THWARTING DISTRIBUTED DENIAL OF SERVICE ATTACKS USING NORMAL DISTRIBUTION AND WEIBULL THEOREM

V.ShyamalaDevi [1] **Dr. R. Umarani [2]**

Associate Professor / Department of MCA, KSRCT,Tiruchengode , Tamilnadu
Associate Professor / Department of MCA, Saradha Womens, Salem, Tamilnadu

Abstract

DDoS attacks have become common place on the Internet today. These attacks occur when a hacker gains control of a number of hosts on the network and directs large volumes of traffic from those hosts to one or more target hosts. They often use Botnets in such attacks. Botnets are large collections of computers infected by worm's s that are taken over and remotely controlled by hackers to send spam, propagate viruses, or launch denial of service attacks. The number of compromised hosts on the Internet can be staggering in the hundreds of thousands. Service providers are regularly needed to protect and mitigate the attacks which occur on their networks. This paper discusses about the attacking flow on the Internet and ways that service providers can prevent or mitigate damages from the attack threats.

This paper focuses the attacks and discusses the mitigating techniques to prevent, legitimate packet dropping in a service provider environment. Implementing a new thing Normal distribution with weibull theorem derives a sample of n number of packets is obtained from victim machine and those packets are tested by normal distribution method to find the actual intruder who attacked the victim machine. The slope of the weibull plot, beta, (β), determines which member of the weibull failure distributions best fits or describes the data. The slope, β, also indicates which class of failures is present.

$\beta < 1.0$ indicates infant mortality
$\beta = 1.0$ means random failures (independent)
$\beta > 1.0$ indicates wear out failures

keywords : DDoS, Normal Distribution, Weibull, THM, DoS

1.Introduction to DoS / DDoS

Dos attacks are a result of exploiting the vulnerabilities present in the system or architecture. This can bring down any service be overloading it or injecting. Resulting is no further request processed by the server.

a)Types or Levels of Dos Attacks

Logic Attacks: These attack can exploit vulnerabilities in network software such as web server or the underlying TCP/IP stack.

Protocol Attacks: Exploiting a specific feature or implementation bug of some protocol installed at the victim in order to consume excess amounts of its resources. Protocols here are rules that are to be followed to send data over network. DDoS is a much more powerful attack than a normal DoS, the attack is being generated by one system. An amplifying network might be used to bounce the traffic around, but the attack is still originating from one system. A DDoS takes the attack to the next level by using agents and handlers. DDoS attackers have joined the world of distributed computing. One of the distinct differences between DoS and DDoS is that a DDoS attack consists of two distinct phases. First, during the preattack, the hacker must compromise computers scattered across the Internet and load software on these clients to aid in the attack. After this step is completed, the second step can commence. The second step is the actual attack. At this point, the attacker instructs the masters to communicate to the zombies to launch the attack.

Figure 1, the DDoS attack allows the attacker to maintain his distance from the actual target. The attacker can use the master to coordinate the attack and wait for the right moment to launch. Because the master systems consume little bandwidth or processing power, the fact that these systems have been compromised will probably not be noticed. After the zombies start to flood the victim with traffic, the attack can seem to be coming from everywhere, which makes it difficult to control or stop.

2. Methods for Mitigating the DDoS Threat

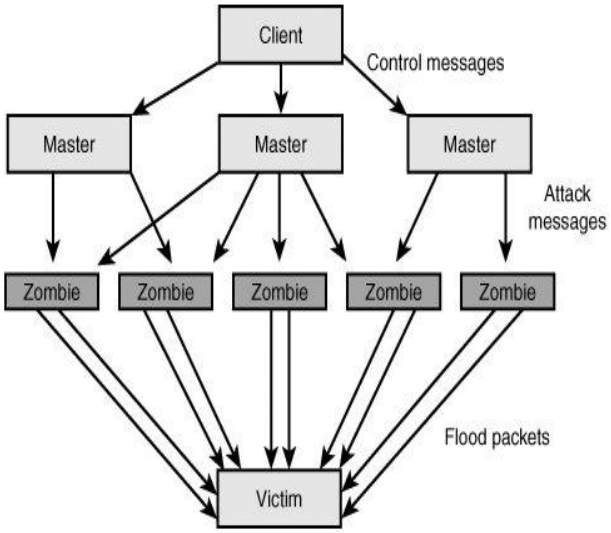


Figure 1. DDoS attack

Taking on DDoS attacks requires a new approach that not only detects increasingly complex and deceptive assaults but also mitigates the effects of

the attack to ensure business continuity and resource availability.

Complete DDoS protection is built around three key themes:

1. Mitigate or identified, not for detection.
2. Accurately distinguish the good traffic from bad traffic to preserve the continuity, just to detect the overall presence of an attack using Normal distribution Technique.
3. Improve the performance to deploy the new application (weibull theorem) in upstream to protect all points of vulnerability.

2.1 Generation of Attack Mechanism

Most of the researches are carried out mainly in the two areas namely Attack Traceback and Attack Mitigation. This study focuses mainly on the Mitigation of IP addresses used for the information transmitted on the attack process. It will be carried out after an attack has been launched and it will prevent the forthcoming attacks to the system. Attack Traceback addresses the problem of collecting the information about individual packet forwarding agents and collating this data to obtain an approximate Internet router-level graph (attack tree rooted at the victim); whereby tracing the routing path that any packet has taken, provides sufficient basis for attack attribution (attack tree leaves). The Attack traceback is necessary for cleansing zombie attackers, while also being of critical forensic value to law enforcement. The major sources of attacks are due to the increase in the accessing of the network resources by an outside unauthorized

user. These users are from different geographical regions and from different countries. This makes the traceback process difficult in the real time situation.

The information transmitted from a router consists of the source address and the destination address along with the information content. The advantage of the proposed method is that it will split the entire network into various sub-networks, and helping to identify the attacker in an easy manner with the use of geographic information. The geographical information helps to trace the system of source of attack that residing in the network. The result is compared with the other existing traceback schemes.

In the present generation, a DDoS attack poses a serious threat to a large number of organizations. The reason is due to the number of systems involved in accessing the internet is increasing day by day in a rapid manner. Due to this, the traffic and the information access become difficult. Also the availability of space for providing the IP address to the systems has gone beyond the limit. The preventive measures against this attack is also a major difficult task due to various reasons like an increase in traffic, availability of latest technologies for packet transmission and increase in the usage of internet among the people.

To overcome these problems, few basic countermeasures against the attacks namely detection, mitigation, prevention are considered. The basic step to carry out is to identify the source of the attack and to check whether the attack is happening in the network or not. It is identified with the help of information packets and their rate of arrival to destination machine. These packets are classified into two basic categories such as the valid packets from the legitimate user and the attack packets from the source of the attacker as shown figure 2. The process of identifying the attack packets, carried out with the help of normal distribution

When the victim machine feels congestion in traffic, the reason for this may be, over flow of information packets and some other factors. Dilemma over the congestion may be due to more number of packets sent by the hackers. In this situation, a proper

rescue mechanism has to take up and deal with the traffic congestion in order to make a smooth flow of packets in the network. The information flow in the network should be monitored frequently in order to achieve high efficiency. For finding out the attack, it has to be identified whether the packet received by the receiver is legitimate or illegitimate packets deliberately sent by hackers. To achieve this, Trust management Helmet (TMH) along with normal distribution is applied in order to rescue from the problem. The reason for choosing the normal distribution is to regulate the flow of the packet in the network and in order to make the network traffic less.

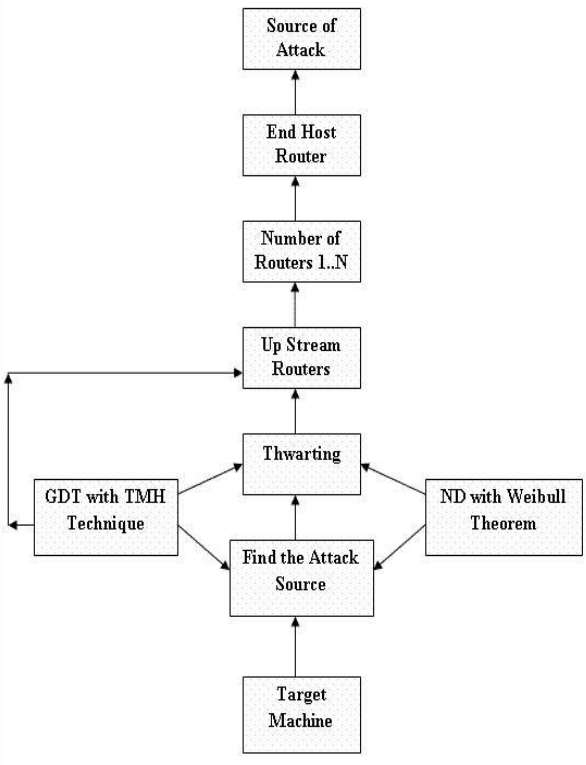


Figure 2 Data flow diagram for thwarting a DDoS attack

2.2 Attack Detection using Normal Distribution

Normal Distribution is the process of finding the probability of failure, undesirable event in a large group of quantity or in an augmentation of information. It is practically impossible to calculate and qualify all items in given specified time. To calculate large amount of such items, normal distribution is being used in various quantitative study. Even though various distributions of calculations are available, to derive a pattern from large quantity, normal distribution is opted.

In this work, normal distribution is applied to identify illegal packets sent by intruders. A sample of n number of packets is obtained from victim machine and those packets are tested by normal distribution method to find the actual intruder who attacked the victim machine. All illegal packets being routed through upstream router are blocked and legal packets are allowed to reach its destination uninterruptedly. The

equation which describes the normal distribution is

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \quad (1)$$

$$\overline{X} = \frac{\sum_{k=0}^{n} x_k}{n} \quad (2)$$

Z= Normal Distribution for packets transmission in victim machine.

$\overline{X}$ = sample packet in victim machine

μ = The population packets in victim machine

σ = Standard Deviation of population packets

n=Total number of packets in the sample

X_k=k^{th} value of Selected sample packets

The equation 1 gives the normal distribution where Z is the normal distribution of sample packets transmission. Sample packets are the packets which are selected from 'n' number of packets to test for its attack and mean of those sample packets at victim can be found by using the formula 2.

A statistical hypothesis test is a method of making decisions using data, whether from a controlled experiment or an observational study (not controlled). In statistics, a result is called statistically significant if it is unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level. In probability, these decisions are almost always made using null-hypothesis tests. Hypothesis testing allows us to use sample data to test a claim about a population, such as testing whether a population mean equals sample mean.

3. Distribution Sampling

Hypothesis testing: Often want to know the likelihood that a given sample has come from a population with known characteristic(s)

Define H_0

Test H_0

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

normal distribution with mean 0, standard deviation 1

Example.

$\overline{X}$ = 104.0

$H_0 : \mu = 100$

$\sigma_{\overline{X}} = 3$

z = (104 – 100) / 3 = 1.33

α = 0.05

therefore retain H_0 $t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$

for a given mean and sd, normal distribution is completely defined there are a family of t curves, depending on degrees of freedom n – 1 degrees of freedom associated with deviations from a single mean with infinite degrees of freedom, t = z ,H_0 : μ = 100

$\overline{X}$ = 120

n = 25

s_x = 35.5

$$s_{\overline{X}} = \frac{s_x}{\sqrt{n}} = \frac{35.5}{\sqrt{25}} = 7.1$$

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{120 - 100}{7.1} = 2.82$$

df = 24

α = 0.05

3.1 Significant Values

The Sample values of the statistic beyond which the null hypothesis will be rejected are called significant values. Two types of test Two tailed test and one tailed Test

Tests in One Tail and Two tail

When two tails of the sampling distribution of the normal curve are used, the relevant test is called two tailed test.The alternative hypothesis H_1: $\mu_1 != \mu_2$ is taken in two tailed test for H_0:$\mu_1=\mu_2$.

When only one tail of the sampling distribution of the normal curve is used, the test is described as one tail test.

For H_0: $\mu_1= \mu_2$, the formulated alternative hypothesis is either H_1: $\mu_1> \mu_2$ or H_1:$\mu_1<\mu_2$.

3.2 Testing a Hypothesis Function

1) Formulate H_0 and H_1.
2) Choose the level of significance α.
3) Compute the test statistic Z, using the data available in the problem
4) Pick out the critical value at α % using Zα.

Draw conclusion: if $|Z| < Z\alpha$, accept H_0 at α % level. Otherwise reject H_0 at α % level.

3.3 Single Proportion Test:

Single proportion test utilizes the information about the traffic packets and check whether the travelling path consists of attack packets. If x is the number of items possessing a certain attribute in a sample of n items, then the sample proportion p = x/n. Consider a sample of size n with proportion P taken from a population. Let P be the population proportion. To test whether

a) The difference between sample proportion P and population proportion p is significant or not.

b) The sample has been chosen from the population, we proceed as follows.

Let the null hypothesis be H_0: p=P i.e., p has a specified value.

The alternative hypothesis is H_1: p! = P

The equation 3 which describes the normal test is

$$Z = \frac{p-P}{\sqrt{pq/n}} \qquad (3)$$

If $\alpha = 0.05$ is the level of significance, we compare the calculated Z with value 1.53.

$|Z|<1.53$, H_0 is accepted. Otherwise it is rejected.

In general, $|Z|<3$, H_0 is accepted, Otherwise H_0 it is rejected.

In a web server, a sample of 100 packets is drawn. 92 packets are attack packets and remaining is legitimate packets. In general, both attack packets and legitimate packets are equally distributed in the particular web server at 5% level of significance.

$$Z = \frac{p-P}{\sqrt{pq/n}}$$

Z=Normal Test for Proportion.

P= Proportion of Population

p=Sample Proportion of attack packets

x=No of attack packets

n=Total no of samples packets from the population

H_0: p=1/2 (Legitimate packets and attack packets are equally distributed)

H_1: p>1/2 (Large number of Attack packets in that group) Level of(Attack Packet) Significance Fixed as 5%

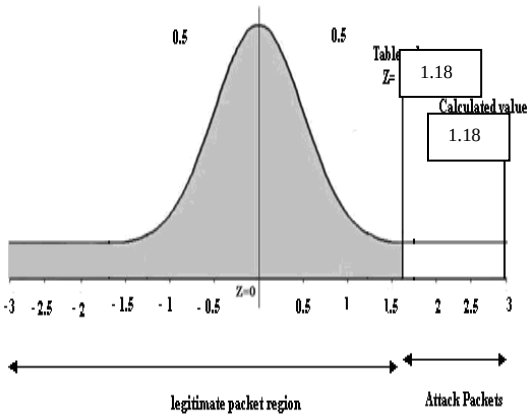


Figure 2 Data flow diagram for thwarting a DDoS attack

Since Z (Calculated value)>Z (Tabulated value) so H_0 is Rejected, H_1 is accepted; there is more number of attack packets in that web server at the time of testing.

Once the attack packet is identified, the information about the packets can be combined and forwarded to the upstream router as Regional Identification Mark (RIM) value. In addition to the Normal Distribution, the Trust Management Helmet is used to make sure whether the packets identified using Normal Distribution are attack packets. The next section discusses how thwarting is carried out.

4. Trust management Helmet (TMH)

Trust Management Helmet is a lightweight mitigation mechanism to mitigate session flooding DDoS attack that uses trust to differentiate legitimate users from attackers. The trust of clients is evaluated based on their visiting history and used to schedule the service to their requests. For every established connection it records four aspects of trust to the user: short-term trust, long-term trust, negative trust and misusing trust, which are used to compute an overall trust that helps in determining whether to accept the request a client for next connection.

The information stored on the clients is termed as license and is given to each client for communication and the computed values are stored in it. Each time the client establishes a connection and its license is checked by the server to establish connection and updated

The identification information, such as ID and IP, must be stored at the client license. The state variables for trust computation can be stored at the client or at the server. The client provides the license whenever he requests a connection. TMH verifies the license by first checking whether the request originates from the IP address included in the license and whether the last access time matches the server's log, then

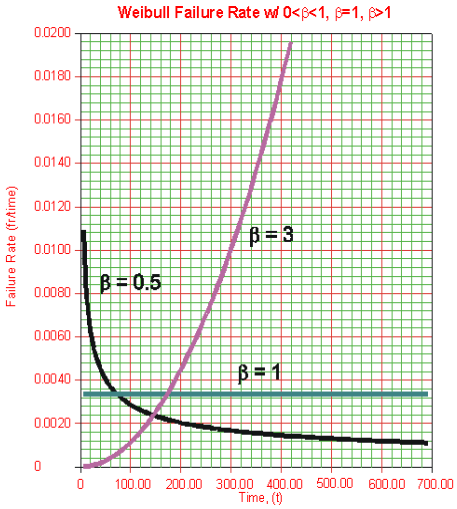
validating it, if the hash H agrees with the hash computed using the license and the server password. Connection request without a license will be treated as from new users and a new license will be issued if TMH decides to accept it.

To reduce the processing overhead brought by TMH, a short-term blacklist should be implemented. The blacklist records the list of clients whose trust values are too low. When a client's trust drops below some threshold, he is recorded into the blacklist with an expiration time. He is then banned from accessing the server until his blacklist record expires. Since the TMH mitigation mechanism is deployed on the server, a session connection request first reaches TMH and it checks whether the client is blacklisted; if not, it computes the trust of the client and uses trust-based scheduling to schedule the connection request to the server.TMH can also be used among multiple servers termed as Collaborative trust Management and the collaborating TMHs can take either or both actions. Exchange blacklist: When a TMH receives a blacklist (periodically), it merges the received blacklist into its own. Exchange the trust values of clients: As a client may visit the same server multiple times within a period, only the latest overall trust logged by TMH is exchanged periodically.

4.1 Weibull Distribution Theorem

The Weibull shape parameter, β, is also known as the Weibull slope. This is the value of β is equal to the slope of the line in a probability plot. Different values of the shape parameter can have marked effects on the behavior of the distribution. In fact, some values of the shape parameter will cause the distribution equations to reduce to those of other distributions. The most important aspects of the effect of β on the Weibull distribution. As is indicated by the plot, Weibull distributions with $\beta < 1$ have a failure rate that decreases with time, also known as infantile or early-life failures. Weibull distributions with β close to or equal to 1 have a fairly constant failure rate, indicative of useful life or random failures. Weibull distributions with $\beta > 1$ have a failure rate that increases with time, also known as wear-out failures. These comprise the three sections of the classic "bathtub

curve." A mixed Weibull distribution with one subpopulation with β < 1, one subpopulation with β = 1 and one subpopulation with β > 1 would have a failure rate plot that was identical to the bathtub curve.



The weibull reliability function given by

$$\lambda(R(T) = e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta}}$$

where Γ(*) is the gamma function. The gamma function is defined as:

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$$

The cumulative hazard function for the Weibull is the integral of the failure rate A more general three-parameter form of the Weibull includes an additional waiting time parameter μ (sometimes called a shift or location parameter). The formulas for the 3-parameter Weibull are easily obtained from the formulas by replacing *t* by (*t* - μ) wherever *t* appears. No failure can occur before μ hours, so the time scale starts at μ, and not 0. If a shift parameter μ is known (based, perhaps, on the physics of the failure mode), then all is do the subtract μ from all the observed failure times and/or readout times and analyze the resulting shifted data with a two-parameter Weibull.

5. Simulation Experimental results

This study helps to analyze the packet information and filter it based on the available information. It feeds the information in the packet only once when it enters into the first router in the network. Figure 1 discusses the computational burden and scalability comparison with different aspects (Packet Detection using Normal distribution). The experimentation is conducted with two networks containing 24 nodes interconnected with one another.

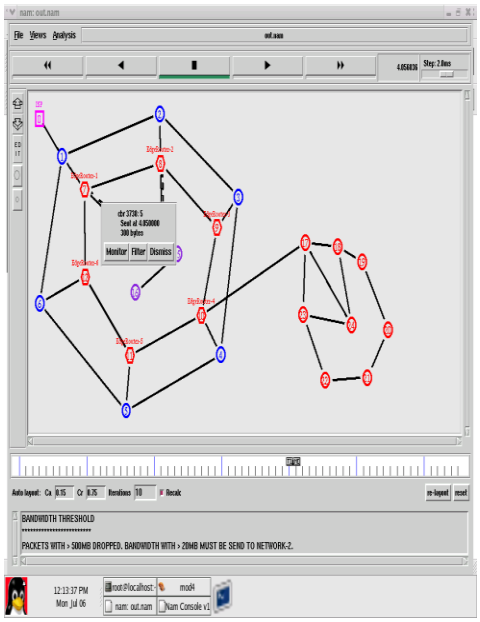
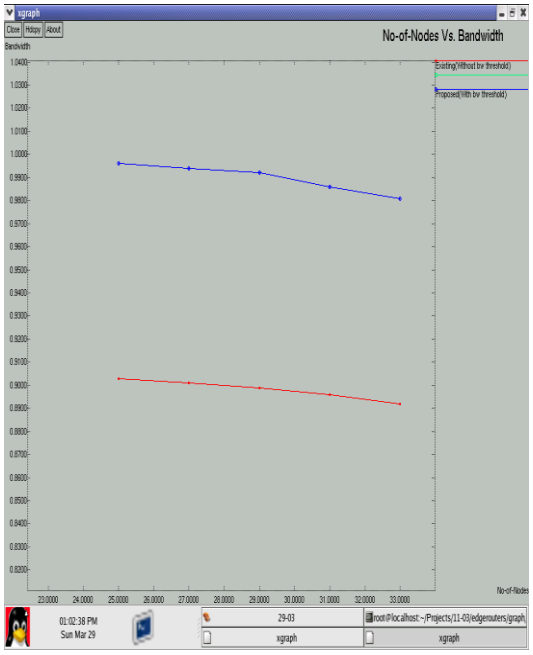
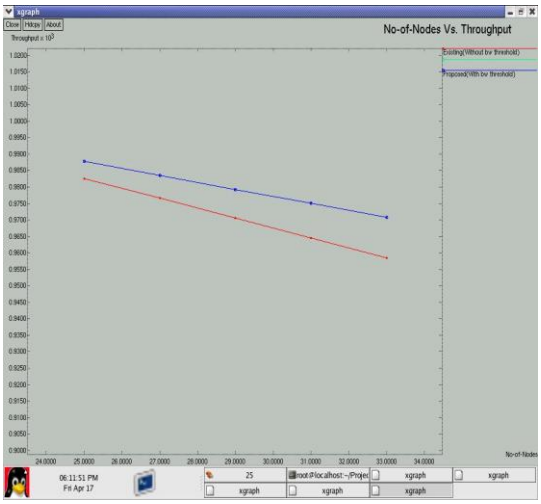


Fig 5.1 Packet Detection using Normal Distribution



Graph 1 Network Utilization Vs Number of nodes in the network



Graph 2: Throughput Vs Number of Nodes

Graph 1 shows the network utilization on the application of the DDoS resistive scheme in the network communication for flooding and packet marking attack. As the number of nodes increases, network utilization decreases for both existing and proposed scheme. However when compared to existing method the network utilization is optimized in the proposed method. Graph 2 depicts the output of the simulation by varying the nodes there is an appreciable change in the throughput of the data communication. As the number of nodes increases, throughput decreases.

6. Conclusion

These papers presents a system for defending against the DDoS and investigate the traffic rate analysis and mitigate mechanism along with Normal distribution, weibull application to avoid the maximum number of failure rates over the entire flow of network . Compared the various technique, the Thwarting process is effectively achieved with the help of GDT technique. The Trust management Helmet along with ND plays a major role in thwarting the attack. It reduces the overhead at the router and increases its performance. In future, to determine the reliability of this method in different network settings. This frame work traffic becomes prohibitive and valuable to detect flood attack through the internet.

References

1. Y. Chen, K. Hwang, Spectral Analysis of TCP flows for defence against reduction-of-quality attacks, in: The 2007 IEEE International Conference on Communications (ICC'07), June 2007, pp. 1203–1210.

2. Hongli Zhang, Zhimin Gu, Caixia Liu, Tang Jie, Detecting VoIP-specific denialof- service using change-point method, in: 11th International Conf. on Feb. 2009, pp. 1059–1064.

1. F. Yi, S. Yu, W. Zhou, J. Hai, A. Bonti, Source-based filtering algorithm against DDOS attacks, International Journal of Database Theory and application 1 (1) (December 2008) 9–20.

2. K. Lu, D. Wu, J. Fan, S. Todorovic, A. Nucci, Robust and efficient detection of DDoS attacks for large-scale internet, Computer Networks 51 (September 2007) 5036–5056.

3. S. Yu, W. Zhou, R. Doss, Information theory based detection against network behavior mimicking DDoS attack, IEEE Communications Letters 12 (4) (April 2008) 319–321.

4. H. Sengar, H. Wang, D. Wijesekera, S. Jajodia, Detecting VoIP floods using the Hellinger Distance, in: IEEE Transactions on Parallel and Distributed Systems, vol. 19, No. 6, June 2008. pp. 794–805.

5. Y. Jin, E. Sharafuddin, Z.-L. Zhang, Unveiling core network-wide communication patterns through application traffic activity graph decomposition, in: Proceedings of ACM SIGMETRICS, 2009.

6. A. Chen, L. Li, J. Cao, Tracking cardinality distributions in network traffic, in: Proceedings of IEEE INFOCOM, 2009.

7. X. Meng, G. Jiang, H. Zhang, H. Chen, K. Yoshihira, Automatic profiling of network event sequences: algorithm and applications, in: Proceedings of IEEE INFOCOM, 2008.

8. P. Dhungel, D. Wub, K. Ross, Measurement and mitigation of BitTorrent leecher attacks, Computer Communications 32 (2009) 852–1861.

9. G. Neglia, G. Reina, H. Zhang, Availability in BitTorrent Systems, in: Proceedings of INFOCOM'07, 2007.

10. D. Karger, M. Ruhl, Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems, in: Proceedings of IPTPS'04, 2004.

11. A. Rao, K. Lakshminarayanan, S. Surana, R. Karp, I. Stoica, Load Balancing in Structured P2P Systems, in: Proceedings of IPTPS'03, 2003.

12. Shen, C. Xu, Locality-aware and churn-resilient load-balancing algorithms in structured peer-to-peer networks, IEEE Transactions on Parallel and Distributed Systems (2007).

13. P. Jordi, G. Pedro, PlanetSim: An Extensible Simulation Tool

14. A. Montresor, M. Jelasity, PeerSim: A Scalable P2P Simulator, in: Proceedings of P2P'09, 2009.