A Framework for Classifying Denial of Service Attacks–Extended*

Alefiya Hussain John Heidemann Christos Papadopoulos
ISI-TR-2003-569b
Date: 25 Feb 2003, Updated: 25 June 2003
{hussain,johnh,christos}@isi.edu

ABSTRACT

Launching a denial of service (DoS) attack is trivial, but detection and response is a painfully slow and often a manual process. Automatic classification of attacks as single- or multi-source can help focus a response, but current packet-header-based approaches are susceptible to spoofing. This paper introduces a framework for classifying DoS attacks based on header content, transient ramp-up behavior and novel techniques such as spectral analysis. Although headers are easily forged, we show that characteristics of attack ramp-up and attack spectrum are more difficult to spoof. To evaluate our framework we monitored access links of a regional ISP detecting 80 live attacks. Header analysis identified the number of attackers in 67 attacks, while the remaining 13 attacks were classified based on ramp-up and spectral analysis. We validate our results through monitoring at a second site, controlled experiments, and simulation. We use experiments and simulation to understand the underlying reasons for the characteristics observed. In addition to helping understand attack dynamics, classification mechanisms such as ours are important for the development of realistic models of DoS traffic, can be packaged as an automated tool to aid in rapid response to attacks, and can also be used to estimate the level of DoS activity on the Internet.

1. INTRODUCTION

The Internet connects hundreds of millions of computers across the world running on multiple hardware and software platforms. It serves uncountable personal and professional needs for people and corporations. However, this interconnectivity among computers also enables malicious users to misuse resources and mount denial of service (DoS) attacks against arbitrary sites.

In a denial of service attack, a malicious user exploits the connectivity of the Internet to cripple the services offered by a victim site, often simply by *flooding* a victim with many requests. A DoS attack can be either a *single-source* attack, originating at only one host, or a *multi-source*, where multiple hosts coordinate to flood the victim with a barrage of attack packets. The latter is called a distributed denial of service (DDoS) attack. Sophisticated attack tools that automate the procedure of compromising hosts and launching attacks are readily available on the Internet, and detailed instructions allow even an amateur to use them effectively.

Denial of service attacks cause significant financial damage every year, making it essential to devise techniques to detect and respond to attacks quickly. Development of effective response techniques requires intimate knowledge of attack dynamics, yet little information about attacks in the wild is published in the research community. Moore et al provide insight into the prevalence of DoS activity on the Internet [26], but their analysis is based on back-scatter packets and lacks the level of detail required to study attack dynamics or generate high-fidelity models needed for DoS research. Monitoring tools today can detect an attack and identify basic properties such as traffic rates and packet types. However, because attackers can forge most packet information, characterizing attacks as single- or multi-source and identifying the number of attackers is difficult.

In this paper, we develop a framework to classify attacks based on header analysis, ramp-up behavior and spectral analysis. First, we analyze the header content to get a rapid characterization of the attackers. Since headers can be forged by the attacker, we develop two new techniques to analyze packet stream dynamics using the ramp-up behavior and the spectral characteristics of the attack traffic. The absence of an initial ramp-up suggests a single attacker, whereas a slow ramp-up (several hundred milliseconds or more) suggests a multi-source attack. Since ramp-up is also easily spoofed, we identify spectral characteristics that distinguish single- from multi-source attacks and show that attackers cannot easily spoof spectral content without reducing attack effectiveness. We describe the algorithms used in our framework in Section 4 and discuss robustness to counter-measures in Section 7.

The contribution of this paper is an automated methodology for characterizing DoS attacks that adds new techniques of ramp-up and spectral analysis, building on the existing approach of header analysis. In addition to providing a better understanding of DoS attack dynamics, our work has several direct applications. This identification framework can be used as part of an automated DoS detection and response system. It can provide the classification component of a real-time attack analysis system to aid network administrators in selecting an appropriate response depending on the type of ongoing DoS attack. For example, if an attack consists of only a single source using traceback to identify the culprit is trivial, but as the number of attackers increase traceback becomes rapidly intractable. Thus one application of our framework is to judiciously decide if activation of traceback is appropriate during a particular attack. This analysis can also be used to create and validate models of DoS and DDoS attacks for simulation and experimentation. Finally, long-term automated measurements of DoS attacks can be used to estimate the level of DoS attack activity in the Internet. We describe these applications in Section 8.

We evaluated our framework on traffic collected from two peering links at Los Nettos, a regional ISP in Los Angeles. Over a period of five months we observed and analyzed 80 attacks. We could classify 67 attacks as single- or multi-source with header analysis; the remaining 13 attacks were classified based on ramp-up and

*This paper is an extended version of the orginal paper that will appear in SIGCOMM 2003, Karlsruhe, Germany. The research is based on work supported by DARPA via the Space and Naval Warfare Systems Center San Diego under Contract No. N66001-00-C-8066 ("SAMAN"), by NSF under grant number ANI-9986208 ("CONSER"), by DARPA via the Fault Tolerant Networks program under grant number N66001-01-1-8939 ("COSSACK") and by Los Alamos National Laboratory under grant number 53272-001.

spectral behavior. We validate our algorithm and conclusions in three ways. First, we monitor a second site at University of Southern California and compare the observed attack dynamics. Second, to understand the spectral characteristics of attacks we conduct a series of experiments with synthetically generated attack traffic sent over a wide-area network and with real attack traffic generated using attack tools on an isolated testbed. Finally, we use simple numerical simulations to improve and confirm our understanding of the underlying causes for differences in spectral behavior. Our validation methodology is detailed in Section 6.

2. RELATED WORK

Denial of service attacks attempt to exhaust or disable access to resources at the victim. These resources are either network bandwidth, computing power, or operating system data structures. Research on denial of service attacks is primarily focused on attack detection and response mechanisms. Attack detection identifies an ongoing attack using either anomaly-detection [15, 27, 41] or signature-scan techniques [29, 31]. Most response mechanisms attempt to alleviate the damage caused by the attack by taking reactive measures like reducing the intensity of the attack by blocking attack packets [19, 23, 27], or localizing the source of the attack using traceback techniques [4, 9, 32, 33, 34]. Besides the reactive techniques discussed above, some systems take proactive measures to discourage DoS activity, for example, both CenterTrack [36] and SOS [21] use overlay techniques with selective re-routing to prevent large flooding attacks. In this paper, we use a simple anomaly-detection technique that that identifies an attack if the number of sources connecting to the same destination exceeds a pre-defined threshold. We develop a unique framework to identify single- and multi-source attacks that can be used as part of an automated attack detection tool to select an appropriate response mechanism. The additional information provided by our framework can be used to decide the location of filters and if traceback should be activated.

Many techniques have been proposed to detect an ongoing DoS attack. Cisco routers provide support for attack detection making use of RMON [38] and Netflow [37] data, that can be processed offline to detect an attack. Multops exploits the correlation of incoming and outgoing packet rates at different level of subnet prefix aggregation to identify attacks [15]. Wang provides a rigorous statistical model to detect abrupt changes in the number of TCP SYN packets as compared to the TCP SYN ACK packets [41]. Bro, an intrusion detection system uses change in (statistical) normal behavior of applications and protocols to detect attacks [29] while Cheng use spectral analysis to detect high volume DoS attack due to change in periodicities in the aggregate traffic [10]. All the above techniques are based on *anomaly-detection* which is faster than static *signature-scan* techniques used by Snort [31]. Snort has one main disadvantage; new attacks that do not have well-defined signatures may go undetected until the signature is defined.

Response to an attack consists of localizing the attackers and reducing the intensity of the attack. The SPIE system can traceback individual packets within a domain using packet digests [33]. On the other hand, Burch and Cheswick propose a technique to traceback to the source by flooding routes to the victim and observing change in the attack rates [9]. IP Traceback [32, 11, 34] and ICMP traceback [4] provide mechanisms to identify the source of the attack using packet marking at routers. Most of these mechanisms require large scale deployment over the Internet to be effective and as the number of attackers increase, the number of packets and computational time required to identify the attacker increases drastically (SPIE is the exception). In this paper we propose a framework to identify the presence of single- or multi-sources in an attack based on local attack stream information. If an attack consists of only a single attackers, using traceback to identify the culprits is trivial, but as the number of attackers increase traceback becomes rapidly intractable. Thus the additional information provided by our framework can be used to judiciously decide the response mechanism.

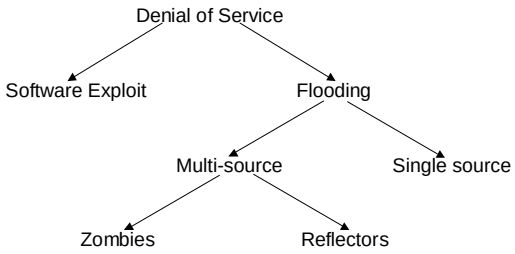


Figure 1: Classification of DoS attacks based on volume of packets and number of attackers. In this paper we analyze flooding attacks.

To reduce the intensity of an attack, Mahajan et al propose an aggregate congestion control and pushback technique to identify and throttle the attack flows [23]. Pushback is a cooperative technique that allows routers to block an aggregate upstream. On the other hand, D-WARD uses TCP-based rate control at the first hop to prevent attackers from participating in an attack [19]. Packet filters are the best line of defense during an attack [43]. Filtering decisions are typically based on source and destination addresses, port numbers or packet contents. Once an attack is classified as single or multi-source using the proposed framework, network operators can strategically deploy packet filters to block the attack packets.

Beside attack detection and response mechanisms, it is important to understand DoS attack prevalence and attack dynamics on the Internet. Moore et al used backscatter analysis and detected 12,805 attacks during a period of 3 weeks [26]. The backscatter technique allows detection of attacks that uniformly spoof source addresses in the complete IP address space. Many attack tools use reflection techniques, subnet spoofing, or do not spoof source addresses [16, 30]. The backscatter technique will not detect these attacks. In this paper, we develop an alternate approach where we extrapolate the attack activity observed at Los Nettos to the Internet (Section 8.3).

Signal processing technique have been used previously to analyze malicious network traffic. Cheng et al use spectral analysis to detect high volume DoS attack due to change in periodicities in the aggregate traffic [10] while Barford et al use flow-level information to identify frequency characteristics of DoS attacks and other anomalous network traffic [3]. Further, wavelets and other signal processing techniques have been extensively used to analyze both wired and wireless network traffic [8, 28, 42]. Prior research in the DoS area has applied signal processing to detect ongoing attacks. In this paper we analyze the spectral behavior of the attack stream to provide information regarding the presence of multiple attackers.

3. ATTACK TAXONOMY

To launch a DDoS attack, a malicious user first compromises Internet hosts by exploiting security holes, many of which are openly disclosed by software vendors. The malicious user then installs attack tools on the compromised host (also known as a *zombie*), that now becomes available to attack any victim on command. With full control of the zombie the attacker can construct any packet including illegal packets, such as packets with incorrect checksums, incorrect header field values, or an invalid combination of flags.

The different types of denial of service attacks can be broadly classified into *software exploits* and *flooding attacks*. Flooding attacks can be further classified into single- and multi-source attacks based on the number of attackers. This classification is depicted in Figure 1 and explained next.

3.1 Software Exploits

These attacks exploit specific software bugs in the target's operating system or applications, and can potentially disable the victim machine with a single or a few packets. A well known example is the *ping of death*, that causes the operating system to crash by sending a single large ICMP echo packet. Similarly, the *land attack* sends a single TCP SYN packet containing the victim's IP address in both the source and destination address fields, resulting in an endless loop in the protocol stack. Such attacks can only be prevented by diligently applying software updates. While software-exploit attacks are important, this paper focuses on flooding attacks, since they cannot be addressed by software fixes.

3.2 Flooding attacks

Flooding attacks are the result of one or more attackers sending incessant streams of packets aimed at overwhelming link bandwidth or computing resources at the victim. Based on the location of the observation point, we classify flooding attacks as single-source attacks when a single zombie is observed flooding the victim, and as multi-source when multiple zombies are observed, as shown in Figure 2(a) and Figure 2(b) respectively. In both cases, we may misclassify sources if our observation point misses some zombies such as the dotted zombies. Multiple attackers may be summoned for an attack to increase firepower, or to evade detection. In both attack classes, the master can install attack tools on the host machine that can generate illegal packets. Examples include the TCP NULL attack that generates packets with no flags set, the Xmas attack that has all TCP flags set, and attacks that use packets with a non-existent IP protocol number [2]. Several canned attack tools are available on the Internet, such as Stacheldraht, Trinoo, Tribal Flood Network 2000, and Mstream that generate flooding attacks using a combination of TCP, UDP, and ICMP packets [12]

A significant percentage of captured attacks consist of a single source. Moore et al detected 14% of all DoS attacks were directed toward home machines using either dial-up or broadband access [26]. CERT also reports most DoS attacks on the Internet are from a single source to a single victim [16]. Thus, a single high bandwidth zombie can potentially generate enough packets to overwhelm a victim.

The third type of attack is the *reflector* attack (Figure 2(c)). Such attacks are used to hide the identity of the attacker and/or to amplify an attack [30]. A reflector is any host that responds to requests, such as web servers or ftp servers, that respond to TCP SYN requests with a SYN-ACK reply, or hosts that respond to ICMP echo requests with ICMP echo replies. Servers may be used as reflectors by spoofing the victim's IP address in the source field of the request, tricking the reflector into directing its response to the victim. Unlike direct zombie attacks, reflector attacks require well-formed packets to solicit a reply. If many reflector machines are employed, such an attack can easily overwhelm the victim without adversely affecting the reflectors or triggering the local IDS. Reflectors can also be used as amplifiers by sending packets to the broadcast address on the reflector network, soliciting a response from every host on the LAN. Unlike zombies which represent improperly secured hosts, reflectors are often hosts intentionally providing Internet services, and so reflector attacks may be more difficult to prevent.

Let P ={attack packets}, $P_i \subset P, P = \bigcup_{i=2}^{n} P_i$
If $\forall\, p \in P$
    ID value increases monotonically **and**
    TTL value remains constant
    then **Single-source**
elseif $\forall\, p \in P_i$
        ID value increases monotonically **and**
        TTL value remains constant
        then **Multi-source with n attackers**
else **Unclassified**

Figure 3: Pseudo code to identify number of attackers based on header content.

4. ATTACK CLASSIFICATION

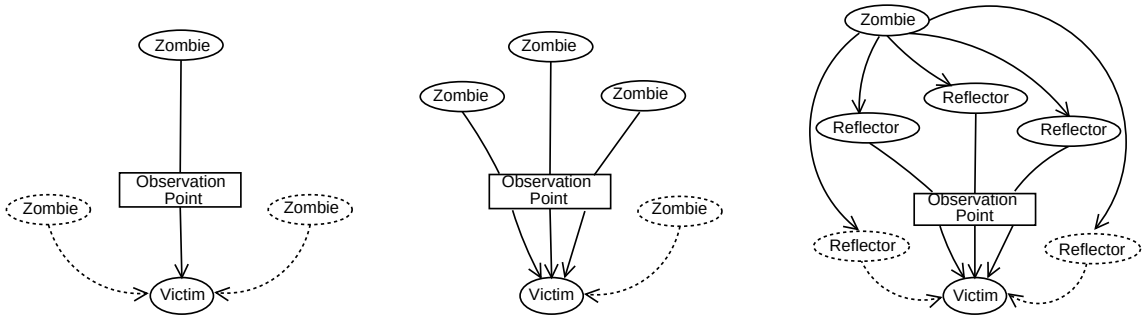
Our framework classifies attacks using header contents, transient ramp-up behavior, and spectral characteristics. This three-pronged approach is necessary to deal with an increasing level of difficulty in classifying attacks depending on the level of IP header spoofing present in an attack. If the source address in the attack packets is not spoofed, classifying an attack as single- or multi-source becomes a simple matter of counting the distinct sources present in the attack stream. When the source address is spoofed, we must look at other header fields (such as ID and TTL) for clues. Finally, when the entire IP header is spoofed, we resort to ramp-up and spectral analysis for classification. Next, we describe these stages in more detail.

4.1 Header Contents

Most attacks spoof the source address concealing the number of attackers. However, other header fields, such as the fragment identification field (ID) and time-to-live field (TTL), can be indirectly interpreted to provide hints regarding the number of attackers. Such techniques have been used before to identify multiple interfaces on routers [35] and count number of hosts behind a NAT box [5]. These techniques work because many operating systems sequentially increment the ID field for each successive packet. As a result, all packets generated by the same host will contain monotonically increasing ID values. In addition, assuming the routes remain relatively stable during the attack, the TTL value will remain constant for the same source-destination pair. Thus for attacks where the ID and TTL fields are not forged we use the algorithm outlined in Figure 3 to estimate the number of attackers and classify attacks as single- or multi-source.

We estimate the number of attackers by counting the number of distinct ID sequences present in the attack. Packets are classified as belonging to the same sequence if their ID values are separated by less than *idgap* (we use an *idgap* of 16) and the TTL value remains constant for all packets. We allow for some separation in idgap to tolerate moderate packet reordering. In high volume attacks the ID value typically wraps around within a second. Therefore using a small idgap also limits collisions during sequence identification. If a packet does not belong to an existing sequence, it forms the beginning of a new sequence. In most cases, attack packets arrive close to each other and have a idgap of one. An attack sequence must consist of at least 100 packets to identify a distinct attacker.

Some attacks have short silence periods during the attack. After a silence period, packets may form a new attack sequence that should be considered as a continuation of an old sequence, but would not be identified as such due to the strict idgap. To bridge these silence periods we coalesce such streams into one stream if they are within 500ms of each other. Finally, since many operating systems do not send the ID value in network byte order, we infer byte-order from



(a) Single-source (b) Multi-source (c) Reflector

Figure 2: Flooding attacks are classified as (a) single-source, (b) multi-source, or (c) reflected based on the number of attackers and their location, with respect to the observation point and victim.

the first 10 packets observed.

Many attack tools spoof the source IP address but allow the operating system to fill in default values for other fields [12]. These tools are susceptible to ID analysis. We are not aware of any attack tools that attempt to coordinate the ID field over a distributed set of attackers. In fact, differences in RTT and available bandwidth make it inherently difficult to coordinate packet streams from multiple hosts such that their ID fields consistently arrive in order without reducing the rate (and hence effectiveness) of the attack.

Some attack tools forge all header contents, including both the ID and the TTL field. For such attacks it is impossible to distinguish between a single or multiple sources based on header information alone, making it essential to use additional techniques.

4.2 Ramp-up Behavior

In a multi-source attack, a master typically activates a large number of zombies by sending a trigger message that either activates the zombies immediately or at some later time. When observed near the victim, this distributed activation of zombies results in a *ramp-up* of the attack intensity due to the variation in path latency between the master and the zombies and weak synchronization of local clocks at the zombies. In contrast, single-source attacks do not exhibit a ramp-up behavior and typically begin their attack at full strength. Thus, the presence of a ramp-up provides a hint as to whether the attack is single- or multi-source. This method cannot robustly identify single-source attacks since an intelligent attacker could create an artificial ramp-up from a single site. To our knowledge, current attack tools do not attempt to do so.

4.3 Spectral Analysis

A more robust method for classifying attacks as single- or multi-source is to consider their spectral characteristics. We observed attack streams have markedly different spectral content that varies depending on the number of attackers. In this section, we present our methodology for analyzing the spectral characteristics of an attack stream; in Section 5.5 we present several examples with intuition why it works.

Spectral analysis requires treating the packet trace as a time series. We divide the attack stream into 30 second segments, defining $x(t), 0 \leq t < 30,000$ as the number of attack packet arrivals in each 1ms interval. Since non-stationarity can taint spectral analysis, we discard segments that show initial ramp-up or abrupt changes (perhaps due to a change in number of attackers). We use linear least-square regression to compute the slope of $x(t)$ and verify that the difference between the slope and zero is statistically insignificant within a 95% confidence interval [6]. Further, we condition $x(t)$ by subtracting the mean arrival rate before proceeding with spectral analysis. The mean value results in a large DC component in the spectrum that does not provide any useful information for our classification framework.

For stationary segments, we compute the power spectral density by performing the discrete-time Fourier transform on the autocorrelation function (ACF) of the attack stream. The autocorrelation of an attack stream is a measure of how similar the attack is to itself shifted in time by offset k [6, 7]. When $k = 0$ we compare the attack stream to itself, and the autocorrelation is maximum and equal to the variance of the attack stream. When $k > 0$ we compare the attack stream with a version of itself shifted by lag k. The autocorrelation sequence $r(k)$ at lag k is

$$c(k) = 1/N \sum_{t=0}^{N-k} (x(t) - \bar{x})(x(t+k) - \bar{x}); \quad (1)$$

$$r(k) = c(k)/c(0) \quad (2)$$

where $\bar{x}$ is the mean of $x(t)$ and N is the length of the attack stream $x(t)$. The power spectrum $S(f)$ of attack obtained by the discrete-time Fourier transform of the autocorrelation sequence of length M:

$$S(f) = \sum_{k=0}^{M} r(k) e^{-\imath 2\pi f k} \quad (3)$$

The highest frequency observable by this procedure is 500Hz, since we consider 1ms intervals and the Fourier transform is symmetric. Intuitively, the spectrum $S(f)$ captures the *power* or strength of the attack stream contains at a particular frequency.

Once we generate the spectrum, we need a technique to compare the spectral characteristics of different attacks. Therefore, for each attack we define the cumulative spectrum $P(f)$ as the amount of power in the range 0 to f. We normalize $P(f)$ by the total power to get the normalized cumulative spectrum (NCS), $C(f)$ [7]. Finally, we define quantile $F(p)$ as the frequency at which the NCS captures p percent of the power. Formally:

$$P(f) = \sum_{i=0}^{f-1} \frac{(S(i) + S(i+1))}{2}; \quad (4)$$

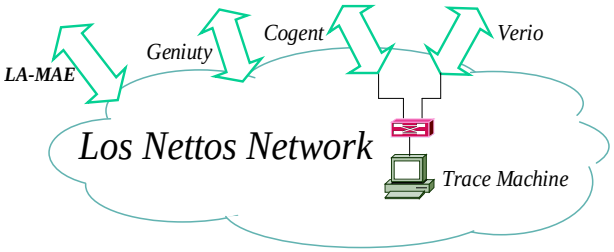


Figure 4: The trace machine monitors two of the four peering links at Los Nettos.

$$C(f) = \frac{P(f)}{P(f_{max})}; \tag{5}$$

$$F(p) = \min_{0 \le f \le f_{max}} f \text{ such that } C(f) \ge p \tag{6}$$

We use $F(p)$ as a numerical method of comparing power spectral graphs. The key insight is that multi-source attacks shift spectrum to lower frequencies. To quantify this, we pick a quantile of 60% of the power and compare the $F(60\%)$ values of attacks. Our observations indicate single-source attacks have a linear cumulative spectrum due to dominant frequencies spread across the spectrum. This causes $F(60\%)$ to be in the range of 240–296Hz. In contrast, multi-source attacks have localization of power in lower frequencies resulting a $F(60\%)$ in the range of 142–210Hz. Although we use a 60% quantile, our choice is somewhat arbitrary. The important characteristic is that it capture the trend in frequency distribution of the spectra. Our results are insensitive to the particular choice of quantile as examined in detail in Section 7.

5. EVALUATION

In this section we present our trace collection infrastructure and our experimental analysis based on attack captured at Los Nettos. Validation of these results is presented the next section.

5.1 Attack Detection

We tested the framework described in Section 4 using attacks captured at Los Nettos, a moderate size ISP located in Los Angeles [22]. We captured 80 large-scale attacks over a period of five months, from July 2002 to November 2002.

Los Nettos has four major peering links with commercial ISP providers. Due to lack of available mirroring capacity, we were able to monitor only two links, as shown in Figure 4. Los Nettos has a diverse clientele including academic and commercial customers. The trace machine is an off-the-shelf Intel P4 1.8GHz, with 1GB of RAM running FreeBSD 4.5. We use a Netgear GA620 1000BT-SX NIC, and modified the driver to support partial packet transfer from the NIC to the kernel. Typical daytime load is 140Mb/s with a mean of 38Kpackets/s. Measurement drops (as reported by tcpdump) are usually below 0.04% during normal operation, rising to 0.6% during attacks that increase packet rates to 100Kpackets/s.

We continuously capture packet headers using tcpdump, creating a trace file every two minutes. Each trace is then post-processed and flagged as containing a potential attack if either of two thresholds are reached: (a) the number sources that connect to the same destination within one second exceeds 60, or (b) the traffic rate exceeds 40Kpackets/s. These thresholds were determined by observing the traffic seen at the observation point. Traces that are not flagged as an attack are discarded. We identify and ignore known servers that would trigger these thresholds through normal traffic. Finally, we manually verify each flagged trace to confirm the presence of an attack. The automated thresholding works reasonably well, but provides a false positive rate of 25–35%. Ongoing attacks that do not meet the thresholds are not identified. We thus miss many small DoS attacks, including some attacks that would incapacitate a dial-up line.

We monitor both inbound and outbound traffic. Since we monitor the two busiest peering links, we believe we capture most of the attack traffic for attacks terminating within Los Nettos, missing only portions from the peering links we do not monitor and from attackers within Los Nettos. For attacks transiting through Los Nettos, our monitoring point may not be exposed to the full intensity of the attack since there may be attackers outside Los Nettos and we do not monitor all external links of Los Nettos. The distinction between transient and terminating attacks becomes important when projecting numbers of attacks in Section 8.3.

Attack Class	# Attacks	Range (packets/s)	Range (kbits/s)
Single-source	37	350–82500	2700–93000
Multi-source	10	300–98000	17000–100000
Reflected	20	340–13000	3000–33000
Unclassified	13	400–68500	12000-66000

Table 1: Number of attacks in each class based on header analysis

Protocol	Packet Type	Attack Class			
		S	M	R	U
TCP	SYN	2	3 (2)	-	7 (5)
	ACK	5	2 (2)	-	3 (2)
	SYN-ACK	9	-	4	-
	no flags	15	1 (1)	-	-
	unusual	5	1	-	-
	state exploit	2	-	-	-
ICMP	echo request	5	-	-	-
	echo reply	1	-	16 (3)	-
	invalid	-	-	-	1 (1)
UDP	all	6 (1)	-	-	5 (4)
Other	ip-proto 0	5	-	-	-
	ip-proto 255	-	3	-	-
	fragmented	1	-	-	3 (3)

Table 2: Detailed analysis of packet headers. *S* indicates single-source, *M* indicates multi-source, *R* indicates distributed reflectors, and *U* indicates unclassified attacks. The number in parenthesis indicates attacks terminating within our ISP while the first number indicates total attacks.

5.2 Packet Headers Analysis

First we classify attacks based on packet header information alone. As shown in Table 1, we classified 67 attacks (all but 13) using this method. Table 2 shows a more detailed breakdown of attacks based on manual analysis with tcpdump. The packet type categories listed in Table 2 are not mutually exclusive since some attack streams carry multiple packet types.

From header analysis we can make several observations about the prevalence of attack techniques in the wild. First, 87% of the zombie attacks use illegal packet formats or randomize fields, indicating the presence of root access on the zombies. Use of TCP protocol was most common, with reflection attacks typically exploiting web servers (port 80) and FTP servers (port 21). In Table 2, TCP *no flags* refers to pure data packets with no flags set, while *unusual* refers to attacks that use non-standard (but not always invalid) com-

binations of TCP flags, such as setting all the flags. *State exploit* refers to attacks that exhaust OS data-structures based on the TCP-state diagram, (such as ESTABLISHED or FIN-WAIT1 states) [1]. Even though TCP-SYN attacks belong to this class, we list them separately since they are common.

ICMP is the next protocol of choice. The echo reply attack was the most popular reflector attack, since most Internet hosts respond to an echo request packet allowing the attacker to choose from the large number of possible reflectors. The remaining ICMP attacks use echo request packet or an *invalid* ICMP code. UDP and undefined protocols were less frequently used in the attacks. Finally, we detected five attacks that use a combination of protocols, such as TCP, ICMP, UDP, and IP proto-0.

5.3 Arrival Rate Analysis

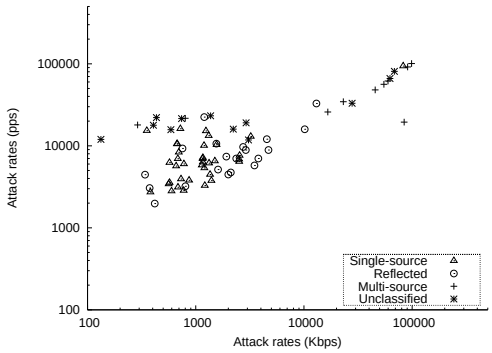
This section investigates the relation between attack rate and attacker population. We captured attacks with peak rates ranging from 300packets/s to 98Kpackets/s. Figure 5 shows the correlation between the attack classes and attack rate. In Figure 5(a) we plot the peak attack rates in Kbits/s on the x-axis against packets/s on the y-axis in logarithmic scale for each attack. Not surprisingly, single-source attacks are clustered toward lower packet rates whereas direct multi-source attacks exhibit higher rates, most likely due to aggregation of traffic from multiple zombies. In reflection attacks, many reflectors are typically employed to generate high attack aggregates without overloading the reflectors. The captured reflection attacks have a much lower intensity than direct multi-source attacks since the observation point might not be exposed to the complete intensity of the attack.

To statistically confirm attack rates of single-, multi-source, and reflected attacks have different means, we performed Kruskal-Wallis one-way ANOVA test [6]. We consider the null hypothesis, H_0; there is no relation between the attack rates and attack class. The alternative hypothesis, H_a states there is a relation between attack rate and class. If H_0 is true, the variance estimate based on within-class variability should be approximately the same as the variance due to between-class variability. This test defines a *F ratio* that evaluates the two variance estimates; if the F ratio is significantly greater than 1, the test is statistically significant, and we can conclude that the means for the three classes are different from each other and reject H_0. It also defines a *p*-value, the probability of observing the sample result assuming H_0 true. Hence a smaller p-value provides higher confidence in rejecting H_0. For the data in Figure 5(a), the F ratio is 37, indicating a strong relation between the attack rates and the attack classes. Further, the p-value is 1.7×10^{-11}, indicating a very low probability of H_0 being correct.

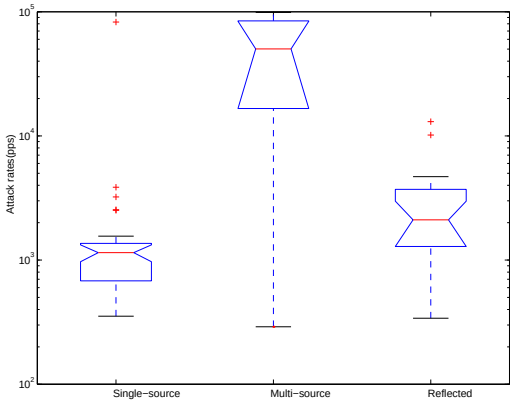
The box plot in Figure 5(b) provides graphical representation of the means of different classes. The lower and upper lines of the box indicate the 25th and 75th percentiles of attack rates making the distance between the top and bottom of the box is the interquartile range The line in the middle of the box is the median attack rate for the attack class. The "whiskers" (lines extending above and below the box) show the range of the attack rate, except for the outliers indicated by a "+" marker. Single-source attacks have the lowest median while the median and range of the multi-source attacks is significantly higher than single-source and reflected attacks. Further analysis of attack rates along with cummulative distribution graphs are provided in Appendix A.

5.4 Ramp-up Behavior Analysis

To identify the presence of multiple sources when the header is forged we measure the attack's ramp-up behavior (changes in the traffic volume of the attack as a function of time). Of the attacks



(a) Attack intensity in packets/s and Kbits/s



(b) Attack intensity in packets/s for each class

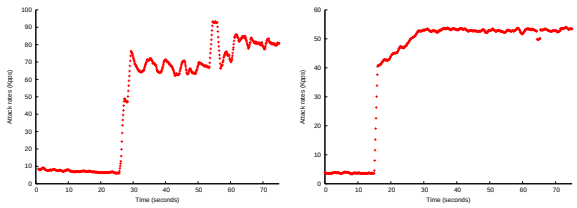
Figure 5: Correlation of attack rates and attack class

we observed, single-source attacks typically exhibit no ramp-up, while all multi-source attacks showed ramp-up behavior, ranging from 200ms to 14s.

Figure 6 illustrates the attack ramp-up for two observed attacks. Figure 6(a) shows an attack where packet headers were not forged, and thus the attacker population was visible. The graph shows a three second ramp-up at about 27s as the number of attackers gradually increase to six. The attack reaches a peak rate of 78Kpackets/s with 14 active sources. We observe a total of 40 unique IP addresses during the attack. Figure 6(b) shows an attack where the last eight bits of the source address are forged. The attack is classified as a multi-source attack since it exhibits a ramp-up, rising from 6Kpackets/s to 52Kpackets/s in 14 seconds. In this attack the source addresses and ID field is spoofed, and all packets have the same TTL value, making it difficult to classify the attack based on header content. The presence of transient ramp-up behavior in the first few seconds of the attack strongly suggests the presence of multiple sources. We also verified it is a multi-source attack via spectral analysis.

5.5 Spectral Content Analysis

In this section we demonstrate that spectral analysis of the attack time-series (described in Section 4.3) can distinguish between single- and multi-source attacks, even if all headers are spoofed. Because the traffic spectrum is influenced by OS and network behavior we argue that it will be difficult for attackers to easily con-



(a) Multiple source addresses observed in attack

(b) Subnet spoofed source addresses

Figure 6: Due to lack of synchronization among the zombies, multi-source attacks exhibit initial ramp-up behavior

ceal their spectrum without reducing attack effectiveness. We review this claim more carefully Sections 6 and 7, for now we present example spectra to illustrate the technique.

We analyze the spectral content of all 67 attacks previously classified by header analysis. Based on observations from these known classes, we conclude that single- and multi-source attacks can be distinguished by their spectra:

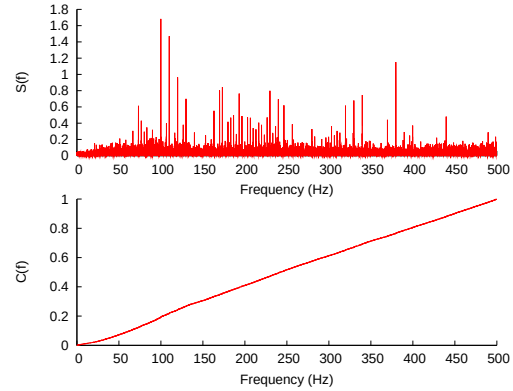
- Single-source attacks include dominant high frequencies creating a linear trend in the normalized cumulative spectrum.
- Multi-source attacks have dominant low frequencies causing the normalized cumulative spectrum to sharply rises at lower frequencies.

Figure 7(a) shows an example of the spectrum of a single-source attack. In this case, the attacker that generates TCP no flag packets at a rate of 1100packets/s. The source addresses are spoofed, but the ID and TTL values clearly indicate a single-source attack (using analysis from Section 4.1). There are noticeable peaks at higher frequencies in the spectrum and the NCS is linear.

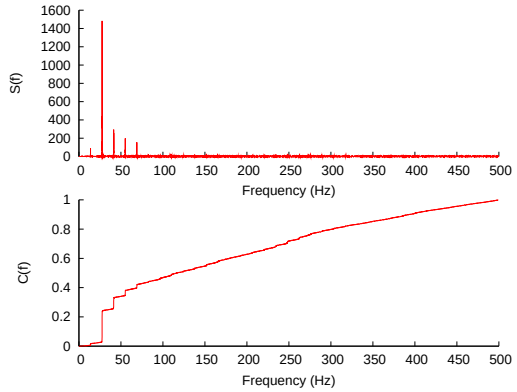
In contrast, Figure 7(b) shows a reflected attack using echo reply packets. Since the source address in reflected attacks is not spoofed, we can count 145 different reflectors located in countries such as Brazil, Japan, Korea, Singapore, and United States. The attack rate is 4300packets/s. Here we observe concentration of power in lower frequencies creating a corresponding shift in the NCS.

The intuition behind the result requires consideration of a single attack source and then the interaction of multiple attackers. We suggest that a single attacker sending at full rate will have high frequency components in the attack traffic because any computer and network interface has a maximum possible transmission rate due to hardware or operating system limits. This rate gives that attacker a basic frequency and harmonics at multiples of that frequency, resulting in some high-frequency components and a basically linear cumulative spectrum. Now consider a collaborative, distributed attack with multiple attackers, each sending as fast as possible. Each attacker will have its own maximum rate and corresponding spectra, but in the aggregate, their traffic will "blur together", because the attackers operate independently at different rates and frequencies, and because each attacker experiences noise from different levels of cross-traffic, loosing high frequency components and causing the lower frequency components to dominate the spectrum. We expand on this intuition in several steps: through experiments in Section 6.2, simple simulations in Section 6.3, and discussion about robustness in Section 7.

Since it is difficult to quantify differences between attacks with a graphical representation of spectrum, we use the $F(60\%)$ value



(a) Single-source



(b) Multi-source

Figure 7: The power spectrum (top) and NCS (bottom) for two example attacks

(from Equation 6) for each attack to isolate the concept of power being concentrated in lower frequencies. Figure 8 plots $F(60\%)$ against the attack rates in packets/s (log-scale). Single-source attacks are concentrated in the center frequency band because their linear normalized cumulative spectrum results in mid-range $F(60\%)$ values. Multi-source attacks, both direct and reflected, are concentrated in the lower frequency band, due to the accumulation of power in lower frequencies. The two classes of attacks also have a significant difference in first-order statistics: single-source attacks have a mean 268Hz and a 95% confidence interval between 240–295Hz, while multi-source attacks have a mean of 172Hz, and a 95% confidence interval between 142–210Hz. We performed the Wilcoxon rank sum test [6] to verify that the two classes have different $F(60\%)$ ranges. The test strongly rejects the null hypothesis, that single- and multi-source attacks have identical dominant frequencies, with a p-value of 7.5×10^{-5}.

We use the spectral analysis described above to classify the remaining 13 unclassified attacks. The spectrum of five attacks match spectral characteristics of single-source attacks, with a $F(60\%)$ located above 240Hz. The remaining eight attacks have spectral characteristics similar to multi-source attacks with localization of power in the lower frequencies. These attacks also exhibit an initial ramp-up lasting from 300ms to 14 seconds corroborating the presence of multiple attackers.

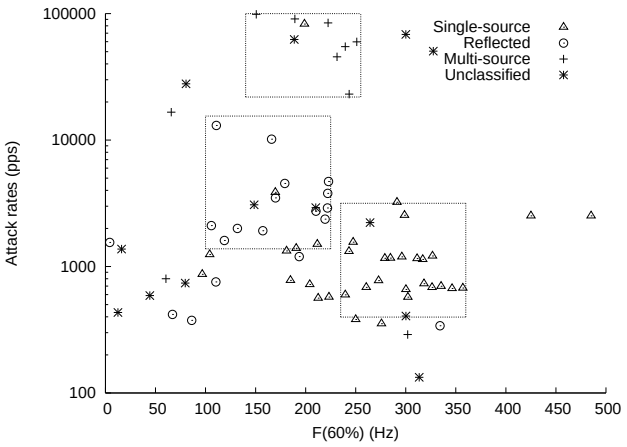


Figure 8: Comparison of $F(60\%)$ against attack rate for each attack class

Protocols	Los Nettos	USC
TCP	84.2%	95.6%
UDP	13.8%	4.10%
ICMP	1.21%	0.118%
other	0.894%	0.175%

Table 3: Percentage of packets observed for each protocol at the two sites

6. VALIDATION

We use three techniques to validate our classification algorithms and understand the nature of our observations. First, we analyze DoS attacks from a second site to confirm that the numbers and types of attacks we identified were not unique to our original observation point. Then we conduct controlled experiments and use simple numerical simulations to understand the physical characteristics behind our classification techniques.

6.1 Observations from an Alternate Site

We deployed a second trace machine at USC's connection to Internet2. The typical daytime load is 112Mbits/s with a mean of 25Kpackets/s. The traffic mix on the Internet2 link is fairly different than what we observed at Los Nettos; see Tables 3 for a breakdown of traffic at each site by protocol. Los Nettos shows much more DNS traffic (due to the presence of the b-root nameserver) and web traffic, while USC shows more "other" traffic due to gaming, file sharing and research that uses atypical or ephemeral ports.

We observed 18 attacks at USC during the months of October and November 2002. Due to the differences in monitoring duration and traffic quantity, it is difficult to compare the absolute number of attacks with our observations at Los Nettos. However, we observed about the same ratio of attacks in each attack class.

Table 4 lists attacks by class as determined by header content. Three attacks were unclassified since they completely randomize the ID value. Table 5 shows a detailed manual analysis of packet headers. Although it is difficult to directly compare with Table 2, we observe a similar set of attacks. Attacks of type TCP SYN-ACK, TCP-unusual and ICMP-illegal were not seen at USC; however, these were not very frequent at Los Nettos either.

Ramp-up and spectral analysis of attacks at USC were similar to attacks observed at our original site, and hence we do not reproduce spectra of individual attacks here. Figure 9 plots $F(60\%)$ against the attack rate (in log-scale) for each attack class. The USC results also indicate the $F(60\%)$ is located in the middle frequency band for single-source attacks, and in the low frequency band for multi-source attacks. The two classes of attacks also have first-order statistics similar to the Los Nettos. The mean for single-source attacks is 292Hz and a 95% confidence interval between 200–380Hz, while multi-source attacks have a mean of 120Hz and a 95% confidence interval between 35Hz–200Hz. One unclassified attack is most likely a single-source attack since it does not show a ramp-up and its $F(60\%)$ is 260Hz. The other two unclassified attacks are similar to each other in many aspects. They exhibit a small ramp-up of 120ms and have low $F(60\%)$ of 12Hz, indicating multiple attackers.

The tendency of multi-source attacks to localize power in lower frequencies is distinctly visible in the summary of $F(60\%)$ frequencies for both sites, Los Nettos in Figure 8 and USC in Figure 9. Based on these observations, we conclude that our results are not distorted by unusual traffic characteristics at Los Nettos and our techniques could be applied to other traffic mixes.

Attack Class	# Attacks	Range (packets/s)	Range (kbits/s)
Single-source	9	1250–54000	1100–10000
Multi-source	3	58700–95000	28000–72000
Reflected	3	2120-2250	1641–2142
Unclassified	3	6170–8500	2600–6500

Table 4: Number of attacks in each class based on header analysis at USC.

Protocol	Packet Type	Attack Class			
		S	M	R	U
TCP	SYN	-	-	-	2
	ACK	3 (1)	-	-	-
	no flag	5	-	-	-
	unusual	3	-	-	-
	state exploit	-	-	-	1
ICMP	echo request	4	-	-	-
	echo reply	-	-	3	-
UDP	all	5	2 (2)	-	-
Other	ip-proto 0	4	-	-	-
	ip-proto 255	1	1 (1)	-	-
	fragmented	1	-	-	-
	routing	1	-	-	-

Table 5: Detailed analysis of packet headers at USC.

6.2 Experimental Confirmation

To understand the effect of network topology and number of

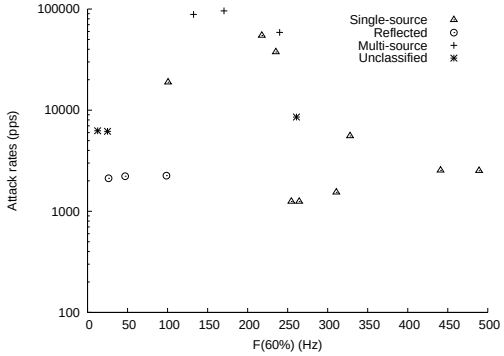
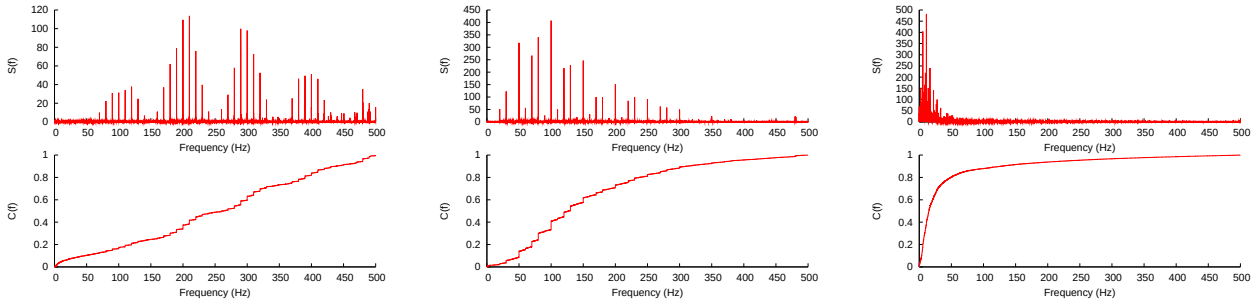
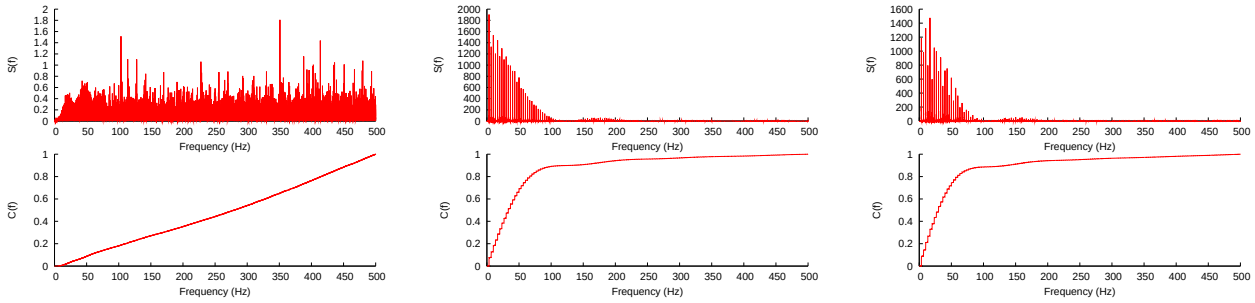


Figure 9: Comparison of $F(60\%)$ against attack rate by attack class for USC attacks.



(a) One Source (b) Two Sources (c) Three Sources

Figure 10: WAN experiments using a clustered topology.



(a) One Source (b) Two Sources (c) Three Sources

Figure 11: WAN experiments using a distributed topology.

sources on attack traffic we carried out controlled experiments over the Internet varying both these parameters. We placed synthetic attackers at universities and research labs on both coasts of the United States (at ISI East, UCLA, UCSB, UCSD, UMass, and USC). We measured traffic at a target while varying the number of sources from 1–5 considering two topologies: a *clustered* attack, where all attackers reside on the same LAN segment and are well connected to the target via a high bandwidth, low latency link, and a *distributed* topology where attackers are widely distributed with attackers on both coasts. Although it is not possible to control Internet traffic, we repeated these experiments multiple times during heavy and light network utilization, during peak weekday hours and early morning/weekends (as measured local to the target). The victim and the observation point were located on the same Ethernet segment, connected via a hub. The traffic traces were captured using tcpdump [18]. Each synthetic DoS attacker was an Iperf [39] UDP source sending 50 byte packets at a rate of 1Mbits/s and each experiment was run for 100 seconds. The hosts in the experiments have different operating speeds and all run variants of Linux.

Figure 10(a) shows the clustered topology with only one sender. We see strong peaks in the high frequency ranges. This behavior is an inherent characteristic of a host sending at a rapid pace. All computers run at certain frequencies due to clocks in the CPU, the network card, and the operating system. We therefore believe that this pattern will be present in any host that is sending as rapidly as possible.

Looking across Figure 10 we see how the spectrum changes as the number of sources increase from 1 to 3 with all sources on the same Ethernet segment. The dominant spectral characteristics tend to shift toward low frequencies as the number of sources increase, with $F(60\%)$ at 300Hz, 150Hz, and 21Hz for 1, 2 and 3 sources respectively. In Section 6.3 we examine this effect more closely to show that it is due to multiple attackers operating out of phase with each other.

To examine the effect of network topology we repeated this experiment with each source at different locations around the Internet. Figure 11(a) shows the spectrum of a single synthetic attacker located at UMass. The spectrum lacks the distinct peaks of Figure 10(a). We believe this smoothing is due to a larger amount of cross traffic and more variation in transit time than with a single attacker in the clustered topology. The normalized cumulative spectrum is robust to this effect, with both single-source attacks showing nearly linear trends.

Comparing Figure 11(a) to Figures 11(b) and 11(c), we see a shift in the spectrum to lower frequencies, with $F(60\%)$ at 43Hz and 35Hz for 2 and 3 sources, as compared to 328Hz for a single-source. Again, we believe this is due to the presence of multiple, unsynchronized sources.

Figure 12 summarizes the $F(60\%)$ results for 30 experiments conducted at different times of the day. The experiments show a localization of power in lower frequencies as the number of sources increase from 1–5 in both clustered and the distributed topologies. As seen in Figure 12, the $F(60\%)$ is close to 300Hz during single-source experiments, but reduces to 100Hz when more sources are introduced. The experiments indicate that although the absolute value of $F(60\%)$ differs from one experiment to the next, the multi-source attacks always have a lower $F(60\%)$ in both topologies, qualitatively confirming our attack observations at Los Nettos.

To confirm the above results are not due to characteristics unique to the synthetic attack traffic generated by Iperf, we conducted ex-

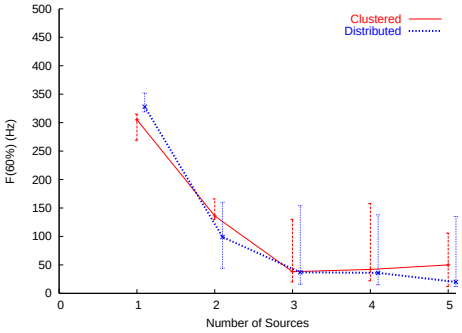


Figure 12: Localization of power as the number of sources increase in both clustered and distributed topologies. The error bars show 95% confidence intervals for the experiment.

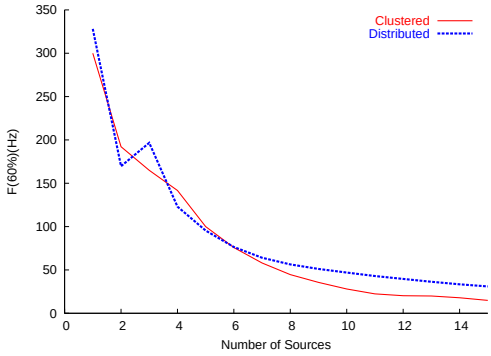


Figure 13: The effect of aggregation of multiple sources at different phases.

periments with real DoS attack tools on a dumbbell-shaped topology consisting of 12 hosts, four hubs and two Cisco routers (we could not deploy the attack tools on the clustered and distributed topologies due to signature-based IDS monitoring tools). The testbed provides a low latency (less than 1ms), high bandwidth (100Mbits/s) connection between the attackers and the victim. We generated attack traffic using three DoS tools: punk, stream, and synful, and web-based background traffic with WebStone [40]. The results are discussed in Appendix A.

These experiments confirm the presence of multiple attackers changes the attack spectrum, and that the $F(60\%)$ is a reasonably robust discriminator between single- and multi-source attacks. They do not completely explain, however, the reasons why multiple attackers shift the spectrum; we consider that next.

6.3 Understanding Multi-Source Effects

Although Section 6.2 confirms the validity of using spectral analysis to discriminate between single- and multiple-sources, it does not explain *why* spectral content is a good discriminator. To understand the physical meaning behind the shift in $F(60\%)$ to lower frequencies, we considered three hypotheses:

1. Aggregation of multiple sources at either slightly, or very different rates.
2. Bunching of traffic due to queuing behavior (analogous to ACK compression [25], but for data).
3. Aggregation of multiple sources, each at different phase.

To investigate these hypotheses we perform simple numerical simulations. Due to space constraints we omit plots support our rejected hypotheses; interested readers are referred to Appendix A. To test Hypothesis 1, we aggregate a *scaled* attack trace with the original attack trace to simulate aggregation of multiple attackers at different rates. If $a(t)$ represents the packet arrival sequence in the original trace, we multiply the time-stamp by a *scaling factor* s, with artificial added jitter denoted by ϵ, to generate a scaled trace. Therefore the aggregate trace is given by:

$$a_1(t) = a(t) + a((s+\epsilon)t) \tag{7}$$

We use the packet trace from the single-source clustered experiment (Figure 10(a)) and vary the scaling factor from 0.5 to 2 representing attackers with rates varying from twice to half the original attack rate respectively (ϵ is uniformly distributed between 1–5μs). The scaled trace is then aggregated with the original attack trace using the approach defined by Kamath et al. [20]. If Hypothesis 1 is true, then a change in the attack rate should cause a corresponding change in $F(60\%)$. However, we observe $F(60\%)$ remains nearly constant even when aggregated with an attacker with dissimilar attack rates. Hence we reject Hypothesis 1.

To test Hypothesis 2 we capture a packet arrival sequence on the attacker host and filter the arrival sequence to delay transmission until p packets (p varies from 5–15) have arrived, sending out all the packets at once. The spectra created by this process has a cluster of prominent frequencies around 320Hz with very little power (less than 15%) in the lower frequency band. The normalized cumulative spectrum has a sharp rise between 300–320Hz which is unlike spectra we have observed earlier. We therefore discarded Hypothesis 2.

To test Hypothesis 3, we aggregate a *shifted* attack trace with the original attack trace to simulate aggregation of attackers at different phases. If $a(t)$ represents the packet arrival sequence in the original trace, we add a *phase* ϕ, with jitter ϵ, to generate a shifted trace. Therefore the aggregate trace is given by:

$$a_3(t) = a(t) + a(t + \phi + \epsilon) \tag{8}$$

We vary the phase from 1–200ms, representing the difficultly of attackers to start and remain synchronized. If Hypothesis 3 is true, then changes in attacker phase should cause a corresponding change in $F(60\%)$, but we observe $F(60\%)$ remains nearly constant even when aggregated with an attacker with dissimilar phase demonstrating that phase alone (Hypothesis 3) does not cause the shift.

Finally we consider a variation on Hypothesis 3 and aggregate multiple streams each slightly out of phase. To test this hypothesis we aggregate shifted attack traces with the original attack trace to simulate aggregation of multiple attackers at different phases. If $a(t)$ represents the packet arrival sequence in the original trace, we generate the shifted trace by:

$$a_{3b}(t) = \sum_{i=2}^{n} a(t + i\phi) \tag{9}$$

We vary the number of attackers n from 2–15 with a 1ms phase shift between each attacker. If the hypothesis is true, then we should observe a drop in $F(60\%)$ as the number of attacker increase. Figure 13 plots the number of sources against $F(60\%)$ when using packet traces from both Figures 10(a) and 11(a). In both cases we observe a drop in the $F(60\%)$ as the number of sources increase indicating that phase along with aggregation of multiple sources is most likely the cause of localization of power in the lower frequencies. This result is consistent with the attack traffic observed at Los Nettos and USC.

In summary, these experiments support our claims that: (a) High-rate single-source attacks have an inherently linear cumulative spectrum. (b) High-rate multi-source attacks cannot maintain this linearity; lower frequency components dominate due to aggregation of multiple attackers starting out of phase. (c) Cross traffic can decrease the prominence of individual frequencies, but the cumulative spectrum is robust to its effects. While these results apply to high-rate attackers, it is possible for attackers to affect their spectral characteristics by changing their attack rate. We examine this issue in the next section.

7. SENSITIVITY ANALYSIS

Network security is an arms race: both attack tools and defenses evolve in relation to each other. Thus an important consideration of our framework is its robustness to improved attack tools. In fact, our ramp-up and spectral analysis techniques were motivated by limitations of header analysis in the face of packet spoofing.

Although header analysis was successful at classifying 83% of the attacks we observed, this percentage may drop as more sophisticated tools become available. Even though source addresses are forged, currently most attack tools neglect randomizing the ID field. However, it is easy for attackers to spoof this field and even standard operating systems are randomizing the ID field when the packet is not fragmented to discourage OS fingerprinting [14]. Further, it may be possible to synchronize ID values in low volume attacks using out-of-band communication to make it appear monotonically increasing and evade correct classification [13]. The use of TTL is somewhat more robust (assuming stable routing), since attack packets with very low TTL values will fail to reach the victim. Statistical analysis of TTL values may be helpful in determining attacker distance in spite of spoofing. Unfortunately usefulness of this approach will be limited because a distance of a few hops quickly encompasses much of the Internet. We expect evolution of attack tools to increase dependence on more advanced classification techniques based on spectral content.

Even though none of the observed single-source attacks exhibit an initial ramp-up, it can be easily generated by an attacker that gradually increases the attack rate emulating a multi-source attack and effectively triggering more complex response mechanisms. On the other hand, in large multi-source attacks we believe an initial ramp-up will be quite difficult to conceal. The duration of the ramp-up may vary based on the zombie clock skew and differences in the zombie-victim network distance, but masking the ramp-up by accounting for both sources of variability would require fair sophistication.

Spectral analysis is more robust to attacker manipulation than header analysis. We believe the characteristics of high-rate attack traffic are inherent; they cannot be avoided by single- or multi-source attackers sending at maximum rate. Further, it is not practical for a multi-source attacker to synchronize geographically distributed attackers to create spectral characteristics similar to single-source attacks. Accomplishing comparable levels of synchronization requires not only tight time synchronization between attacking hosts but also measurement and accounting for the varying propagation and queuing delay between each attacker and the victim. It may be possible for a single-source attacker to masquerade as a multi-source attack if it is willing to reduce its attack rate. A single-source can generate packets in bursty, on-off patterns by introducing a delay between packets and creating dominant low frequency contents in its spectrum. As future work we will investigate spectral analysis techniques based on uneven sampling that will be more robust to such attack patterns [28] and study the effect of packet loss and queuing on the attack spectra.

Finally, we consider the sensitivity of the 60% quantile used to in attack classification. The technique used in this paper is based on testing for localization of power in lower frequency band (under 200Hz). Any mid-range quantile can capture this characteristic of the trace; our selection of 60% is somewhat arbitrary. To verify this we compared classification of the attacks we captured and found comparable results for quantiles between 45–65%. However, at very low or high quantiles (less than 40% or more than 70%) many attacks are incorrectly classified because the measure becomes overly influenced by variance at either end of the spectra.

8. APPLICATIONS

There are several applications of our results, including automated attack detection, developing synthetic models of attack traffic and inferring the amount of DoS attack activity in the Internet. Although details of these applications are outside the scope of this paper, we briefly discuss each next.

8.1 Automating Attack Detection

A robust automatic attack detection tool is useful in guiding response systems (both manual and automated) in installing filters [19, 27] or for use in aggregate congestion control for flash crowds [23]. Although we use simple filters for first-pass detection of an attack, the approaches we develop help *classify* attacks as from single or multiple sources. Discrimination between single and multi-source attacks is useful in selecting the appropriate response mechanism since some mechanisms are more expensive when dealing with multiple attackers compared to single attackers (for example, traceback [32]). We developed an automated tool that takes the attack trace and carries out spectral analysis to classify it. While this demonstrates the feasibility of such a tool, but we have not yet integrated this tool with other detection systems. We are also exploring the possibility of developing attack profiles to identify similar *attack scenarios* consisting of same the attack tool and zombies used in repeated attacks.

8.2 Modeling Attacks

Many simulation studies of DoS attack detection and response use fairly simple traffic models such as constant bit-rate sources with fixed size packets. Such models fail to capture the nuances of attack traffic. Although real attack tools are easy to obtain and can be used in a testbed, there remain questions about how to support large numbers of attack machines and how to configure a testbed to reproduce attacks similar to those in the wild. Clearly, to create more realistic synthetic DoS traffic both in simulation and testbeds, we need a better understanding of attack dynamics.

To our knowledge there have been no published studies of detailed characterization or models of DoS attack traffic. Studies based on back-scatter observe attacks indirectly, and thus do not capture fine-grained details of the attack dynamics [26]. Given the many modes of failure an attack can cause (such as hardware failures, exploitation of software glitches, misconfiguration), it is important to create faithful reproductions of real attacks. Although not the focus of this paper, we include some statistics about the kinds of attacks we see in the wild. Future work may use our tools as part of a broader study to better characterize DoS attacks, laying down the groundwork for the development of more realistic attack models.

8.3 Inferring DoS Activity in the Internet

Using our detection tools, we captured 80 DoS attacks in Los Nettos over five months. If we consider these attacks to be a sample of DoS activity in the Internet as a whole, we can project attack ac-

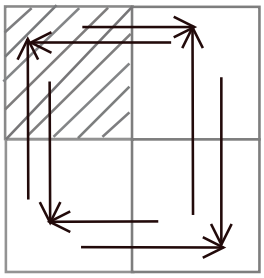


Figure 14: Limitations when extrapolating Los Nettos DoS activity to the Internet.

tivity to the public Internet. Such a projection should be considered *extremely* rough due to the small fraction of the Internet and the relatively few attacks we observed, and the assumptions required by such an estimation. However, we believe that the methodology proposed below coupled with a larger future monitoring effort can provide a reasonable Internet-wide estimate of attack counts.

To arrive at our rough projection, we first compare the size of the monitored address space to the Internet. We monitor about 0.105% of the advertised Internet address space, determined by comparing the size of the routing table advertised by Los Nettos to the size of the advertised Internet address space as reported by Route Views [24] on December 15, 2002. We assume that addresses are consumed uniformly and that attacks are targeted uniformly in both Los Nettos and the general Internet. Given these assumptions, we can scale our observations to the Internet accordingly (multiplying by a factor of 1900).

We observe DoS attacks that both transit and terminate in Los Nettos. Since distributed DoS attacks depart from many sources to attack a victim, we would expect that transiting attacks are more prevalent than terminating attacks. For example, in Figure 14 we monitor a quarter of the address space. If we measure unique victims of any attacks *involving* the shaded area, we observe 3 attacks and project 12, overestimating by a factor of three. If instead we observe only attacks *terminating* in the area, we get an accurate estimate of 4. In general, projections from transit traffic identify an upper-bound on the number of attacks, since it may overestimate by the minimum of the scale-up factor or the number of distributed attackers. Although, it is difficult to quantify the false negative rate of our detection mechanism, we believe it captures most large attacks but misses many small attacks. Therefore, the number of attacks observed can be considered a lower bound.

Based on these assumptions, Table 6 projects our observations to the Internet as a whole. Clearly these projections are tentative, since we know that there were at least 10 attacks in October but we project none. One point of comparison is the work of Moore et al. where they observe backscatter from 12,805 attacks in 3 weeks [26]. Direct comparison between their observation and ours is extremely difficult since the methodology and classes of counted attacks are very different, but it is somewhat reassuring that both their observation and our estimate are roughly the same order of magnitude.

Although the observations are very rough and require many assumptions, we believe this methodology will be useful at approximating attack prevalence if we can increase the size and duration of the monitored region. We are working on doing both.

Month	**In Los Nettos**		**In the Internet**
	transiting	terminating	terminating (projected)
July	18	6	11400
Aug	12	1	1800
Sept	10	5	9500
Oct	10	0	0*
Nov	9	6	11400

Table 6: Extrapolating Los Nettos DoS activity to the Internet.

9. CONCLUSION

This paper presented a framework to classify DoS attacks into single- and multi-source attacks. In addition to using packet headers to classify the attacks, we develop two new approaches: initial ramp-up transients and spectral analysis. These approaches depend only on information in the attack packet stream, and we believe the spectral characteristics of attacks cannot be altered without reducing attack rates.

We evaluated our framework on 80 attacks captured from two peering links at a moderate-size, regional ISP. We validated our framework with attacks captured at a second monitoring site, and through experiments with synthetic attacks on a wide-area network, and real attack tools on an isolated testbed. We used experiments and simulations to explain the underlying reasons for the difference in attack characteristics.

DoS attacks are constantly evolving, and currently there is a dearth of detailed information regarding attack dynamics. We suggested several applications of the techniques developed in this paper: to develop an automatic detection and response system based on number of attackers, to enable fine-grained analysis of attack patterns and topologies, and to infer global DoS activity.

Acknowledgments

We would like to thank Jim Pepin, Walter Prue and Sanford George of Los Nettos, and Brian Yamaguchi of USC, for helping us obtain traces, and for discussions about handling DoS attacks. We would like to thank Kimberley Claffy, David Moore, Elizabeth Belding-Royer, Bing Wang, Don Towsley, Deborah Estrin, and Colin Perkins for providing access to their lab machines for our WAN experiments. Rohit Agarwal helped with the testbed experiments and identifying attack tools. In addition we would like to thank Craig Partridge, Edmond Jonckheere, and the anonymous reviewers for their useful comments on the earlier version of this paper.

10. REFERENCES

[1] M. Allman, V. Paxson, and W. Stevens. TCP congestion control. RFC 2581, Internet Request For Comments, April 1999.

[2] Incident Detection Analysis and Response. http://ki.sei.cmu.edu/idar.

[3] Paul Barford, Jeffery Kline, David Plonka, and Ron Amos. A signal analysis of network traffic anomalies. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop*, Marseilles, France, November 2002.

[4] Steven Bellovin. ICMP traceback messages. Work in Progress: draft-bellovin-itrace-00.txt.

[5] Steven Bellovin. A technique for counting nated hosts. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop*, pages 112–208, Marseille, France, November 2002.

[6] George Box, Gwilym Jenkins, and Gregory Reinsel. *Time series analysis: forecasting and control*. Prentice-Hall, Upper Saddle River, New Jersey, 1994.

[7] Ronald Bracewell. *The Fourier Transform and Its Applications*. Series in Electrical Engineering. McGraw-Hill,

New York, NY, 1986.

[8] Andre Broido, Evi Nemeth, and kc Claffy. Spectroscopy of DNS Update Traffic. In *Proceedings of the ACM SIGMETRICS*, San Diego, CA, June 2003.

[9] Hal Burch and Bill Cheswick. Tracing anonymous packets to their approximate source. In *Proceedings of the USENIX LISA*, pages 319–327, New Orleans, USA, Decemeber 2000. USENIX.

[10] Chen-Mou Cheng, H.T. Kung, and Koan-Sin Tan. Use of spectral analysis in defense against DoS attacks. In *Proceedings of the IEEE GLOBECOM*, Taipei, Taiwan, 2002.

[11] Drew Dean, Matt Franklin, and Adam Stubblefield. An algebraic approach to IP traceback. In *In Proceedings of Network and Distributed Systems Security Symposium*, San Diego, CA, February 2001.

[12] David Dittirch. DDoS Attacks and Tools. http://staff.washington.edu/dittrich/misc/ddos.

[13] Julio Escobar, Craig Partridge, and Debra Deutsch. Flow Synchronization Protocol. *ACM/IEEE Transactions on Networking*, 2(2):111–121, April 1994.

[14] Fyodor. Remote OS detection via TCP/IP stack fingerprinting. http://www.insecure.org/nmap/, October 1998.

[15] Thomer M. Gil and Massimiliano Poletto. MULTOPS: A Data-Structure for bandwidth attack detection. In *Proceedings of the USENIX Security Symposium*, pages 23–38, Washington, DC, July 2001.

[16] Hevin Houle and George Weaver. Trends in denial of service technology. CERT Coordination Center at Carnegie-Mellon University, October 2001.

[17] Alefiya Hussain, John Heidemann, and Christos Papadopoulos. A Framework for Classifying Denial of Service Attacks. In *Proceedings of ACM SIGCOMM 2003*, Karlsrhue, Germany, August 2003.

[18] Van Jacobson, Craig Leres, and Steven McCanne. tcpdump - the protocol packet cpature and dumper program. http://www.tcpdmp.org.

[19] Peter Reiher Jelena Mirkovic, Greg Prier. Attacking DDoS at the source. In *Proceedings of the IEEE International Conference on Network Protocols*, Paris, France, November 2002.

[20] Purushotham Kamath, Kun chan Lan, John Heidemann, Joe Bannister, and Joe Touch. Generation of high bandwidth network traffic traces. In *In Proceedings of MASCOTS*, pages 401–410, Fort Worth, Texas, USA, October 2002. IEEE.

[21] Angelos D. Keromytis, Vishal. Misra, and Dan. Rubenstein. SOS: Secure Overlay Services. In *Proceedings of ACM SIGCOMM 2002*, August 2002.

[22] Los nettos-passing packets since 1988. http://www.ln.net.

[23] Ratul Mahajan, Steven M. Bellovin, Sally Floyd, John Ioannidis, Vern Paxson, and Scott Shenker. Controlling high bandwidth aggregates in the network. In *ACM Computer Communication Review*, July 2001.

[24] D. Meyer. University of oregon Route Views Project. Advanced Network Technology Center web site, http://www.antc.uoregon.edu/route-views.

[25] Jeffrey C. Mogul. Observing TCP dynamics in real networks. Technical Report 92.2, DEC Western Research Laboratory, April 1992.

[26] David Moore, Geoffrey Voelker, and Stefan Savage. Inferring Internet denial of service activity. In *Proceedings of the USENIX Security Symposium*, Washington, DC, USA, August 2001. USENIX.

[27] Christos Papadopoulos, Robert Lindell, John Mehringer, Alefiya Hussain, and Ramesh Govindan. COSSACK: Coordinated Suppression of Simultaneous Attacks. In *In Proceeding of Discex III*, Washington, DC, USC, April 2003.

[28] Craig Partridge, David Cousins, Alden Jackson, Rajesh Krishnan, Tushar Saxena, and W. Timothy Strayer. Using signal proceesing to analyze wireless data traffic. In *Proceedings of ACM workshop on Wireless Security*, pages 67–76, Atlanta, GA, September 2002.

[29] Vern Paxson. Bro: A system for detecting network intruders in real-time. *Computer Networks*, 31(23–24):2435–2463, Decemeber 1999.

[30] Vern Paxson. An analysis of using reflectors for Distributed Denial-of-Service Attacks. *ACM Computer Communications Review (CCR)*, 31(3), July 2001.

[31] Martin Roesch. Snort - lightweight intrusion detection for networks. http://www.snort.org.

[32] Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson. Practical network support for IP traceback. In *Proceedings of the ACM SIGCOMM Conference*, pages 295–306, Stockholm, Sweeden, August 2000. ACM.

[33] Alex C. Snoeren, Craig Partridge, Luis A. Sanchez, Christine E. Jones, Fabrice Tchakountio Stephen T. Kent, and W. Timothy Strayer. Hash-based ip traceback. In *Proceedings of the ACM SIGCOMM*, pages 3–14, San Deigo CA, August 2001. ACM.

[34] Dawn X. Song and Adrian Perrig. Advanced and authenticated marking schemes for IP traceback. In *Proceedings of the IEEE Infocom*, Anchorage, Alaska, April 2001.

[35] Neil Spring, Ratul Mahajan, and David Wetherall. Measuring ISP topologies with rocketfuel. In *Proceedings of ACM/SIGCOMM '02*, August 2002.

[36] Robert Stone. Centertrack: An IP overlay network for tracking DoS floods. In *Proceedings of the USENIX Security Symposium*, pages 199–212, Denver, CO, USA, July 2000. USENIX.

[37] Cisco Systems. Netflow services and applications. http://www.cisco.com/warp/public/732/netflow.

[38] Cisco Systems. Rmon. http://www.cisco.com/warp/public/614/4.html.

[39] Ajay Tirumala, Feng Qin, Jon Dugan, Jim Ferguson, and Kevin Gibbs. Iperf Version 1.6.5. http://dast.nlanr.net/Projects/Iperf/.

[40] Gene Trent and Mark Sake. WebSTONE: The first generation in HTTP server benchmarking.

[41] Haining Wang, Danlu Zhang, and Kang Shin. Detecting SYN flooding attacks. In *Proceedings of the IEEE Infocom*, New York, NY, June 2002. IEEE.

[42] Zhi-Li Zhang, Vinay Ribeiro, Sue Moon, and Christophe Diot. Small-time scaling behaviors of Internet backbone traffic: An empirical study. In *Proceedings of the IEEE Infocom*, San Francisco, CA, April 2003.

[43] E. Zwicky, S. Cooper, D. Chapman, and D.Ru. *Building Internet Firewalls*. 2nd Edition. O'Reilly and Associates, 2000.

Location	CPU (Mhz)	Hop Count	RTT (ms)
UCSB	1800	9	5
UCSD	500	10	7
UCLA	900	11	2
ISIE	900	15	74
UMass	600	16	90
USC1	1800	6	1
USC2	1800	6	1
USC3	1000	6	1
USC4	900	6	1
USC5	500	6	1

Table 7: The WAN testbed. The top block of hosts are used for distributed experiments while the bottom block of hosts are used for clustered experiments.

A. APPENDIX: ADDITIONAL ANALYSIS AND VALIDATION

In this section we present the results that were omitted due to space constraints in the SIGCOMM paper [17].

A.1 Analysis of Attack Duration and Rate

This section investigates the relation between attack rate and attacker population. We captured attacks with peak rates ranging from 300packets/s to 98Kpackets/s. In addition on the analysis provided in Section 5.3, Figure 15 shows the cumulative distribution of the attack duration and peak attack rates in packets and kbits/s. In some attacks packet rates vary over the attack duration. An increase in the attack rate is usually due to addition of new machines or the addition of new type of attack. We also observed a reduction in attack rate, possibly due to withdrawal of sources or installation of filters by network operators.

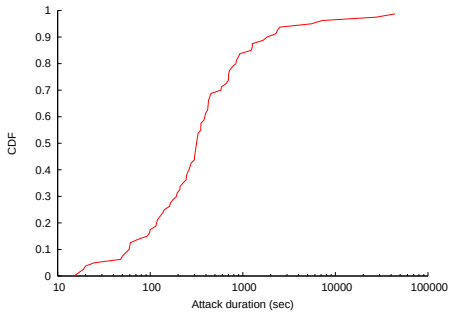
A.2 Validation

In this section we present details regarding the experimental methodology and discuss simulation results in more detail.

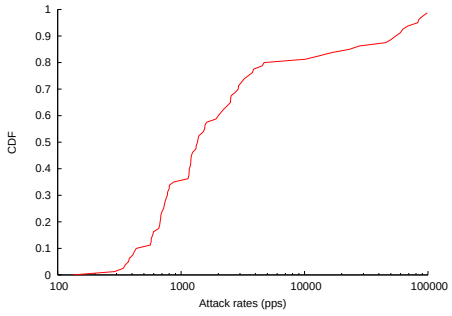
A.2.1 Experimental Confirmation

To understand the effect of network topology and number of sources on attack traffic we carried out controlled experiments over the Internet varying both these parameters. We placed synthetic attackers at universities and research labs on both coasts of the United States (at ISI East, UCLA, UCSB, UCSD, UMass, and USC). Table 7 provides a complete list of all the hosts, their operating speeds, the number of hops and RTT from the victim.

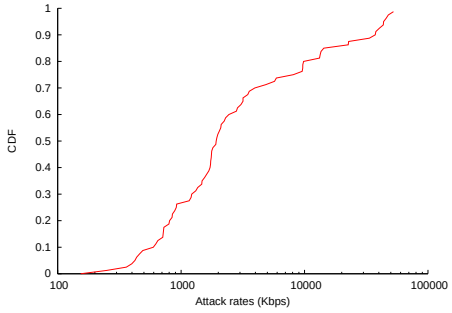
To confirm the observed spectral behavior is not due to characteristics unique to the synthetic attack traffic generated by Iperf, we conducted experiments with real DoS attack tools on a dumbbell-shaped topology consisting of 12 hosts, four hubs and two Cisco routers The testbed provides a low latency (less than 1ms), high bandwidth (100Mbits/s) connection between the attackers and the victim. We generated attack traffic using three DoS tools: punk, stream, and synful, and web-based background traffic with WebStone [40]. Figure 16 shows that all three attack tools produced spectral characteristics similar to Figure 12, both in the single- and multi-source experiments. We observe the single-source spectra (Figure 16(a), (b), and (c)) created by attack tools show strong characteristic high frequencies and linear normalized cumulative spectra. The power in the higher frequencies begins to reduce and the NCS shows a localization of power in lower frequencies as more attackers are added (Figure 16(c), (d) and (e)). All three attack tools produced spectral characteristics similar to Figure 12, in both single- and multi-source experiments.



(a) cdf of attack duration



(b) cdf of attack rates in packets/s



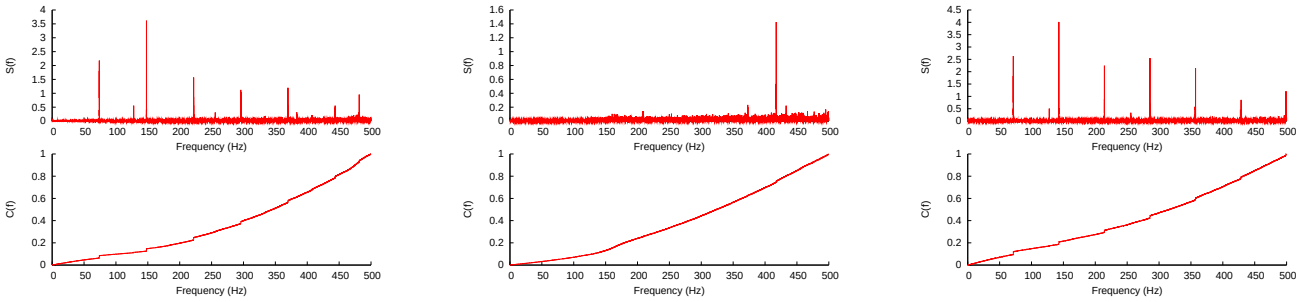
(c) cdf of attack rates in kbits/s

Figure 15: The cumulative distribution of (a) attack duration, and attack rates both in (b) packets/s and (c) kbits/s for 80 attacks

A.2.2 Understanding Multi-source Effects

To understand the physical meaning behind the shift in $F(60\%)$ to lower frequencies, we considered three hypotheses: (a) Aggregation of multiple sources at either slightly, or very different rates. (b) Bunching of traffic due to queuing behavior (analogous to ACK compression [25], but for data). (c) Aggregation of multiple sources, each at different phase. This section presents the graphs for the results discussed in Section 6.3

To test Hypothesis 1, we aggregate a *scaled* attack trace with the original attack trace to simulate aggregation of multiple attackers at different rates. Figure 17(a) plots the scaling factor s against $F(60\%)$. We observe $F(60\%)$ remains nearly constant even when aggregated with an attacker with dissimilar attack rates. However,



(a) Punk: One Source

(b) Stream: One Source

(c) Synful: One Source

(d) Punk: Three Sources

(e) Stream: Three Sources

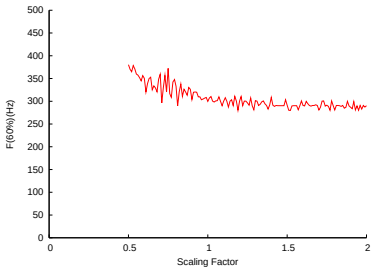
(f) Synful: Three Source

Figure 16: Testbed experiments using real attack tools.

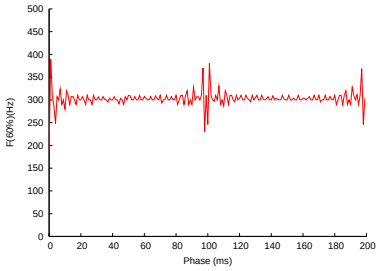
we observe $F(60\%)$ remains nearly constant even when aggregated with an attacker with dissimilar attack rates. Hence we reject Hypothesis 1.

To test Hypothesis 2 we capture a packet arrival sequence on the attacker host and filter the arrival sequence to delay transmission until p packets (p varies from 5–15) have arrived, sending out all the packets at once. The spectra created by this process has a cluster of prominent frequencies around 320Hz which is unlike spectra we have observed earlier. We therefore discarded Hypothesis 2.

To test Hypothesis 3, we aggregate a *shifted* attack trace with the original attack trace to simulate aggregation of attackers at different phases. Figure 17(b) plots the phase ϕ against $F(60\%)$. We observe $F(60\%)$ remains nearly constant even when aggregated with an attacker with dissimilar phase demonstrating that phase alone (Hypothesis 3) does not cause the shift. in $F(60\%)$, but we observe $F(60\%)$ remains nearly constant even when aggregated with an attacker with dissimilar phase demonstrating that phase alone (Hypothesis 3) does not cause the shift.



(a) Aggregation of two sources at different rates



(b) Aggregation of two sources at different phases

Figure 17: Testing the Hypotheses with numerical simulation