.

7. Hang, B., R. Hu and W. Shi, “An Enhanced SYN Cookie Defence Method for TCP DDoS Attack”, *Journal Of Network*, Vol. 6, No. 8, 2011.

8. Miller, L. C., *DDoS For Dummies*, John Wiley & Sons, Inc., New Jersey, NJ, USA, 2012.

9. Zargar, S., J.Joshi and D. Tipper, “A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks”, *Communications Surveys Tutorials*, 2013.

10. Maregeli, C., “A Study On TCP–SYN Attacks And Their Effects on A Network Infrastructure”, *Faculty of Electrical Engineering, Delft University*, 2010.

11. Zhang, G. and M. Parashar, “Cooperative Defence against DDoS Attacks”, *Journal of Research and Practice in Information Technology*, 2006.

12. Sachdeva, M., G. Singh, K. Kumar and K. Singh, “DDoS Incidents and their Impact: A Review”, *The International Arab Journal of Information Technology*, Vol. 7, 2012.

13. Poongothai, M. and M. Sathyakala, “Simulation and Analysis of DDoS Attacks”, *International Conference on Emerging Trends in Science, Engineering and Technology*, 2012.

14. Cisco, “Internet Protocols”, *http://docwiki.cisco.com/wiki/internetprotocols*, 2015.

15. Rui, X., M. Wen-li and Z. Wen-ling, “SYN Flooding Detecting using Negative Selection Algorithm based on Eigenvalue Sets”, *E-Business and Information System Security*, 2009.

16. Rouse, M., “UDP (User Datagram Protocol) definition”, *http://searchsoa.techtarget.com/definition/UDP*, 2015.

17. Rui, X., M. Wen-li and Z. Wen-ling, “Defending Against UDP Flooding by Negative Selection Algorithm based on Eigenvalue Sets”, *International Conference on Information Assurance and Security*, 2009.

18. Sun, C., J. Fan and B. Liu, “A Robust Scheme to Detect SYN Flooding Attacks”, *Communications and Networking in China*, 2007.

19. Nashat, D., “Router based detection for low-rate agents of DDoS attack”, *International Conference on High Performance Switching and Routing*, 2008.

20. Singh, J., M. Sachdeva and K. Kumar, “Detection Of DDoS Attacks Using Source Ip Based Entropy”, *International Journal of Computer Science Engineering and Information Technology Research*, Vol. 3, 2013.

21. Bhuyan, M. H., D. K. Bhattacharyya and J. K. Kalita, “Information Metrics for Low-rate DDoS Attack Detection : A Comparative Evaluation”, *Contemporary*

Computing (IC3) International Conference, 2014.

22. Xu, T., D. He and Y. Luo, "DDoS Attack Detection Based on RLT Features", *International Conference on Computational Intelligence and Security*, 2007.

23. Wang, D., Z. Yufu and L. Lie, "A Multi-core Based DDoS Detection Method", *Computer Science and Information Technology*, Vol. 4, 2010.

24. Rahmani, H., N. Sahli and F. Kamoun, "DDoS flooding attack detection scheme based on F-divergence", *Computer Communications*, Vol. 35, 2012.

25. Lee, S. M. and D. S. Kim, "Detection of DDoS attacks using optimized traffic matrix", *Computers and Mathematics with Applications*, 2012.

26. Sharma, S., S. Sahu and S. Jena, "On Selection of Attributes for Entropy Based Detection of DDoS", *Advances in Computing, Communications and Informatics*, 2015.

27. Shi, H., "Adaptive Packet Context-Constrained KL Divergence Model for Intrusion Detection", *Journal Of Networks*, Vol. 9, 2014.

28. Salem, O., F. N. Abdesselam and A. Mehaoua, "Anomaly Detection in Network Traffic using Jensen-Shannon Divergence", *Wireless Networks Symposium*, 2012.

29. Oo, T. T. and T. Phyu, "Analysis of DDoS Detection System based on Anomaly Detection System", *International Conference on Advances in Engineering and Technology*, 2014.

30. Guo, R., H. Yin, D. Wang and B. Zhang, "Research the Active DDoS Filtering Algorithm Based on IP Flow", *Int. J. Communications, Network and System Sciences*, Vol. 7, 2009.

31. Qian, L. and B. E. Carpenter, "A Flow-Based Performance Analysis of TCP and TCP Applications", *International Conference on Computer and Network Technol-*

ogy, Vol. 14, 2011.

32. Hu, C. H. and Y. Zhou, “A More Accurate Scheme to Detect SYN Flood Attacks”, *IEEE INFOCOM Workshops*, 2009.

33. Liu, H. and M. Kim, “Real-Time Detection of Stealthy DDoS Attacks Using Time-Series Decomposition”, *Communications (ICC), 2010 IEEE International Conference*, 2010.

34. Tajer, J., “Detecting Flooding Attacks Using Power Divergence”, *Kaspersky Academy IT Security for the Next Generation*, 2012.

35. “University of California Trace Format”, *lasr.cs.ucla.edu/ddos/traces/*, 2015.

36. “Network Embedded Security Using In-Network Packet Marking”, *wisnet.seecs.nust.edu.pk/projects/nes/datasets.html*, 2015.

37. Mehdi, S., J. Khalid and S. Khayam, “Revisiting Traffic Anomaly Detection using Software Defined Networking”, *International Symposium on Recent Advances in Intrusion Detection (RAID)*, 2011.

38. Satrya, G. B. and R. L. Chandra, “The Detection of DDOS Flooding Attack Using Hybrid Analysis in IPv6 Networks”, *International Conference on Information and Communication Technology*, 2015.