An Autonomous Framework for Early Detection of Spoofed Flooding Attacks

Malliga Subramanian and Tamilarasi Angamuthu
(Corresponding author: Malliga Subramanian)

Department of Computer Science and Engineering, Kongu Engineering College
Perundurai, Erode, Tamil Nadu 638 052, India
(Email: mallisenthil@yahoo.com)
(Received Jan. 22, 2008; revised and accepted May 9, 2008)

Abstract

One of the challenging tasks on the Internet is differentiating the attack traffic from legitimate traffic. Tackling this challenge would aid in the detection of Denial of Service/Distributed DoS (DoS/DDoS) attacks. In this paper, we propose a flow profiling scheme that adopts itself to detect these flooding attacks by monitoring the trends in the current traffic. Moreover, our scheme filters most of the traffic, which are found to be suspicious, at the source end, thus avoiding flooding at the target. The scheme distinguishes itself from other source end defenses in the manner in which it gathers and profiles the statistics. Information entropy, a measure to find correlation among traffic flows, is used. We made this attempt to infer the current state of the dynamic network. The result of correlation is then used to support the evidences which justify the necessity of filtering the packets. We use Theory of evidence to improve the decision making with regard to filtering. We implemented and tested our scheme using network traffic traces and found the results to be appreciable.

Keywords: DoS, DDoS, flooding, information entropy, theory of evidence, traffic profiling

1 Introduction

According to CERT [2], DoS attacks can be defined as [9, 10]:

1) Occupancy of limited resources or difficult to renew such as network bandwidth, data structure or memory of a system.

2) Changeable or damage network data, for instance, delete system configuration, shutdown web services etc.

These flooding attacks have attracted the attention of many networking research groups in the recent years. In spite of that, it is sill a hard problem to locate their origin and defend against them. During the recent years, the network users have faced many unpleasant events due to these bandwidth or flooding attacks. For instance, 8 out of 13 DNS servers were brought down by these flooding attacks in October 2003 [12]. Furthermore, the attackers exploit the inherent weakness in the Internet Protocol. The attackers, who employ spoofing of source IP addresses, further harden the defensive mechanisms. The spoofing disables the defensive systems from locating the source of these flooding attacks.

The systems which detect and prevent these bandwidth attacks can be deployed at three different locations in the wide area network like Internet, namely at source end, at the intermediate routers and at the target side [15]. To enhance the victim's ability to detect the attacks, many systems, which are deployed at the victim, have been advocated in the recent past. Most of the defensive systems are employed at the victim network since it is the one which suffered from these attacks at most. These systems protect the victim from the flooding attacks and reduce the impact of these attacks on the victim. Such techniques, which employ the victim end defense, are presented in [4, 11] and [14]. Detecting the bandwidth attacks at the victim side suffers from few drawbacks [23], like detecting abnormal traffic flow at the victim is very hard until the attacking is at the bursting situation and making it hard for the victim to execute a post detection response at the accumulation of sheer volume of the attack packets.

The defensive mechanisms that are deployed at the core and intermediate routers would enjoy the benefits of observing huge traffic. Such mechanisms include IP traceback, logging etc. IP traceback, proposed in [1] and [17], refers to locating the true origin of any packet sent over the network. But, by spoofing, this is very much complicated. To aid in the process of traceback, various Probabilistic and Deterministic Packet Marking (PPM and DPM) schemes have been proposed [15] and [17]. These schemes use the identification field of IP header for mark-

ing. The PPM and DPM mechanisms defer the way they mark the packets. PPM marks the packets based on probability wherein DPM, all the packets that pass through the routers are marked. These markings are used by the victim to construct the attack path to find the origin of the attack. In general, these methods require considerable amount of marked packets to reconstruct the attack path.

Logging refers to buffering the packets at the intermediate routers and using them to infer about the attack path. To reduce the storage needs, methods like [19] have been proposed. These methods store the digest of the packets rather than the packets itself. Nevertheless, these routers are required to operate at high speeds and be rich in resources.

To overcome the drawbacks of intermediate and victim end defensive mechanisms, the approaches that give early indication of the flooding attacks at the source end have been suggested. The objective of the DoS defensive systems deployed at the source network is to identify and deny the packets sent by the attackers from reaching the Internet, thus avoiding them from flooding the victim and thus the network. For handling the flooding attacks near the source end, the key issue to be considered is the detection of the attacks. But, the attacks appear different at the source end than at the victim. It is quite inevitable to detect the attacks at the victim, since the entire abnormal flow targets at the victim. But, because of the distributed nature of the attack, the same traffic, at the source side, appears to be legitimate traffic. If those traffics are identified correctly, then the attacks can be prevented at the place of origination itself. Hence an ideal place to detect and prevent the DoS attacks should be as close to the source as posible.

Recent researches have proposed such schemes. Few of these schemes run on 3-way handshake between the client and server during the TCP communication and some of them are based on flow profiling. The schemes like D-WARD, Ingress and Egress filtering and some traffic profiling schemes [5, 14, 20, 21, 22] and [23] are examples for the source end defenses.

There are few hard challenges while testing the ability of a scheme for detecting the flooding attacks at the source. The first point to be pondered is the detection of the illegitimate traffic near the vicinity of the source, since the traffic is highly dispersed and small compared to the sheer volume of the traffic at the victim. The other point to be considered is that the detection should be as soon as possible without giving a false alert. Consuming large quantities of resources at the victim would be the characteristics of DoS attacks. To do so, the attackers try to compromise the hosts and spoof the packets with random IP addresses and packet sizes. In order to send huge volume of traffic, they would also establish many connections to the victim. By observing the changes in the traffic pattern, we can deduce about the attacks. In this paper, we adapt a detection approach, which observes the packet contents. We monitor the source IP address and packet size distribution of the packets. The reasons to choose these two parameters are, it is well known from [20] and [21] that monitoring the arrival of new source IP addresses would be useful in detecting bandwidth attacks. But, instead of monitoring the number of new source IP addresses and making decision just based on the excessive number of IP addresses, we use the information entropy to observe the increase in new IP addresses. The other heuristic, packet size distribution of the packets from a source network, is to further support our belief. The belief is used as a foundation for our scheme. We put together these evidences to detect the DoS attacks.

The rest of the paper is organized as follows: The related researches on source end defensive systems for DoS attacks are discussed in Section 2. Section 3 gives a presentation on Information entropy and Theory of Evidence and our way of using these mathematical concepts. Our system for detection of flooding attacks is enumerated in Section 4. In Section 5, we evaluate the proposed approach using simulation and present the results of the same. This Section also provides the comparison of the proposed system with other source oriented defensive systems. The results of the simulation experiments are interpreted in Section 6. The conclusions are provided in Section 7.

2 Source-end Defenses for DoS

This section gives a discussion on the existing solutions for source end defense. An ideal place to stop these attacks is the place closest to the source. There are definite advantages for source end defenses over the intermediate and victim end defenses [15]. These include small collateral damage, less traffic, effective utilization of scarce resources, dispersed protection mechanisms etc. Nevertheless, there are some hard facts about the detection in case of source end defenses. Detecting the attack near its starting is hindered by its distributed nature. Because of its distributed behavior, it is hard to gather statistics about the anomalous and misbehaving traffic near the source network. Besides, an attack may be launched using legitimate requests, thus monitoring the traffic going out may not give correct alert.

There are some research efforts that investigated the effectiveness of the defenses at the source. In [13], Mirkovic et al. proposed a system architecture that is located at the source network, which autonomously observes, detects, responds and puts an end to the flooding attacks originating from the source network. This system uses egress filtering [3] and request/response rate of TCP communication and discards outgoing traffic that does not follow the traffic policies. In the approach proposed by Wang et al. [22], the protocol behavior of the TCP SYN-FIN and SYN-RST pairs are used to detect the flooding attacks that exploit the TCP's 3-way handshaking mechanism. But these schemes fail for the packets that spoof the source addresses belonging to the source network it-

self. We also have approaches that have their foundation on the statistics of the traffic from the source network. These detection techniques are classified into anomaly-based and misuse based [8] and [15].

The mechanisms that run on anomaly-based maintain the patterns of normal system behavior. Each packet is monitored and compared with the stored patterns to discover the anomaly. In contrast to this, misuse-based or pattern detection approaches store the signatures of the known attacks in a database. Then the current traffic is compared with the database entries to find the patterns matching. The obvious drawback of misuse detection approaches is that they can only detect known attack patterns and are not for detecting new attacks that do not match with stored patterns. The system presented in [5] is based on filtering on packet type and rate. It uses disproportional packet rates to and from the hosts and the victim as a heuristic to determine the bandwidth attacks. This system can be deployed at both source and victim end.

This paper adopts an anomaly-based approach that uses information entropy to detect DoS attacks. We use the information present in the packets to detect the anomaly. We adopt to use the source IP address as a parameter to detect anomaly. Our approach maintains the list of source IP addresses that appeared in the traffic frequently from the source network and observes for the other IP addresses.

The paper [20] proposes a similar approach, wherein the system monitors the number of new source IP addresses in a prescribed time period to detect the DoS attacks and uses CUSUM algorithm to detect the changes in number of new IP addresses. Moreover, this scheme is based on the assumption that DoS attacks generally use a large number of spoofed IP addresses. In case, the attackers reduce the number of spoofed IP addresses below the threshold, then the system would run on the assumption that when the number of spoofed IP addresses decreases, then there would be increase in number of packets and detect the attacking source subsequently. But this can not be a valid assumption always.

Moreover, the new IP addresses can be easily determined by egress/ingress filtering proposed in [3]. The decision making is further enhanced by monitoring another variable namely packet sizes. We have strong reasons for choosing these two parameters. From [21], we understood that most of the IP addresses received during a specific interval have already appeared and also analyzed the traces of traffic from [6] and [16] and found that the traces have a set of packet sizes being used repeatedly. We design a probabilistic model on these two heuristics to aid in decision making. We use this model to support our belief since the model measures uncertainty associated with random variables. We provide a closer look on the proposed model in the following sections.

3 Mathematical Representation of the Proposed Model

Since the Internet is viewed as system that does not lend itself to be represented as a functional model and also exhibits dynamic and stochastic behavior, it is very hard to understand the system state. During our study, we found that the behavior of the system, that shows randomness and uncertainty, can be understood only by defining probabilistic models for the system. One such probabilistic representation is Information entropy. We modified and used the principles of information entropy to predict the behavior of the network which is dynamic. After having predicted the current state of the network, Dempster Shafer's Theory of Evidence, a mathematical theory using belief and plausible reasoning, is used to support the belief of current state. In [18], the authors have used this theory to collect the reports from different sensors, which include a sensor that gathers details about TCP, UDP and ICMP packet rates and a sensor that collects and analyses SNMP data, to infer the current status about system being monitored. We follow similar approach to use our evidences from the probabilistic model to enhance the decision making with regard to DoS attacks.

3.1 Information (or) Shannon Entropy

According to information theory, the information entropy is a measure of randomness and uncertainty associated with a random variable. It measures the average information contained in piece of data. The information entropy of a random variable S, that takes the possible values $(X_1, X_2, X_3 \ldots X_n)$ is given as

$$\begin{aligned} H(S) &= E(I(S)) \\ &= \sum P(X_i) \log(\frac{1}{P(X_i)}) \\ &= -\sum P(X_i) \log(P(X_i)), \end{aligned}$$

where $I(S)$ is the information contained in S, which is a random variable to be monitored and $P(X_i)$ is the probability mass function of S. We use the properties of entropy to predict the current state of the system. Few remarkable properties are as follows:

1) Continuity. This property indicates that the measure should be continuous (i.e. changing the value of one of the probabilities by a considerable amount should change the entropy also). We use this property to find the changes in source IP addresses and standard packet size distribution and the probability of packets from those addresses with unusual packet sizes.

2) Maximal. If all outcomes are likely equal, then the entropy should be maximal. And, also, the entropy increases when the number of outcomes increases. This property is useful in finding the increase in new IP addresses (i.e. the reduced uncertainty would result in lower entropy).

3.2 Modified Information Entropy

We modified the calculation of $H(S)$ to improve the detection accuracy of the proposed scheme. The changed $H(S)$ is given as

$$H(S) = -\sum(SIP_i)P(X_i)\log P(X_i), \quad (1)$$

where SIP_i is source IP address. The motivation behind the modification is to overcome the problems in detection of DoS attacks even if only few source IP addresses are spoofed with equal distribution of packets like the normal traffic. For every source IP address of the network an equivalent ID is maintained by the edge router for the purpose of calculating the entropy.

3.3 Theory of Evidence

Theory of Evidence is a mathematical theory proposed by Arthur P. Dempster and Glenn Shafer and based on belief function and plausible reasoning. This theory combines several pieces of evidences to measure the probability of an event. This theory finds its application in various fields like statistics, decision making, risk analysis etc. [18].

Let S be the universal set of all possible states of the system under consideration (i.e. $s_1, s_2, s_3 \ldots s_n$ belong to S). The system states are mutually exclusive. The power set of S is the set of all possible subsets of S including null set. The elements of power set are used to represent the preposition that system may be in any of the possible subsets. By virtue, the mass function of an empty set is zero (i.e. $m(\Phi) = 0$). The remaining subsets have mass function values that add up to the sum of 1.

$$\sum_{A \subseteq S} m(A) = 1 \text{ and also } 0 \leq m(A) \leq 1.$$

The mass $m(A)$ expresses the proportion of available evidence that claims that the actual state is A, but not to particular subset of A (i.e. the $m(A)$ represents only set A and makes no claim about any of the subsets of A). By definition, each subset of A has its own mass.

3.4 Rule of Combination

Rule of Combination is used to combine the independent sets of mass assignments. This rule emphasizes on the agreement between multiple sources and ignores all the evidences that conflict. The joint mass for 2 masses, say m_1 and m_2, is calculated as

$$\begin{aligned} m_{12}(\Phi) &= 0 \\ m_{12}(A) &= m_1(A) + m_2(A) \\ m_{12}(A) &= \frac{1}{k-1} \sum_{B \cap C = A \neq \Phi} m_1(B) m_2(C), \end{aligned} \quad (2)$$

where $k = \sum m_1(B)m_2(C)$ for all $B \cap C \neq \Phi$ and it is the measure of the amount of conflict between 2 masses and $1 - k$ is the normalization factor, which would ignore conflicts.

The theory of evidence allows for belief about a preposition to be expressed as two values belief and plausibility. These two values are related as:

$$\begin{aligned} Belief(A) &\leq P(A) \leq Plausibility(A), \\ &\text{where } A \text{ is set and} \\ Belief(A) &= \sum_{B \subseteq A} m(B). \end{aligned}$$

The belief for a system to be in a state A is expressed as the sum of all the masses of subsets of A. In other words, it is an evidence for the hypothesis that the system is in state A. The Plausibility(A) is defined as the sum of all the masses of sets B that intersect with A and given as:

$$Plausibility(A) = \sum_{B \cap A \neq \Phi} m(B).$$

These two values can be related to each other as:

$$Plausibility(A) = 1 - Belief(A).$$

In general, the measures of belief and plausibility represent lower and upper bound that supports the given hypothesis.

Using Equation (2), the system state can be inferred by combining the observations that we obtain from the measures of belief that correspond to various hypothesis. Hence the theory of evidence would be useful in deducing the system which is uncertain and dynamic.

4 Detection System for DoS Attacks

The objective of the proposed system is to prevent the suspicious packets from flooding the victim. The best location is at the vicinity of the source network. Hence we prefer to locate our detection mechanism at the edge router. We propose an anomaly based detection mechanism that analyzes the traffic flowing out of the network. The packets contents are used to collect the statistics and analyze. For analyzing the traffic, we consider two parameters namely, source IP address and packet size distribution. In [7], Jung et al. observed that most of the IP addresses were new during bandwidth attacks. The source IP based packet filtering, proposed in [20], monitored the number of new source IP addresses rather than the entire traffic. But there exist ingress and egress filtering methods that examine the source IP addresses. These mechanisms check for the IP addresses that do not belong to the source network. If attackers spoof the IP addresses belonging to the source network, then these schemes do not see any invalid packets. To detect the spoofing attacks, in our proposal, we maintain the list of addresses that are frequently used by source network and also a list of addresses that belong to the source network, but not

being used frequently. The intention of the attackers to use the IP addresses, which belong to source network but being used occasionally, motivated us to maintain two separate lists. Instead of counting number of addresses that are spoofed, we use information entropy to find the abnormal increase in new IP addresses that are witnessed by the edge router. To interpret an unpredictable system, different mathematical formulations and approaches exist. We use information entropy, which is a measure of uncertainty associated with any system which is random and dynamic.

We define a source IP Monitoring Agent (IPMA), which collects the traffic statistics. The monitoring agent observes all the outgoing traffic and maintains a list of frequently used valid IP addresses. The list is updated periodically to reflect the knowledge obtained through new observations by the agent. To infer the information contained in the data received by the agent, we calculate threshold information entropy, $H(S)_{ipthres}$ as expressed in Equation (1) for frequently used IP addresses. This is used as a threshold for finding the increase in new IP addresses. To avoid the false alarm, we define lower and upper values for $H(S)_{ipthres}$.

Then, for every Δt, the information entropy, $H(S)_{ipcalc}$ for all IP addresses received, is calculated again using Equation (1) to infer the current state. If $H(S)_{ipcalc}$ falls behind lower and upper bound of $H(S)_{ipthres}$, then it is understood the attack is on progress. If it happens the attackers spoof the frequently used IP addresses, then $H(S)_{ipcalc}$ falls within lower and upper bound of $H(S)_{ipthres}$, then the attack traffic would be misinterpreted as legitimate. Hence, to further strengthen the detection capability, we add another parameter, the packet size distribution. While analyzing, the network traffic traces, we found that traffic from a source exhibits similar traffic characteristics. One such similarity we found, is the packet size distribution packets from a source have had fixed and usual sizes. While spoofing the valid source addresses, the attackers tend to send the packets with random sizes since they are unaware of valid packet size distribution of the source addresses which they have spoofed. This could be a useful indication for the detection of flooding attacks. To use this detection feature, we use a Packet Size Monitoring Agent (PSMA) and observe the sizes of packets during the statistics collection and maintain a list of valid packet sizes from each source IP address. For the valid packet sizes from every IP address, the monitoring agent calculates the threshold information entropy $H(S)_{ippsizethres}$. Then the lower and upper entropy are set to avoid false alert. For every Δt and each packet size received from every source IP address, $H(S)_{ippsizecalc}$ is calculated using Equation (1). By using the $H(S)_{ippsizecalc}$, the list of hosts whose $H(S)_{ippsizecalc}$ breached the boundaries of $H(S)_{ippsizethres}$ is found. Then, based on these observations, it is concluded that these hosts have involved in flooding attacks.

The system then fuses the evidences or knowledge it collected from the two monitoring agents to interpret the current state of the system. For deciding the action to be taken with regard to the evidences obtained, the system uses the theory of evidence. The monitors provide the evidences for the interpretation of current system state. Each monitor assigns a mass function, m, to the hypothesis that supports or provides an evidence to infer the current system state, based on the calculated information entropy. Mass function (m) is the degree of belief about a hypothesis. The two monitoring agents find the mass functions, m_1 and m_2, which represent the evidence about the spoofed IP addresses and use of invalid packet size. These mass functions are calculated as

$$\begin{aligned} m_1 &= \text{number of spoofed hosts/total number of hosts;} \\ m_2 &= \text{number of hosts who sent invalid packet} \\ &\quad \text{sizes/total number of hosts.} \end{aligned} \tag{3}$$

Periodically, the assigned mass function from each monitor is given to the Decision Agent (DA) that makes decision. For every Δt, the agent collects the mass function from the monitoring agents and makes decision based on the evidences that support particular system state. The model of the proposed system is given below:

1) Monitor the traffic and collect the statistics on source IP addresses and packet sizes from every IP address. Calculate $H(S)_{ipthres}$ for all IP addresses and $H(S)_{ippsizethres}$ for the packet sizes from every IP address. For every Δt.

2) Gather the traffic trends. Calculate $H(S)_{ipcalc}$ and $H(S)_{ippsizecalc}$ for the IP addresses and packet sizes received during Δt. This is done by IPMA and PSMA respectively.

3) If $upper(H(S)_{ipthres}) < H(S)_{ipcalc}$ and $H(S)_{ipcalc} < lower(H(S)_{ipthres})$ then attack is on progress. Use this as evidence E_1.

4) Compare each $H(S)_{ippsizecalc}$ with the bounds of $H(S)_{ippsizethres}$. If any $H(S)_{ippsizecalc}$ exceeds the limits, then the corresponding host is the spoofed host. Use this as E_2.

5) Using Equation (3), assign mass functions to each evidence obtained and fuse them on DA.

6) Based on the mass functions, decide the hosts whose addresses have been spoofed and then reduce the sending rate of those hosts.

5 Implementation and Results

5.1 Experiment Setup

The system described earlier has been implemented to evaluate its performance with regard to the detection of flooding attacks. The Network Simulator (NS2) is used to simulate the realistic flooding attack scenarios and to examine the effectiveness of the detection scheme. To

evaluate the scheme, we applied the network traces we obtained from [16] to generate the traffic on legitimate hosts and attached File Transfer and Constant Bit Rate traffic generator from NS2 for generating TCP and UDP packets on the spoofed hosts. Hence the simulation generates normal traffic and the attack traffic that involves source IP address that are not frequently used by normal traffic and the attack traffic with invalid packet sizes.

5.2 Data Structure

The central data structure is the packet size distribution table (T_{psize}) that stores all the relevant traffic from the network. This table contains an entry for each IP address and the $H(S)_{ippsizethres}$ calculated for all the valid packet sizes from each IP address. The table needs to be updated periodically whenever some valid packet sizes are observed for the IP addresses. And, also a control variable that has the $H(S)_{ipthres}$ for all IP addresses that are valid and frequently used, is maintained and updated periodically.

5.3 Estimation of Information Entropy for IP Addresses and Packet Size Distribution

To calculate the $H(S)_{ipthres}$, we ran the simulation without generating the attack traffic. For all valid IP addresses, we calculated their contribution to the total volume of traffic and then the information entropy using Equation (1). The lower and upper entropy have been also set. Meanwhile, $H(S)_{ippsizethres}$ has been calculated for valid packet sizes for every IP address and stored them in T_{psize}. To evaluate the performance of the proposed system, we ran the simulation with attack traffic having IP addresses that are not used by source network frequently and with invalid packet sizes for some IP addresses. For this run, we calculated $H(S)_{ipcalc}$ and $H(S)_{ippsizecalc}$ for every address. Then, by using these values, we found the hosts involved in attack traffic generation.

5.4 Estimation of Mass Function

We, then, assign mass functions, m_1 and m_2 as described earlier in Equation (3) to deduce system state. In our implementation we considered the following states for the system S (i.e. S ={normal, IP addresses attacks, packet size distribution attacks}). The two monitors, PSMA and IPMA, have been programmed to detect the IP addresses and packet size distribution attacks and assign mass functions that act as evidences for the hypothesis that the system is in one of the two attack states. The two hypothesis for attack states include, h_1 = {flooding with IP addresses} and h_2 = {flooding with invalid packet sizes}. The decision agent combines the reported evidences that support the hypothesis about the system state. By comparing, the mass functions m_1 and m_2, the agent decides the current state about the system and acts accordingly.

To make the system more effective, we updated the T_{psize} table maintained by the PSMA and hence the entropy values based on the periodic observations. Once it has been detected the system is under attack, the packets generated by those suspicious hosts are dropped at the edge router there by reducing the transfer rate of the packets from the sources that contributed to the attack and thus reducing the congestion at the outgoing link and hence at the victim. Thus the traffic burst at the victim is controlled by dropping the packets at the source end that seem to launch the flooding attacks.

5.5 Tuning of Parameters and Performance Evaluation

We ran the simulation extensively to evaluate the effectiveness of the proposed scheme on improving the utilization of network bandwidth and reducing the congestion due to the attack traffic at the outgoing link and hence at the victim. The success of the scheme mainly depends on the choice of various control parameters such as the selection of list of frequently used source addresses, valid packet size distribution for every host, duration (Δt), the rate of reduction of traffic from hosts identified as attackers(δr) and the lower and upper entropy for $H(S)_{ipthres}$ and $H(S)_{ippsizethres}$. We conducted the experiments by setting different values for different parameters. Below, in Tables 1 and 2, we present two scenarios of the system with graphs generated along with the two sets of values for the control parameters. By running the traffic traces from the Internet, we observed the packet size distribution has a set of sizes that is repeatedly used [6]. To find the realistic packet size distribution for our network, we used the traces of the traffic generated by the network under normal traffic.

For the above two settings, we have calculated the legitimate and attack traffic arrived at the victim and shown in the Figure 1. We found for the settings of the entropy values in the Table 1, the performance of the system to be appreciable. This setting provides fair treatment to the legitimate traffic while preventing the attack traffic. These values are set to prevent false rejection. The entropy values are so chosen, whenever the source addresses, not frequently used, are spoofed and/or generate invalid packet sizes, the calculated values for the entropy, $H(S)_{ipcalc}$ and $H(S)_{ippsizecalc}$ would fall behind these bounds. We found this setting by careful learning through various observations empirically. For the settings shown in Table 1, the views of the network are presented below together with comparison of the proposed system with other source end techniques. For the same scenario, the behavior of the network with and without our detection system is shown below. The Figure 2 shows the amount of attack traffic arrived at the victim. This is the case, where both the traffic begin at almost same time.

Figures 3 and 4 show the results where the attack traffic starts sometime later after the normal traffic is on progress and when the normal traffic is about to be over

Table 1: Trial I - First setting of control parameters

S.No.	Parameters	Values
1	Δt	60 units
2	No. of frequently used source addresses	50
3	No. of spoofed hosts	3
4	Drop rate(δr)	0.5
5	$H(S)_{ippsizethres}$	**150**
6	Lower & upper entropy for $H(S)_{ipthres}$	$\pm$ 5
7	Lower & upper entropy for $H(S)_{ippsizethres}$	$\pm$ 150
8	Packet size distribution (T_{psize})	set of valid packet sizes from each host

Table 2: Trial II - Second setting of control parameters

S.No.	Parameters	Values
1	Δt	60 units
2	No. of frequently used source addresses	50
3	No. of spoofed hosts	3
4	Drop rate(δr)	0.5
5	$H(S)_{ippsizethres}$	**100**
6	Lower & upper entropy for $H(S)_{ipthres}$	$\pm$ 10
7	Lower & upper entropy for $H(S)_{ippsizethres}$	$\pm$ 150
8	Packet size distribution (T_{psize})	set of valid packet sizes from each host

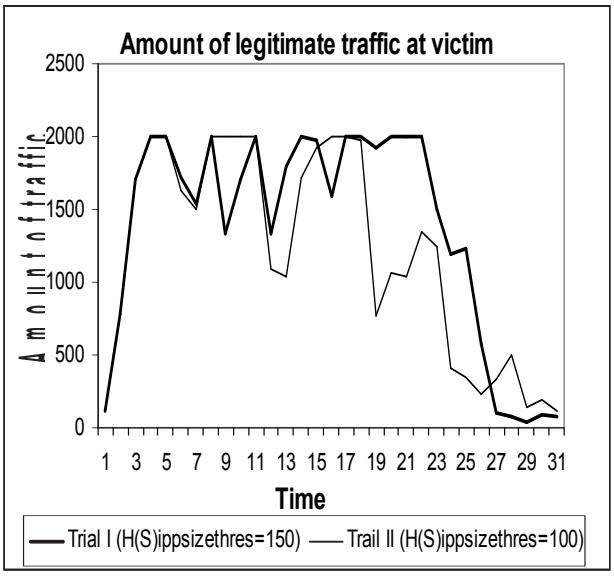

Figure 1: Traffic at victim

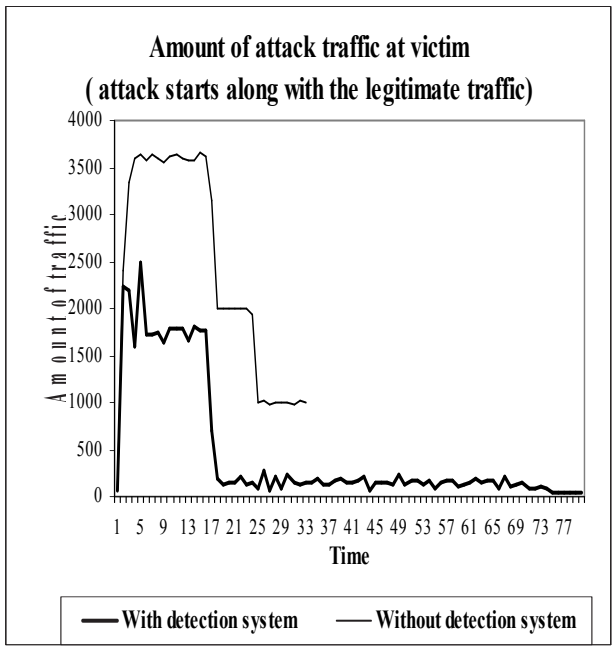

Figure 2: Traffic at victim

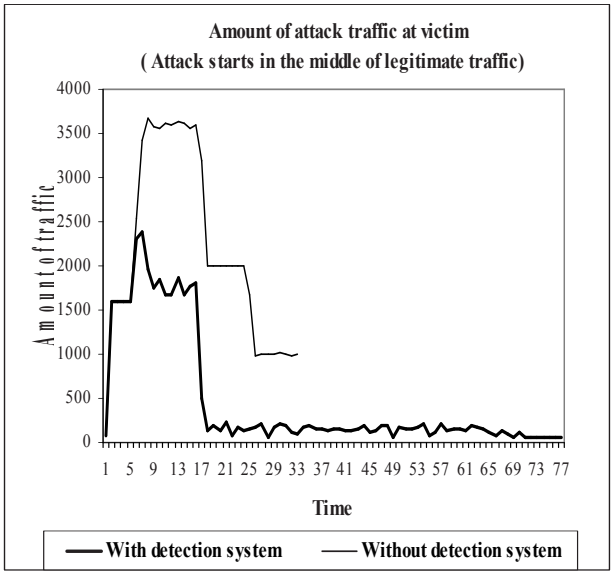

Figure 3: Traffic at victim

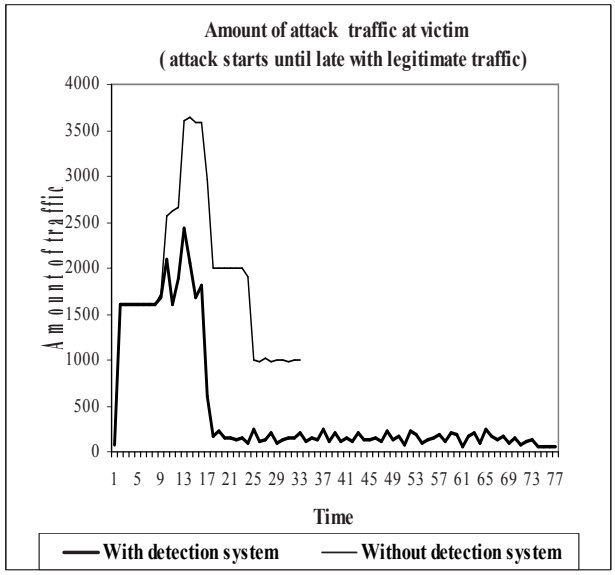

Figure 4: Traffic at victim

respectively.

The above three figures show the amount of attack traffic that is allowed to reach the victim after dropping of the attack packets.

5.6 Summary of Comparison with other Source end Defenses

In the recent past, few source-oriented defenses have been addressed for detecting DOS attack at the edge routers of the network. These systems monitor the outgoing traffic from the source network, gather the statistics and act accordingly. We have compared our system with the systems proposed in [5, 13] and [20]. We ran the above mentioned defenses on the simulator to determine the quantity of the attack traffic filtered by each of them. Each method has been considered in turn to find the traffic dropped during the simulation period, which generated the traffic constituting both legitimate and attack. For simulating the traffic, the traces from NLANR [16] and LBL [6] web sites have been used. We made the comparisons based on the traffic, both legitimate and attack, for both TCP and UDP packets, arrived at the victim and presented them below. First, we present the results of the experiments conducted for TCP packets. The Figures 5 and 6 show the amount of traffic that was allowed to reach the victim by our system and the system in [20]. The graphs shown illustrate that both the methods did not affect much of the legitimate traffic, but the proposed method performs better in regulating the attack traffic also. Reference [20] uses number of new IP addresses as a measure to monitor the attack traffic. If a disgruntled user tends to reside and spoof the address within the source network, then this system could not locate such spoofed attacks. This is so because the spoofed addresses are not new IP addresses, but they belong to the network itself.

A promising technique namely DWARD in [13] has also been considered for comparison. This system detects the DoS attacks based on the responses from the victim and classifies the traffic as attack, suspicious and normal. Based on the classification, the rate cut is imposed on the specific hosts. If the victim is able to send reply back, then the system assumes that the victim is not exposed to attacks. But this need not be true always, that is, the systems of other networks may suffer from degradation of services. Moreover, this system also reduces the sending rate of the legitimate hosts when no response or response below the threshold is seen on the incoming traffic. There is also a chance of reducing the rate of the hosts even when no flooding attacks are aimed at the victim. This would be the case, when the replies are spoofed. The traffic allowed to reach the victim by proposed and DWARD systems are presented the Figures 7 and 8. In this case too, performance of our system is appreciable.

Next we considered another source oriented approach proposed in [5] namely MULTOPS. This system uses disproportional packet rates to and from the hosts and the victim as a heuristic to determine the bandwidth attacks. MULTOPS does not aim in detecting the UDP flooding attacks. As has been shown in the Figures 9 and 10, this system also degrades the service provided to the legitimate hosts in addition to the reduction of the attack traffic.

Similar comparisons were conducted on the UDP flooding attacks. The DWARD in [13] fixes a lower bound on the number of packets from a connection and upper bound on the number of connections to the victim as heuristics to find the flooding attacks. Any traffic breaching these limits would be classified as attacks and rate limit is imposed on those hosts. If the attackers are aware of

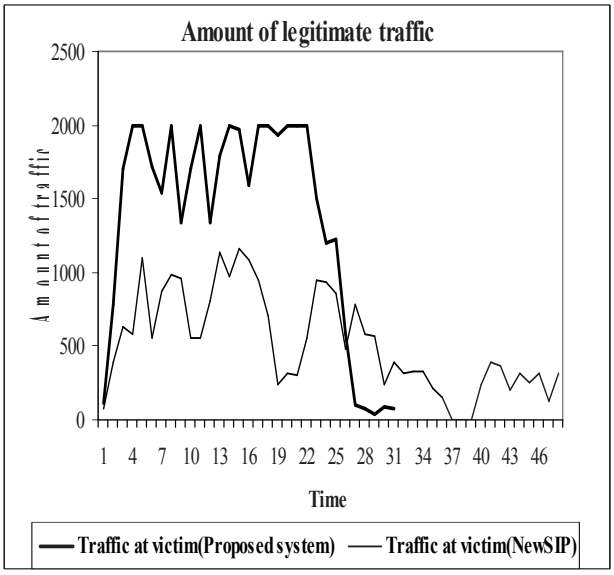

Figure 5: Amount of legitimate traffic at victim (Proposed system Vs NewSIP

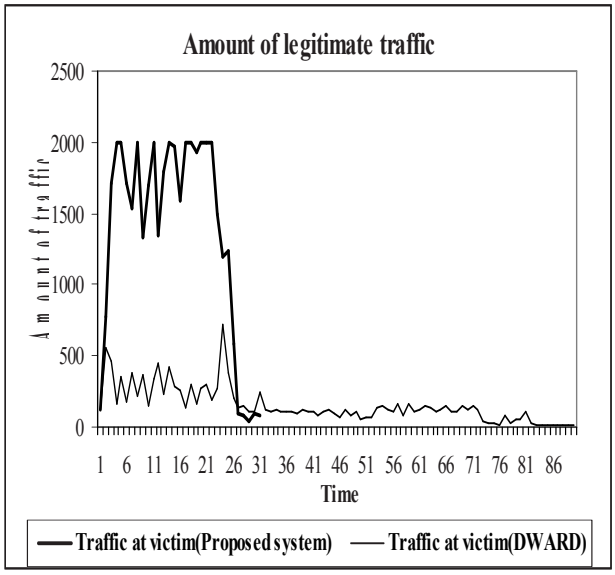

Figure 7: Amount of legitimate traffic at victim (Proposed system Vs DWARD

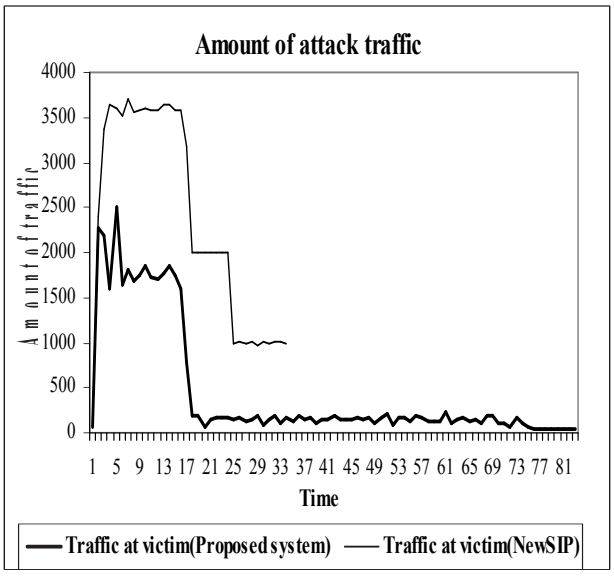

Figure 6: Amount of attack traffic at victim (Proposed system Vs NewSIP

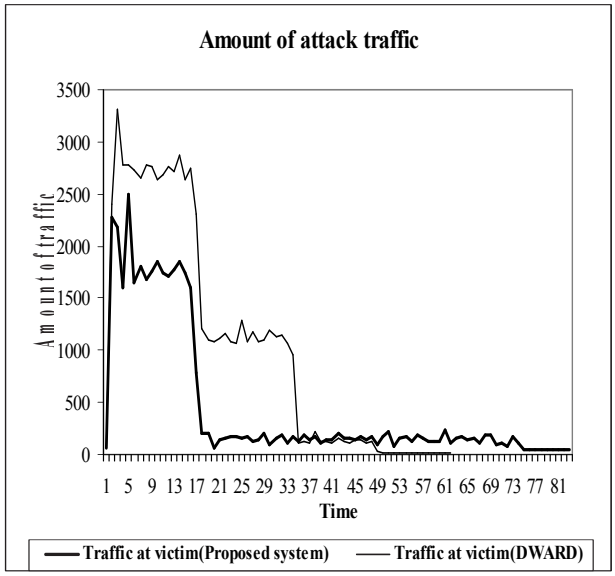

Figure 8: Amount of attack traffic at victim (Proposed system Vs DWARD

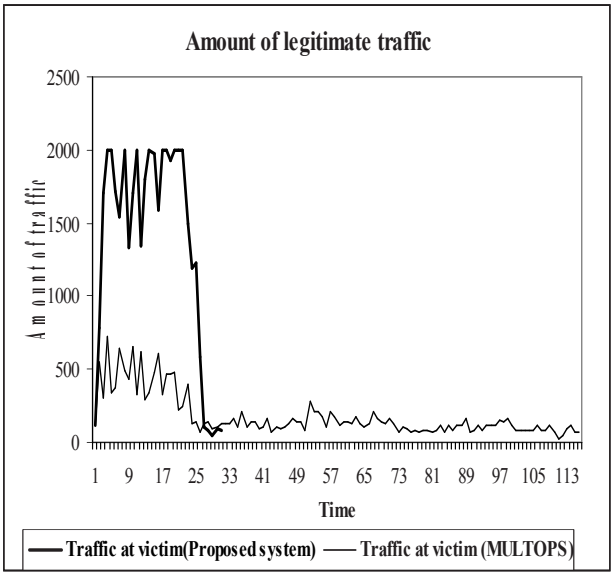

Figure 9: Amount of legitimate traffic at victim (Proposed system Vs MULTOPS

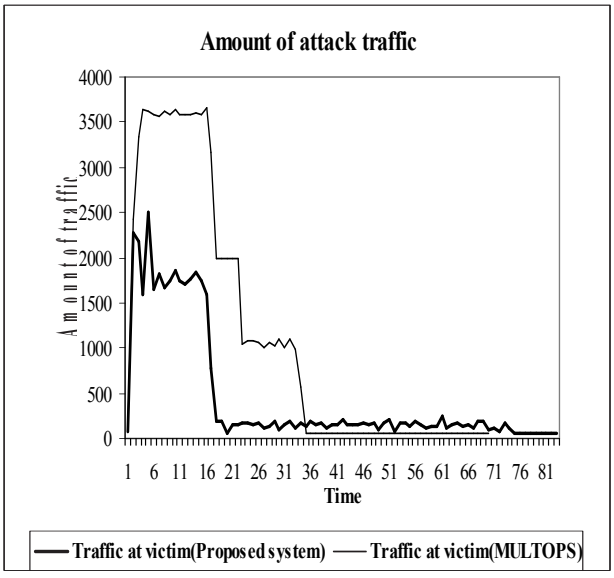

Figure 10: Amount of attack traffic at victim (Proposed system Vs MULTOPS

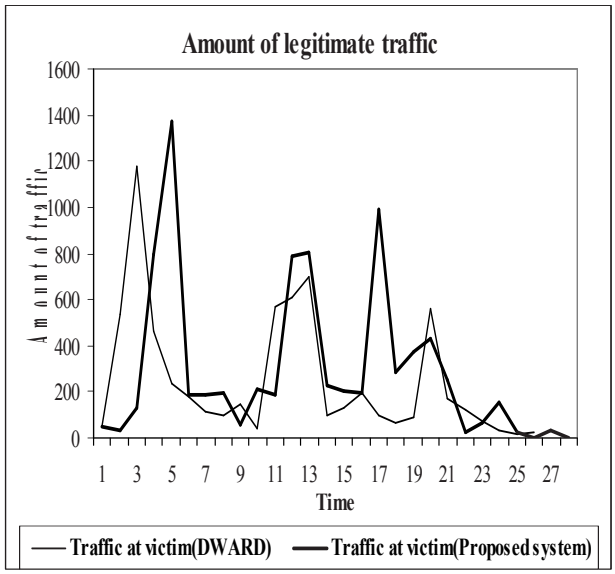

Figure 11: Amount of legitimate traffic (UDP) at victim (Proposed system Vs DWARD

the presence of the detection system, they tend to generate traffic with in these limits, thus flooding the network. Since our system monitors the source IP and packet size distribution and the decision is made based on the calculated entropy for these two variables, even a less amount of UDP flooding would be easily detected. This is shown in the Figures 11 and 12.

5.7 Deployment Gain

Till now, only few systems have been deployed to prevent the hosts from doing harmful actions at source side. Moreover, these systems mostly use 3-way handshaking of TCP communication to detect SYN Flooding, the rate of flow between outgoing and incoming UDP, ICMP, TCP traffic to find the attack. They would fail for spoofed attacks from within the network. This does not mean the source end defense is not beneficial. We have strong reasons for the deployment of our systems at the source end. The victim would be able to detect about flooding only at the bursting stage. By the time it understands flooding, resources would be in scarce. Consequently, it would not be able to issue post detection query. It is a general perception that preventing DoS attacks at the source itself is quite tough, because of the distributed nature of these attacks. But we have studied the behavior of the network carefully and proposed a system that could be deployed at the source end. The system understands the network and uses its interpretations to identify the attacks. Following the identification of the attacks, the traffic from the attackers is dropped, there by reducing the intention of the attackers to flood the network and hence the victim. The system could be especially used to thwart a class of attacks that involves employing unused addresses

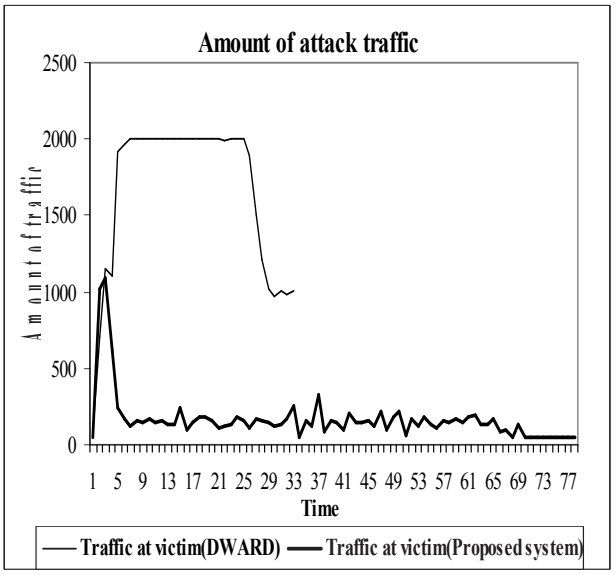

Figure 12: Amount of attack traffic (UDP) at victim (Proposed system Vs DWARD

and unusual packet sizes with objective of degrading the services to the valid requests.

6 Discussion

It is apparent from Section 5 that there are few factors which affect the system's effectiveness. For the scenario presented in Section 5, we considered Δt as 60 units of time. The value of Δt has impact on the quickness of the response. The shorter the Δt, soon we would be able to identify the attack. If it is too large, the attack traffic might flow out and overwhelm at the victim. The selection of $H(S)_{ippsizethres}$ and lower and upper entropy for $H(S)_{ipthres}$ and $H(S)_{ippsizethres}$ also plays vital role. This has been already discussed in Section 5. The improved entropy calculation in Equations (1) ensures the calculated entropy, whenever there are sources addresses found that are not in the frequently used list and used unusual packet sizes, would cross these bounds. With standard way of calculating the entropy, this would not be possible. This is so, if the spoofed traffic has the same amount of traffic like normal, then the $H(S)_{ipcalc}$ for the spoofed traffic would resemble normal. To overcome this, we considered a modification in the entropy calculation and this change has worked out well too.

Care must be taken when the attackers try to spoof the frequently used source addresses. But, our system is resistant to this fact too. In such case, packet size distribution would come for rescue. Thus our system is tolerant to false negatives. Suppose the legitimate traffic itself sends packets with unusual sizes. Then the system generates false positive and drops the packets sent by legitimate hosts. This situation would be handled by monitoring the traffic from those hosts periodically and updating the T_{psize} table using the knowledge obtained so as to help the edge router not to drop the packets from these hosts. This requires further training.

6.1 Deployment Overhead

While analyzing the overhead involved in the implementation of the proposed system, we found there are few issues to be handled well for the proper functioning of the system. One of those aspects includes the list of valid packet sizes from the entire source address space. The size of valid packets may subject to change. Under that condition, the T_{psize} should be updated; otherwise the traffic would be misclassified. But the list having frequently used addresses would not be of much in length. The edge router needs to check every packet that passes through it. This may incur an additional overhead on the router. But this is unavoidable for any source end based defenses.

7 Concluding Remarks

Many solutions have been suggested to handle the flooding attacks which vary in deployment location, filtering techniques etc. A DoS solution would be appreciable if the attacks are detected and prevented as early as possible, but at the same time, providing a fair treatment to the legitimate traffic. In this paper, we proposed such an effective filtering mechanism running at the source end. Our system is dynamic, in the sense that it updates the statistics for filtering, thus minimizing the attack traffic and maximizing the legitimate flow. It automatically detects and rate cuts the flooding attacks. It can be integrated with other defensive systems also.

Our evaluation of the proposed system is oriented towards a single source network with very few parameters. It is not a complete solution. We also have several other parameters for further study. Few of them include, excessive volume of ICMP traffic, high packet fragmentation, too many connections between sources and destinations etc. These aspects may contribute to the future work. We believe our system is one more step towards an automated defense system for detection of DoS attacks.

Acknowledgments

We are greatly indebted to Lawrence Berkeley National Laboratory and National Laboratory for Applied Network Research for having provided the repository of traces of Internet traffic from their site.

References

[1] A. Belenky, and N. Ansari, "IP traceback with deterministic packet marking," *IEEE Communications Letter*, vol. 7, pp. 162-164, Apr. 2003

[2] CERT Coordination Center, *Denial of Service Attacks*. (http://www.cert.org/tech-tips/denial-of service.html)

[3] P. Ferguson, and D.Senie, *Network Ingress Filtering: Defeating Denial of Service Attacks which Employ IP Source Address Spoofing*, RFC 2827, May 2000.

[4] S. Floyd, and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397-413, Aug. 1993.

[5] T.M. Gil, and M. Poletter, "MULTOPS: A data structure for bandwidth attack detection," *Proceedings of USENIX Security Symposium*, pp. 23-38, Aug. 2001.

[6] Internet Traffic Archive, *BC Traces and LBL-PKT Traces*. (http://ita.ee.lbl.gov/html/contrib/BC.html and http://ita.ee.lbl.gov/html/contrib/LBL-PKT.html)

[7] J. Jung, B. Krishnamoorthy, and M. Rabinovich, "Flash crowds and denial of service attacks: Characterization and implications for CDNs and web sites," *Proceedings of 11th World Wide Web Conference*, pp. 293-304, May 2002.

[8] A. Kulkarni and S. Bush, "Detecting distributed denial-of-service attacks using Kolmogorov complexity metrics," *Journal of Network and Systems Management*, vol. 14, no. 1, Mar. 2006.

[9] J. Kurian and K. Sarac, "Defending network-based services against denial of service attacks," *International Journal of Network Security*, vol. 9, no. 2, pp. 186-200, 2009.

[10] M. C. Lee, Y. J. He, and Z. Chen, "Towards improving an algebraic marking scheme for tracing DDoS attacks," *International Journal of Network Security*, vol. 9, no. 3, pp. 204-213, 2009.

[11] S. Liu, Y. Xiong and P. Sun, "On prevention of the denial of service attacks : A control theoretical approach," *IEEE Systems, Man and Cybernetics, Information Assurance and Security Workshop*, West Point, New York, June 2000.

[12] D. McGuire, and B. Krebs, "Attack on internet called largest ever," Oct. 2002. (http://washingtonpost.com/wp-dyn/articles/A828-2002Oct22.html)

[13] J. Mirkovic, and P. Reiher, "D-WARD: A source-end defense against flooding denial-of-service attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 2, no. 3, pp. 216-232, July 2005.

[14] J. Mirkovic, *D-WARD: DDoS Network Attack Recognition and Defense*, Ph.D. Dissertation Prospectus, Computer Science Department, University of California, Los Angeles, Jan. 2002.

[15] P. Mittal, *Defense against Distributed Denial of Service Attacks*, A seminar report, IIT, Guwahati, India, 2005.

[16] National Laboratory for Applied Network Research, *NLANR Packet Traces*. (http://pma.nlanr.net/Traces/traces/long)

[17] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Practical network for IP traceback," *IEEE/ACM Transactions on Networking*, vol. 9, no. 3, pp. 226-237, June 2001.

[18] C. Siaterlis, and B. Maglaris, "Towards multisensor data fusion for DoS detection," *Proceedings of SAC' 04*, pp. 439-446, Nicosia, Cyprus, Mar. 2004.

[19] A. C. Snoren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, B. Schwartz, S. T. Kent, and W. T. Strayer, " Single-packet IP traceback," *IEEE/ACM Transactions on Networking*, vol. 10, no 6, pp. 721-734, Dec. 2002.

[20] T. Peng, C. Leckie, and K. Ramamohanarao, *Detecting Distributed Denial of Service Attacks using Source IP Monitoring*, Draft, Nov. 2002. (http://citeseer.ist.psu.edu/peng02detecting.html)

[21] T. Peng, C. Leckie, and K. Ramamohanarao, "Protection from distributed denial of service attacks using history-based filtering," *Proceedings of ICC 2003*, pp. 11-15, Anchorage, Alsaka, USA, Aug. 2003.

[22] H. Wang, D. Zhang, and K. G. Shin, "Detecting SYN flooding attacks," *Proceedings of Annual Joint Conference of IEEE Computer and Communications Societies (INFOCOM)*, vol. 3, pp. 1530-1539, June 2002.

[23] B. Xiao, W. Chen, and Y. He, "A novel approach for detecting DDoS attacks at an early stage," *Journal on Supercomputing*, vol. 36, pp. 235-248, 2006.

Malliga Subramanian has obtained her Master degree in Computer Science and Engineering from Anna University, Chennai, Tamil Nadu, India in the year 2004. Her area of research interest is network and data security. She is doing her Ph.D. in network security. She has 12 years of teaching experience in the field of Computer Science and Engineering. Currently she is working as a Assistant Professor in the Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Tamil Nadu, India. She has presented papers on the research area in national and international IEEE conferences.

Tamilarasi Angamuthu has obtained her Ph.D. in Algebra in 1994 from the Univesity of Madras. She was awarded JRF by UGC in the year 1986. Presently she is working as Professor in the Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Tamil nadu, India. She has published about 30 papers in national and international journals and conferences. Her areas of interest include semi group theory, fuzzy sets and fuzzy logic. She has been guiding Ph.D. and M.Phil. scholars and is also an approved guide of Anna University, Chennai.