DDoS Attacks Impact on Network Traffic and its Detection Approach

[1]Anup Bhange, [2]Amber Syad, [3]Satyendra Singh Thakur
[1]M.tech Scholar, Dept of CSE
[2] Asst Prof, Dept of CSE
[3]Asst Prof, Dept of CSE
[1,2] Patel Institute of Technology, Bhopal
[3]Patel college of Science Technology, Bhopal

ABSTARCT

A Denial of Service (DoS) attack is a malicious effort to keep endorsed users of a website or web service from accessing it, or limiting their ability to do so. A Distributed Denial of Service (DDoS) attack is a type of DoS attack in which many computers are used to cripple a web page, website or web-based service. Fault either in users' implementation of a network or in the standard specification of protocols has resulted in gaps that allow various kinds of network attack to be launched of the type of network attacks, denial-of-service flood attacks have reason the most severe impact. This analysis study on flood attacks and Flash Crowd their improvement, classifying such attacks as either high-rate flood or low-rate flood. Finally, the attacks are appraised against principle related to their characteristics, technique and collision. This paper discusses a statistical approach to analysis the distribution of network traffic to recognize the normal network traffic behavior. The EM algorism is discussed to approximate the distribution parameter of Gaussian mixture distribution model. Another time series analysis method is studied. This paper also discusses a method to recognize anomalies in network traffic, based on a non restricted α-stable first-order model and statistical hypothesis testing.

Keywords:- DDoS Impact, Anomaly Detection Method, α-Stable Model

1. INTRODUCTION

Distributed denial-of-service attacks (DDoS) pose an immense threat to the Internet, and consequently many defense mechanisms have been proposed to combat them. Attackers constantly modify their tools to bypass these security systems, and researchers in turn modify their approaches to handle new attacks. The DDoS field is evolving quickly, and it is becoming increasingly hard to grasp a global view of the problem. This paper strives to introduce some structure to the DDoS field by developing a taxonomy of DDoS attacks and DDoS defense systems. The goal of the paper is to highlight the important features of both attack and security mechanisms and stimulate discussions that might lead to a better understanding of the DDoS problem.

A Denial of Service attack is an attempt by a person or a group of persons to cripple an online service. This can have serious consequences, especially for companies like Amazon and eBay which rely on their online availability to do business. In the not so distant past there have been some large scale attacks targeting high profile internet sites [28, 29, 30, and 31]. Consequently, there are currently a lot of efforts being made to come up with mechanisms to detect and mitigate such attacks. Even though the first denial of service attacks did not take place a long time ago (tools that automate setting up of an attack network and launching of attacks, started appearing in 1998), there are a multitude of denial of service attacks that have been used. Broadly speaking the attacks can be of three forms. a) Attacks exploiting some vulnerability or implementation bug in the software implementation of a service to bring that down. b) Attacks that use up all the available resources at the target machine. c) Attacks that consume all the bandwidth available to the victim machine. The third type of attacks is called bandwidth attacks. A distributed framework becomes especially suited for such attacks as a reasonable amount of data directed from a number of hosts can generate a lot of traffic at and near the target machine, clogging all the routes to the victim. Protection against such large scale distributed bandwidth attacks is one of the most difficult (and urgent) problem to address in today's internet. CERT reports bandwidth attacks as increasingly being the most common form of Denial of Service attacks seen in the internet today.

2. DDoS ATTACK OVERVIEW

A denial-of-service attack is distinguish by an explicit attempt by attackers to prevent legitimate users of a service from using that service [1]. A distributed denial-of-service attack organizes many machines to attain this goal. The service is denied by sending a stream of packets to a victim that either consumes some key resource, thus rendering it unavailable to legitimate clients, or provides the attacker with unlimited access to the victim machine so he can inflict arbitrary damage.

The operating systems and network protocols are developed without concern security engineering which results in providing hackers a lot of insecure machines on Internet. These insecure and unmatched machines are used by DDoS attackers as their army to launch attack. An attacker or hacker regularly inserts attack programs on these insecure machines. Depending upon complexity in logic of fixed programs these compromise machines are called Masters/Handlers or Zombies and are collectively called bots and the attack network is called botnet in hacker's society. Hackers send control instructions to masters, which in turn communicate it to zombies for launching attack.

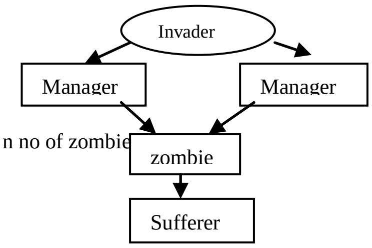

Fig 1. Attack modul

DDOS attacks classified into two broad categories: flooding attacks and Flash Crowd attacks. Flooding DDoS attacks guzzle resources such as network bandwidth by overwhelming bottleneck link with a high volume of packets. Flash Crowd attacks use the predictable behavior of protocols such as TCP and HTTP to the attacker's advantage. The computational resources of the server are tied up by seemingly legitimate requests of the attackers and thus check the server from processing transactions or requests from authorized users.

3. DDOS COLLISION

Distributed Denial-of-service (DoS) attacks have been in the region of for a long time. In the computer network arena, DDoS attacks usually take one of two forms [4]: (a) exploiting bugs in network clients or server applications, in an attempt to crash the application (and possibly the host on which it is running) or (b) Flooding a network server with fake traffic, making it difficult or impossible for the server to receive and process legitimate traffic. The former are typically carried out by using 'buffer overrun attacks' in which a network application is sent a large amount of data which it fails to handle properly, instead overwriting critical information with the excess data.

Some companies do not take security gravely enough and their systems, in general, are easily finding the middle ground and pose a threat not only to the companies themselves but also to anyone else under attack by a hacker through their systems. Defending against DoS attacks involves using secure operating systems such as UNIX which over process protection (to thwart an application crash from crash- in the whole system), keeping up-to-date with security patches and susceptibility alerts (to avoid successively applications which are susceptible to buffer overrun at- tacks) and monitoring and controlling network traffic (to handle flood attacks). DDoS attacks are a new variation on this old theme. A DDoS attack uses net- work flooding, but is harder to defend against because the attack is launched from hundreds or even thousands of hosts concurrently. Rather than appearing as an excess of traffic coming from a single host, a DDoS attack appears in its place as normal traffic coming from a large number of hosts. This makes it harder to identify and control. Even when an attack has been recognized, it can be difficult if not possible to trace back to its genesis as so many compromised hosts are involved. When there are so many hosts implicated, the logistical problems of stemming the attack and recognize its real origin are enormous The Internet was simply not calculated with these susceptibility in mind, and a real solution would involve re-engineering the entire network architecture. This means it is critical to take pre-emptive events to reduce the possibility of these attacks and minimize their collision.

4. INTERNET ARACHITECTURE

The Internet was planned with functionality, not security, in mind, and it was surely very successful in reaching this goal. It offers its accomplice fast, easy and cheap communication mechanisms, enforced with various higher-level protocols that ensure reliable or timely delivery of messages or a certain level of quality of service. Internet design follows the end-to-end paradigm: communicating end hosts deploy complex functionalities to achieve desired service guarantees, while the middle network provides the bare-minimum, best-effort service. The Internet is managed in a 2 distributed manner; therefore no common policy can be enforced among its member. Such design opens several security topics that provide opportunities for distributed denial-of-service attacks:

1. Internet security is extremely mutually supporting. DDoS attacks are usually launched from systems that are weaken through security related compromise. in spite of of how well protected the victim system may be, its susceptibility to DDoS attacks depends on the state of security in the rest of the global Internet [32].

2. Internet resources are limited. Each Internet host has limited resources that can be consumed by a sufficient number of users.

3. Power of many is greater than power of few.

Coordinated and simultaneous malicious actions by some participants can always be detrimental to others, if the resources of the attackers are greater than the resources of the victims.

4. Intelligence and resources are not collocated.

An end-to-end communication example led to locating most of the intelligence wanted for service guarantees with end hosts. At the same time, a wish for large throughput led to the design of high bandwidth pathways in the transitional network. Thus, malicious clients can misuse the plentiful resources of unwitting network for delivery of numerous messages to a victim. Difficulty in tracing back the attack to the source Most (if not all) of the internet runs on top of the TCP/IP protocol. The underlying protocol (IP) is basically connectionless in environment. At every transitional step from the source to the destination, the decision about the next host to forward the packet is finished. All such routing decisions are made on the basis of the destination address. It is thus possible to generate packets with incorrect source IP addresses and use them to launch Denial of Service attacks. This technique is known as IP spoofing. Users with sufficient rights on a system have the ability to make such fake packets.

5. Limited Resources: The infrastructure of the interconnected hosts and networks is including of limited resources. Bandwidth dealing out power and storage capacities is all targets of Denial of Service attacks. If these resources are enlarged by substantial investments, it just raises the bar on the degree an attack must reach to be successful.

5. ANOMALY FINDING METHOD

5.1. Statistical Approach for Network Anomaly Detection*:*

In statistical-based approach include a normal network act and then all traffic that deviates from the normal is noticeable as anomalous. This method is used to study network traffic prototype on an exacting network. By examine network traffic and processing the information with complex statistical algorithms, this systems look for anomalies in the known normal network traffic patterns. All packets are given an anomaly score and if the anomaly score is higher than a certain threshold, the intrusion detection system will generate an alert.

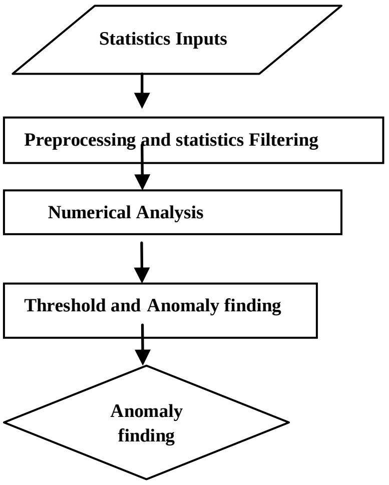
Statistics Inputs
Preprocessing and statistics Filtering
Numerical Analysis
Threshold and Anomaly finding
Anomaly finding

Fig. 2 Statistical Approach for Network Anomaly finding

This approach has a number of advantages [19]. It is capable of sense new unseen attacks like denial of services attacks, worm or virus. It is also capable of sense low intensity slow pace attacks. Another major advantage of this method is that it is potentially easier to sustain than a rule based approach because do not need to uphold and update any record of signature. The basic difficulty with this type of approach is the collection of appropriate threshold value. Problem of false positive and false negative arise due to this value. If value is put low than ratio of false positive increase if value is set too high than the anomalous activities cannot be verify means false negative add to.

5.2 Gaussian Mixture Model

Examine the source data, the network traffic cannot be explained as a Gaussian sharing. The sharing of a Gaussian should be like the shape of ellipse and its residual should be normal. The Gaussian mixture model probability density function is a weighted average of several Gaussian sharing. Here it is taken the Gaussian mixture model with three single Gaussian distribution as an example.

$q(x) = \alpha 1\ g(x; \mu 1, \theta 1) + \alpha 2$
$k(x; \mu 2, \theta 2) + \alpha 3\ g(x; \mu 3, \theta 3)$

The parameter list ($\alpha 1, \alpha 2, \alpha 3$) must satisfy the following condition:

$\alpha 1 + \alpha 2 + \alpha 3 = 1$

The single Gaussian mixture distribution can be represented as:

$q(x; \mu, \sigma 2) = (2\pi)^{-d/2} \sigma^{-d} \exp[-((x - \mu)^T (x - \mu))/2\, \sigma^2]$

The more Gaussian models, the more exact the Gaussian mixture model will be. In the approach argue here it discover the amount of Gaussian distribution will authority the time cost and performance of advance.

5.3 EM Algorithm

EM is an iterative technique for guess the value of some anonymous quantity, given the values of some linked, identified quantity. The process is to first consider that the quantity is instead of as a value in some parameterized probability allocation. The EM procedure is discussed:

Initialize the sharing constraint Repeat until union:

N-Step: estimated the predictable value of the unknown variables, given the current parameter estimate

$P(h`|h) = N[\ln v(V\ |l`)|l, V]$

N-Step: re-estimate the sharing parameters to maximize the similarity of the data, given the predictable estimates of the unknown variables

$L \leftarrow \arg\max P(h`|h)$

At here, the EM algorism is used to guess the mean value of different Gaussian sharing which overlaps with each other to form the Gaussian mixture sharing.

5.4 TIME SLICE WINDOW

The method is the combination of Gaussian model is measured to match the network traffic sharing. Then the EM algorism is used to calculate the mean value of each Gaussian distribution. Considering the data of time series, the data should be divider with time slot. It is called the “window”. The size of the windows should be decided. From the input data the network traffic illustrate the circular fluctuation of date and night. So the time slot window should be the integral times of the 24 hours. In input data 1440 is the circular length. The calculation time delay can be adjusted. The time cost changes greatly with the time delay. Here consider the value is 100.

The calculation window consequence should be:

window_n: tn → tn +1440
window_{n+1} : tn +100 → tn +100 + 1440

5.4.1. Iterative Algorism

Through the application of EM, the mean value of these Gaussian distribution. can be obtained.

Step 1: Calculate the $W[z_{ij}]$ for each hidden variables. Assume the current $l =< \mu 1, \mu 2 \cdot\cdot \mu j >$.

$W[z_{ij}] = q(y = yi\ |\mu = \mu j)/(\Sigma^{2}{}_{n=1}\ q(y = yi\ |\mu = \mu n))$

Step 2:

The maximum similarity approach is considered to calculate the $P` =< \mu`, \mu`, \cdot\cdot\cdot \mu` >$, the $E[z_ij]$ is taken as an estimation of z_ij. then replace the $l =< \mu 1, \mu 2 \cdot\cdot\cdot \mu j >$ with $l` =< \mu`, \mu`, \cdot\cdot\cdot \mu` >$.

$\mu j \leftarrow (\Sigma^{m}{}_{i=1}\ E[z_{ij}\ ]\ xi) / (\Sigma^{m}{}_{i=1}\ E[z_{ij}\ ])$

To repeat these two steps, estimated mean value μj of sharing j can be simply attain

6. DISSCUSSION ON TIME SERIES ANALYSIS

6.2 The Up and Low Bound Method

After calculating the mean value μj , all of them are added into one and check whether the value varies greatly. If so it is considered that there may be some traffic anomaly in the network traffic.

$z_{up}(t) = x(t)` + k * r(t)\ z_{down}(t) = x(t)` - k * r(t)$

x(t)` is the mean value of mean value μj in the latest m Samples;

$x(t)` = (x(t) + x(t - 1) + x(t - 2) + \cdot\cdot\cdot + x(t - m + 1)) / m\ r(t)$ is the standard deviation of mean value μj sum in the latest m samples; $Ai = (x(t - i) - x(t)`)^2$

k is a weighting factor of fluctuation. The z up(t) represents the upper limit of mean value μj sum according to its tendency. The z down (t) represents the down limit of mean value μj sum according to its tendency.

If the value crosses the line a alert will be submitted. The k would be a configurable parameter which associated with the fluctuation range of the normal network traffic behavior.

6.3 The K and D Indicators Approach

The index denotes the association between uppermost value, deprived value of recent days and the value of the last day. This index can reflect the unexpected boost or decrease of the network traffic.

The calculation approach is listed below:
k (n) = 100 *[(C(n) − L5)/(H5 − L5)]
D(n) = 100 * (H3/L3)
In the formulation the C (n) is the value of time stamp n; L5 is the boost value in the most recent 5 times. H5 is the uppermost value in the most recent 5 times. H3 is the sum of (C-L5) in three times. L3 is the sum of (H5 - L5) in three data points.

The K line is more susceptible to the change of the new coming data than the D line. So if the K line passes through the D line, a fluctuation of network traffic is specified. So an alarm will be triggered. At other end, the next cross would be the signal of normal which means the anomaly has passed away.

7. NETWORK TRAFFIC MODELS

Typically, network traffic has been modeled as a Poisson process. Indeed, the Poisson model has been productively used in telephone networks for many years, and so it was innate when telecommunication networks happen to digital and started to send information as data packets. Also, this model has a simple mathematical expression [24], and has only one parameter, λ, which is in turn very instinctive (the mean traffic in packets per time unit). quite a few authors have projected network traffic behavior and presented other models that conquer the limitations which are inherent to Poisson processes, the most distinguished ones probably being that the Poisson model has a fixed association between mean and variance values (both are equal to λ), and that it does not account for high unpredictability or long-range reliance. A few projected models are usually stood on the assumption that network traffic is self-similar in nature, as originally stated in [23]. At this point, it should be obvious that any model for immediate traffic marginal’s must be supple enough to adapt to some properties observed in traffic, that is:

1. Let e (x) is the amount of traffic accumulated at time t. Then, e(x) <=e (x+1) and a (x+1)-e(x)<=N, where N is the network maximum transmission rate.

2. The actuality that at time t there is a certain amount of traffic E (x) does not imply in any way that at time x +1 the amount of traffic lies anywhere near x (t). This is equivalent to say network traffic show the high inconsistency assets.

The last property is also recognized as the “Noah effect” or the endless variance syndrome [24].

At the other side, the primary aforesaid property state the clear fact that network traffic has dense support between 0 and the N. compacted support creates symmetric distributions (Gaussian distributions are symmetric) inappropriate. so, if traffic data think near the maximum transmission rate, a symmetric model would let traffic increments to be larger than physically possible, again, with a non-negligible probability. This also influences the Gamma distribution.

7.1. α-Stable Model

α-stable distributions can be measured as a superset of Gaussian reason and create as the solution to the Central Limit Theorem when second order moments do not there [17], that is, when data can suddenly be different by large amounts as time passes by. This fits nicely to the high unpredictability property seen in network traffic. Moreover, α-stable distributions have an asymmetry parameter which agrees to their PDF to change from totally left-asymmetric to totally right-asymmetric, while genuine Gaussian distributions are always symmetric. This factor makes α-stable distributions of course flexible to the first traffic property (compact support) even when average traffic is almost 0 or very near the maximum theoretical network throughput. In addition, α-stable distributions give an explanation to the limit imposed in [23] about the requirement to aggregate many traffic traces for them to converge to a Gaussian distribution. According to the Generalized Central Limit Theorem [25], which includes the infinite discrepancy case, the sum of n α-stable distributions is another α-stable distribution, although not necessarily Gaussian.

8. CATEGORIZATION of DDoS FLOOD ATTACK

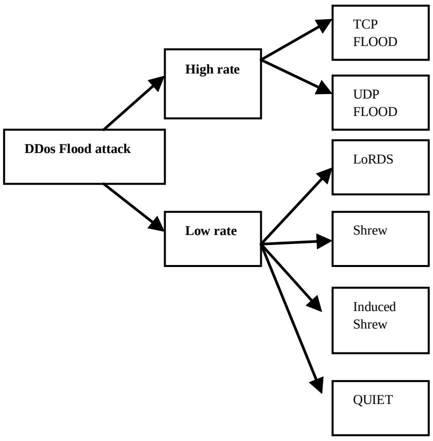

Fig.3 categorization of DDoS Attack

High rate flood attacks**:** at first, flood attacks are high rate flood. This is gifted by generating traffics from many machines, which may number thousands, distributed all over the world. Attack of the flood packets from the attackers will destroy the target hence degrading its routine to the extent of depiction it impractical. The high rate flood attacks examination in this study is the UDP attacks and TCP attacks. They are classify as high rate flood attacks because the attacks are launched by flooding a massive amount of TCP or UDP datagram’s to overpower the victim. Li *et al*. (2008), quantitative behaviors of flood attacks under diverse protocols. Quantitative performance of the attacks become the center in (Li *et al*., 2008) in order to explain the attacks randomly due to the shortage of traffic data of the real attack events. The reason is that in a lot of events of attacks, they will only be statement after the goal machines are already besieged and traffic data is lost.

Low rate flood attacks: opposing to the high rate flood, low rate flood uses carefully skill attack packets. The attack traffic rate is attuned in order to make them hidden by the traditional flood detector which regards high rate of incoming traffic as attack.

9. CONCLUSION

This paper has presented idea about the DDoS Attacks and their impact on network traffic. Here paper studied a DDoS attack to analysis the distribution of network traffic to recognize the normal network traffic behavior. This Paper has also discussed flooding attacks. The EM algorism is discussed to approximate the distribution parameter of Gaussian mixture distribution model. Another time series analysis method is studied. This paper also discussed a method to recognize anomalies in network traffic, based on a non restricted α-stable model and statistical hypothesis testing.

10. REFERENCES

[1] CERT, http://www.cert.com

[2] M. Li. An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern Recognition. Computers & Security, 23(7): 549-558, 2004.

[3] C.S. Sastry, S. Rawat and A.K. Pujari. Network traffic analysis using singular value decomposition and multiscale transforms. Information Sciences, 177(23): 5275-5291, 2007.

[4] M.F. Rohani, M.A. Maarof and A. Selamat. Continuous LoSS detection using iterative window based on SOSS model and MLS approach. In Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, May 2000

[5] H. Hajji. Statistical analysis of network traffic for adaptive faults detection. IEEE Transactions on Neural Networks, 16(5):1053–1063, September 2005.

[6] J. D. Brutlag. Aberrant behavior detection in time series for network monitoring. In LISA '00: Proceedings of the 14th USENIX conference on System administration, pages 139–146, Berkeley, CA, USA, 2000.

[7] D. Rincón and S. Sallent. On-line segmentation of non-stationary fractal network traffic with wavelet transforms and Log-likelihood-based statistics. LNCS, 3375: 110-123, 2005

[8] C. Douligeris and A. Mitrokotsa. DDoS attacks and defense mechanisms: classification and state of- the-art. Computer Networks, 44(5): 643-666, 2004.

[9] P. García-Teodoro, J. Díaz-Verdejo and G. Maciá-Fernández. Anomaly-based network intrusion detection: techniques, systems and challenges. Computers & Security, 28(1-2): 18-28, 2009.

[10] V. A. SIRIS and F. PAPAGALOU. Application of anomaly detection algorithms for detecting syn flooding attacks. In Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '04), volume 4, pages 2050–2054, Dallas, USA, 2004

[11] H. Wang, D. Zhang, and K. G. Shin. Syn-dog: Sniffing syn flooding sources. In Proceedings of the 22th International Conference on Distributed Computing Systems (ICDCS'02), pages 421–429, Washington, DC, USA, 2002. IEEE Computer Society.

[12] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. In Proceedings of the 29th International Colloquium on Automata, Languages and Programming (ICALP '02), pages 693–703, London, UK, 2002. Springer-Verlag.

[13] J. D. Brutlag. Aberrant behavior detection in time series for network monitoring. In LISA '00: Proceedings of the 14th USENIX conference on System administration, pages 139–146, Berkeley, CA, USA, 200

[14] V. Paxson. Bro: A System for Detecting Network Intruders in Real- Time. In Computer Networks, volume 31 (23–24), pages 2435–2463, 1999.

[15] E. S. Page. Continuous inspection schemes. Biometrika, 41:100–115, 1954.

[16] S.X. Wu and W. Banzhaf. The use of computational intelligence in intrusion detection systems: A review. Applied Soft Computing, 10: 1- 35, 2010.

[17] O. Salem, S. Vaton, and A. Gravey. An efficient online anomalies detection mechanism for high-speed networks. In IEEE Workshop on Monitoring, Attack Detection and Migitation (MonAM 2007), November 2007.

[18] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE/ACM Trans. Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994

[19] Y. Guan, A. A. Ghorbani, and N. Belacel, An unsupervised clustering algorithm for intrusion detection. In Proc. of the Sixteenth Canadian Conference on Artificial Intel ligancy (AI 2003), pages 616-617, Halifax, Canada, May 2003. Springer

[20] Huang Kai, Qi Zhengwei, Liu Bo" Network Anomaly Detection Based on Statistical Approach and Time Series Analysis" 2009 International Conference on Advanced Information Networking and Applications Workshops

[21] Federico Simmross, Juan Ignacio, Pablo Cassia-de-la-Higuera, Ioannis A. Dimitriadis" Anomaly Detection in Network Traffic Based on Statistical Inference and α-Stable Modeling" IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 8, NO. 4, JULY/AUGUST 2011

[22] Khadijah Wan Mohd Ghazali and Rosilah Hassan:" Flooding Distributed Denial of Service Attacks-A Review "Journal of Computer Science7 (8): 1218-1223, 2011 ISSN 1549-3636

[23] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE/ACM Trans. Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994.

[24] A. Papoulis, Probability, Random Variables, and Stochastic Processes, third ed., McGraw-Hill, 1991.

[25] G.R. Arce, Nonlinear Signal Processing: A Statistical Approach. John Wiley and Sons, 2005.

[26]Monika sachdeva, gurvinder singh, Krishan Kumar, Kuldeep Singh"DDoS Incident and their Impact"IAJIT2010"

[27] Khadijah Wan, Mohd Ghazali Rosilah Hassan" Flooding Distributed Denial of Service Attacks-A Review "

[28] CNN. Cyber-attacks batter Web heavyweights, February2000/www.cnn.com/2000/TECH/computing/02/09/cyber.attacks

[29] CNN .Immense. Network assault takes down Yahoo, February http://www.cnn.com

[30] Netscape. Leading web sites under attack, February 2000 technews.netscape.com "Journal of Computer Science

[31] CERT coordination center. Denial of Service attacks http://www.cert.org/tech_tips/denial_of_service.html

[32] Distributed Denial of Service (DDoS) Attacks/tools. http://staff.washington.edu/dittrich/misc/ddos