ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, No 5, May 2013

A Statistical Approach to Classify and Identify DDoS Attacks using UCLA Dataset

Thwe Thwe Oo, Thandar Phyu

Abstract— **Nowadays**, Internet is the most well-known and popular thing that is widely used by human beings. It is also an essential part of human life and provides the best and fast communication medium. As many people widely used Internet, i.e., so many user of Internet is increasing, network security attacks are also increasing. Among these security attacks, DDoS (Distributed Denial of Service) is the most serious attack for network security. This attack is not direct attack because it does not enter directly to the system and does not damage it. In this paper, the two proposed algorithms can be classified and identified what types of DDoS attacks by using UCLA data set. At first, packet classification algorithm classifies normal and attack from incoming packets. To get more accurate result, K-NN classifier estimates normal or attacks from results of packet classification algorithm. Finally, the proposed algorithm classifies and identifies types of DDoS attacks.

Index Terms— DDoS, UCLA, network security

I. INTRODUCTION

Many people widely used Internet in all over the world. As the numbers of Internet users are increasing with new and developed services, many security attack threats have become popular. Some serious attacks expose and exploit in many security vulnerabilities. Recently, report for network security breaches indicate that the impact and the damaged costs of attacks are continuously increasing.

The most popular recent network attack trend is the use of network traffic that is flooding attack. An attacker illegally places networks or hosts, without intruding into the hosts. These types of attacks such as Denial of Service(DoS) and Distributed Denial of Service(DDoS) attacks are the serious attacks that can cause the more damage than the other attacks [1,3,6]. A skillful DoS/ DDoS attack exploit the system quickly and it is difficult to trace back the intruder.

This paper focuses on DDoS attack detection and classification method, especially for scanning and flooding attacks. The proposed system analyzes network traffic based on statistical approach by using UCLA data set[7]. The proposed system characterizes field values (packet count, average of packet size, time-interval variance, packet-size variance, and so on).

This paper is organized as follows: Section II introduces some previous researches corresponding to DDoS attacks. Section III presents architecture of DDoS attack and types of DDoS attacks are presented in Section IV. Then the proposed system is presented in Section V and and results in experimental evaluation describes in Section VI and finally the conclusion of this paper.

II. RELATED WORK

In this section, some previous approaches to detect DDoS attacks are discussed. In [5] the proposed system is a statistical approach method based on various features of attacks packets, obtained from study from incoming network traffic and using of Radial Basic Function(RBF) Neural Network to analyze these features. This proposed system can be classified either normal or attack, but that can't be classified and identified what types of attacks.

In[4] this paper presents an abnormal network detecting method and a system prototype and suggests a detection algorithm using changes in traffic patterns that appeared during an attack. This proposed system can identify attacks but that cannot be detected by examining only single packet information.

In[2] this proposed system introduces a method for proactive detection of DDoS attacks by classifying the network status. Then, it investigates the procedure of DDoS attacks and selects variables based on these features and finally, apply the K-Nearest Neighbour(K-NN) method to classify the network status into each phase of DDoS attack.

III. ARCHITECTURE OF DDOS ATTACK

DDoS attack, mainly includes finding hosts, making communication to hosts, compromise hosts and launching attack. The following steps are the detail process of DDoS attack:

(1) Before starting DDoS attack, the attacker searches the hosts in the network which hosts have security vulnerability and weakness. Then the attacker intrudes these vulnerable hosts and they get administration authority to install control programs.

(2) The attacker gives control instruction/command to the handlers that causes the handlers do orders to the agents. Generally, one or more handlers take control the agents.

(3) The agents continuously or periodically send a huge numbers of useless packets to the victim. When the victim receives these huge amounts of packets, they cannot respond

for the entire incoming request. The other normal requests are not able to receive and corresponding reply to them because of the congested network traffic and the victim system crash and slow down.

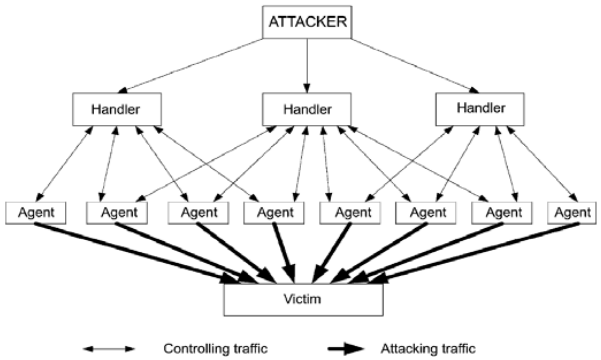

Fig.1.Architecture of DDoS Attack

IV. Types of DDoS Attacks

In this paper, the proposed system emphasizes two types of DDoS attacks; flooding and scanning attacks. There are many types of flooding and scanning attacks such as SYN Flooding attack, ACK Flooding, SYN Scan and ACK Scan, etc.

In a TCP SYN flooding attack, an attacker sends many SYN messages, with spoofed IP addresses, to a single server victim. Although the server replies with SYN/ACK messages, these messages are never acknowledged by the client. As a result, many halfopen connections exist on the server, consuming its resources.This continues until the server has consumed all its resources, hence can no longer accept new TCP connection requests.

ACK Flood attacks use a large number of ACK packets to attack the victims, with all TCP messages being with ACK flag bits. When the host receives a packet with ACK flag bits, the existence of the four-tuple connection expressed by the packet needs to be checked. If the four-tuple connection exists, the host checks whether the state represented by the packet is legal, and then the packet can be passed to the application layer. If the packet is found to be illegal during the inspection (e.g. if the packet's targeted port does not open on the machine) then the host's operating system protocol stack will respond with a RST packet, telling the other side that this port does not exist.

In SYN Scan, send a SYN packet to the targeted destination port. If a host replies with a RST or does not reply, the port is closed. If a host replies with a ACK, close the connection by RSTand identifies open ports.

In ACK Scan, send an ACK packet to the targeted destination port. If a host replies with an RST, a port is most likely open. If a host does not reply but an ICMP destination unreachable packet is received, the port is filtered. Otherwise the port is most likely closed and identifies potentially open and filtered ports.

V. Proposed System

The proposed system involves the following steps:

(1) Collection of packets

(2) Features Extraction

(3) Attack classification by using Packet Classification Algorithm

(4) Estimate using K-NN Classifier

(5) DDoS attack classification

A. *Collection of Packets*

The proposed system collects the incoming packets in every one second. For collecting packets in one second, it does not impact to the classification accuracy. The use of longer flow timeouts is possible; however, there is longer classification response.

B. *Features Extraction*

Feature extraction calculates the selected features for captured packets. These features are proposed by observing the characteristics of DDoS attack packets. These features can be used to recognize and classify incoming attack packets. The proposed features explain as follows:

- Number of packets: Total number of packets from source IP to destination IP. During an attack, the attacker sends a large number of packets to the victim system.
- Number of bytes: Total number of bytes sent from source IP to destination IP. It increases when launching DDoS attack.
- Average packet size: The ratio of number of bytes to number of packets. It increases in attack time.
- Packet rate: Rate of packet per second. For packet rate calculation:

  Packet rate per second = $n_p \times \frac{1}{(t_e - t_s)}$

  n_p = number of packets
  t_e = end packet sent time
  t_s = start packet sent time
- Byte rate: Rate of packets byte per second. For byte rate calculation:

  Byte rate per second = $b_t \times \frac{1}{(t_e - t_s)}$

  b_t = total number of bytes
  t_e = end packet sent time
  t_s = start packet sent time

- Time-interval variance: The attacker sends attack packets in the same time span while launching DDoS attack, so time interval variance will be closer to zero.

  For time-interval variance calculation:

  (i) First, calculating the means:

$$\bar{t} = \frac{\sum t_i}{i} \quad , \ i = 1,2,3,\ldots$$

  (ii) Second, squaring deviation of the means

$$t_c^2 = \frac{\sum \left(t_n - \bar{t}\right)^2}{n}$$

  , c = collection number(c1,c2,c3,…)

  (iii)Third, final calculation of time-interval-variance

$$t_c = \sqrt{\frac{\sum (t_n - \bar{t})^2}{n}}$$

- Packet-size variance: Although normal packets have different packet sizes, attack packet size are the same. So packet-size-variance will be closer to zero.

  For packet size calculation:

  (i) First, calculating the means:

$$\bar{p} = \frac{\sum p_i}{i} \quad , \ i = 1,2,3,\ldots$$

  (ii) Second, squaring deviation of the means

$$p_c^2 = \frac{\sum (p_n - \bar{p})^2}{n}$$

  (iii)Third, final calculation of packet-size-variance

$$p_c = \sqrt{\frac{\sum (p_n - \bar{p})^2}{n}}$$

C. Attack classification by using Packet Classification Algorithm

The algorithm checks the field values of the packet. The diagram in Figure (2) classifies attacks by the field values of the packets.

No: of pkts	Avg pkt size	Time-Inter-rv-al vari-an-ce	Packet size variance	Num-ber of byt-es	Pac-ket rate per sec	Bytes Rate per second	No: of Fla-g pac-ket s	Clas-s
L	L	>0	L	L	<α	L	L	Nor-mal
H	H	<0	<0	H	>λ	H	H	Atta-ck

D.Estimate using K-NN Classifier

It is possible that one field value of the incoming packet may miss with the field value of the algorithm. For example, in normal case, packet size variance value is not (>0) although other field values match to the algorithm. If this case occurs, it cannot be sure that is normal or attack. When this case occurs, the proposed system identifies that this case is “Other” case which is not normal or attacks. To identify “Other” case as normal or attack, use the K-NN classifier to estimate normal or attack.

There are many well-known methods for classifying documents such as SVM, NN, fuzzy logic, and rough set .We choose the k-NN method because this method has features that are suitable for proposed system.These features are: easy implementation, short time computation, and high accuracy.The k-NN algorithm is a similarity-based learning algorithm and is known to be highly effective in various problem domains, including classification problems. Given a test element dt, the k-NN algorithm finds its k nearest neighbors among the training elements, which form the neighborhood of dt.

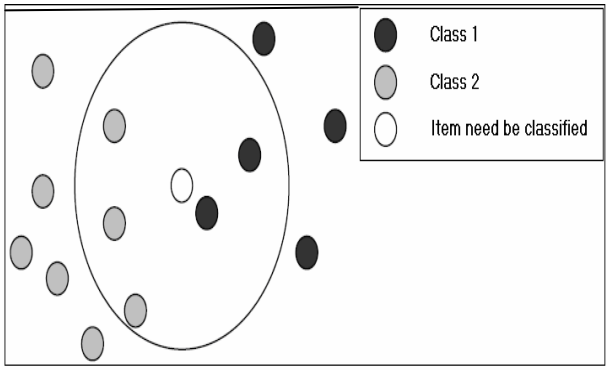

Fig.2.K-NN Classifier

E.DDoS Attack Classification

Finally, the proposed algorithm classifies and identifies what types of attacks.

- IF (no-packets==HIGH) AND (avg-packet-size with same Destination IP==HIGH) AND (protocol==UDP)
  THEN UDP Flooding
- IF (no-packets==HIGH) AND (avg-packet-size with same Destination IP==HIGH) AND (protocol==TCP)
  Begin
  IF(no-ACK with same destination port==HIGH)
  THEN TCP ACK Flooding
  IF (no-SYN with same destination port==HIGH)
  THEN TCP SYN Flooding
  End

- IF(no-packets==HIGH) AND (no-destination-port==HIGH) AND (no-source-IP== LOW)

  Begin

  IF (no-ACK with same destination port==HIGH)

  THEN TCP ACK Scanning

  IF (no-SYN with same destination port==HIGH)

  THEN TCP SYN Scanning

  IF (no-FIN with same destination port==HIGH)

  THEN TCP FIN Scanning

  End

- IF (no-packets==HIGH) AND (no-destination-IP==HIGH) AND (no-destination-port==LOW)

  THEN Network Scanning Attack

VI.EXPERIMENTAL EVALUATION

In this section, the proposed DDoS attack detection is described by the evaluation result. The proposed system classifies DDoS attacks by using UCLA Dataset. Table I shows mathematical calculation for features extraction.

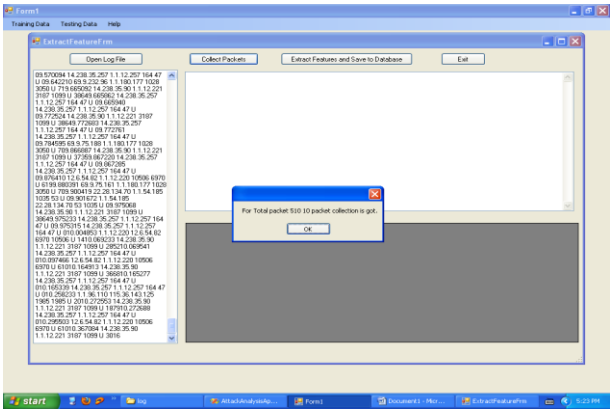

Fig.3. Collection of packet from raw log file

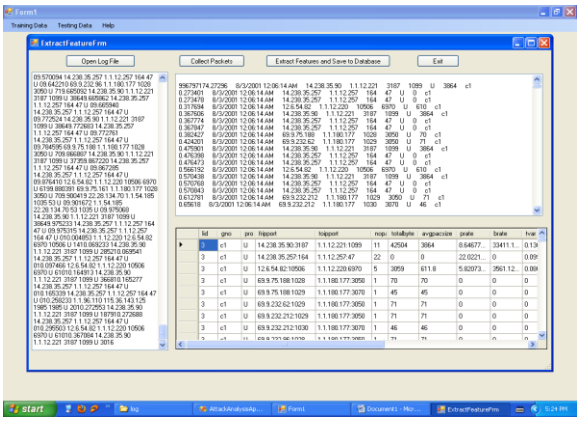

Fig.4. Collection of packets for one second and store into database

VI. CONCLUSIONS

The most serious attack for network security is DDoS (Distributed Denial of Service) attack.The more the rate of the internet usage increases, the more challenge increase for efficient DDos detection system.So there are many challenges for detecting and classifying DDoS attacks. In the proposed system, DDoS attack packets are studied from UCLA data set and extract the features to analyze and classify DDoS attack. The proposed algorithm and packet classification algorithm will be efficient and suitable for classifying DDoS flooding and scanning attacks.

1. Drew Dean, Matt Franklin, and Adam Stubblefield, “An algebraic approach to ip traceback,” Proc. of Network and Distributed System Security Symposium, NDSS '01, San Diego, California, February 2001.
2. Hoai-Vu Nguyen and Yongsun Choi, “Proactive Detection of DDoS Attacks Utilizing k-NN Classifier in an Anti-DDos Framework”, International Journal of Electrical and Electronics Engineering 4:4 2010
3. L. John Ioannidis and Steven M. Bellovin, “Implementing pushback: Routerbased
defense against DDoS att acks,” Proc. of Network and Distributed System Security Symposium, NDSwsS ’02, San Diego,California, February 2002.
4. Myung-Sup Kim, Hun-Jeong Kang, Seong-Cheol Hong, Seung-Hwa Chung, and James W. Hong, Dept. of Computer Science and Engineering, POSTECH, “A Flow-based Method for Abnormal Network Traffic Detection”.
5. Reyhaneh Karimazad and Ahmad Faraahi, “An Anomaly-Based Method for DDoS Attacks Detection using RBF Neural Networks”, 2011 International Conference on Network and Electronics Engineering , IPCSIT vol.11 (2011) © (2011) IACSIT Press, Singapore.
6. Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson, “Practical network support for IP traceback”, Proc.of the 2000 ACM SIGCOMM, Stockholm, Sweden, August 2000.
7. UCLA CSD packet traces. http://www.lasr.cs.ucla.edu/ddos/traces/public/usc.

No	Source IP	Destination IP	Calculated Features						
			Number of packets	Average of packet size	Time-interval variance	Packet-size-variance	Number of bytes	Packet rate	Byte rate
1	14.238.35.257	1.1.12.257	5	27	1.58	3.9	94	0.61	154.89
2	69.9.75.100	1.1.180.177	17	69	3.98	69.88	1140	1.27	682.79
3	12.6.54.82	1.1.12.220	1355	970	0.001	0.02	1636585	34.73	276529.58
4	69.9.232.212	1.1.180.177	1504	994	0.001	0.01	1599453	39.05	286958.36

TABLE I .MATHEMATICAL CALCULATION FOR FEATURES EXTRACTION