Telecommunication Systems 28:3,4, 435–451, 2005

DDoS Attack Detection and Wavelets

LAN LI and GYUNGHO LEE {lli;ghlee}@ece.uic.edu
Department of Electrical and Computer Engineering, University of Illinois at Chicago, 1020 SEO, S. Morgan st., Chicago, IL 60660, USA

Abstract. This paper presents a systematic method for DDoS attack detection. DDoS attack can be considered a system anomaly or misuse from which abnormal behavior is imposed on network traffic. Attack detection can be performed via abnormal behavior identification. Network traffic characterization with behavior modeling could be a good indication of attack detection. Aggregated traffic has been found to be strong bursty across a wide range of time scales. Wavelet analysis is able to capture complex temporal correlation across multiple time scales with very low computational complexity. We utilize energy distribution based on wavelet analysis to detect DDoS attack traffic. Energy distribution over time will have limited variation if the traffic keeps its behavior over time (i.e. attack-free situation) while an introduction of attack traffic in the network will elicit significant energy distribution deviation in a short time period. Our experimental results with typical Internet traffic trace show that energy distribution variance markedly changes, causing a "spike" when traffic behaviors are affected by DDoS attack. In contrast, normal traffic exhibits a remarkably stationary energy distribution. In addition, this spike in energy distribution variance can be captured in the early stages of an attack, far ahead of congestion build-up, making it an effective detection of the attack.

Keywords: distributed denial of service, energy distribution, traffic characterization, wavelet analysis, attack detection

1. Introduction

Distributed denial of service (DDoS) attack has been one of the major attention grabbing security attacks as it explicitly threatens the stability of the Internet. *Computer Economics* [5] estimated that the total economic impact of Code Red was $2.6 billion, and Sircam cost another $1.3 billion. A recent attack via SQL Slammer caused an estimated $1 billion in damage during the first five days as it rapidly spread around the globe [LaMonica, 12]. Unlike denial of service attacks that relies on a specific network protocol or a system weakness, the DDoS attack simply exploits the huge resource asymmetry between the Internet and the victim. A sufficient number of zombies generate huge amounts of "useless" traffic volume towards the victim. Through this "many to one" attack dimension, the DDoS attack is able to block access to the "thoroughfare" reaching the victim, effectively taking the victim off the Internet so that any victim-level of defense becomes irrelevant. In addition, the DDoS attack's strategies of hierarchical attack and IP spoofing make attackers difficult to trace. Although great efforts have been involved in attack detection and prevention, there is still a lack of effective and efficient

solutions to intercept ongoing attacks in a timely fashion, i.e. short enough to prevent traffic build up from DDOS attack.

Several methods have been proposed for attack detection and prevention, such as pattern-based filtering and queue management associated with flow state (e.g., LRU-RED) [Sarvotham et al., 22]. However, common characteristics of DDoS packets cannot be used as general signatures of detection and filtering. Attackers can shape the volume of attack streams and vary all packet fields to avoid exposing their own identity. In addition, even if the detector (or filter) is able to identify the pattern of the attacks, massive amount of traffic may paralyze it and make it ineffective. This is the reason why most current techniques are still unable to withstand large-scale attacks.

DDoS attack can be considered a system anomaly or misuse by which abnormal behavior is imposed on network traffic. Attack detection becomes a change identification of traffic behavior. Traditional anomaly and misuse detections, however, are confined in detecting the deviation from preset reference (e.g., normal traffic pattern) or identifying traffic with a known attack signature. The pattern and signature in use are still on packet or flow level instead of traffic behavior level in which we believe traffic nature is presented. Network traffic characterization could be a good guidance of attack detection, as long as the traffic behavior can be explicitly captured. Recent researches have shown that the time series of aggregated traffic is scale invariant or bursty across a wide range of time scales [Crovella and Bestavros, 8; Paxson and Floyd, 18; Leland et al., 14]. Since time scales can be naturally represented by wavelets [Ma and Ji, 15] and wavelet representation also matches the properties of the bursty network traffic, wavelet-based scaling analysis has been applied to characterize Internet traffic [Tian et al., 26; Ma and Ji, 15]. Analytic study in [Tian et al., 26] shows that variances of wavelet coefficients are determined by the nature of traffic itself. All of these propel us to develop the energy distribution analysis based on wavelets for traffic behavior characterization to detect DDoS attack. Following the wavelet method in [Abry and Veitch, 1; Riedi et al., 19], energy distribution in traffic is defined through the variances of wavelet coefficients on the time series of network traffic.

We applied our traffic behavior characterization with energy distribution to DDoS detection. Our experimental results with Internet traffic trace show that energy distribution variance changes markedly as traffic behavior changes due to DDoS Attack, while normal traffic exhibits a remarkably stationary energy distribution. Furthermore, such change can be captured in a timely manner, i.e. short enough to prevent traffic build up from DDOS attack.

The rest of the paper is organized as follows. We first briefly introduce related work in section 2. We then propose our wavelet based energy distribution analysis in section 3. Normal traffic trace without evident behavior anomaly (including real and simulation trace) has been investigated for its energy distribution. In section 4, through simulation, attack traffic (a typical cause of traffic behavior change) is studied with our energy distribution analysis. Finally, we conclude the paper in section 5.

2. Related work

Several detection methods have been proposed against DDoS attack. Obviously, detecting a DDoS attack is relatively easy at a victim network since attack traffic near the victim is unusually overwhelming. An attack can be captured based on identifying unusually high traffic with certain classification (e.g., packet type). However, the responsiveness of this approach is fairly poor due to the downstream location. Moreover, if an upstream link has been jammed by attack packets, there is not much that can be done on the victim side.

In contrast, attack packets with a spoofed source address can be effectively detected at the attack source side, which is the basis of Network Ingress Filtering (NIF) [Ferguson and Senie, 10]. Routers with NIF drop packets with illegitimate source IP addresses. However, this approach cannot capture attack packets generated by reflectors (here, source addresses are valid) [Chang, 7]. In addition, the effectiveness of this approach significantly depends on the coverage of NIF. Ensuring all ISP networks to install NIF is evidently not practical. Instead of using a source network, route-based packet filtering (RPF) proposed by Park and Lee implements enhanced NIF in an intermediate network [Park and Lee, 17]. RPF validates the route taken by the packets based on the inscribed source and destination addresses and the BGP routing information. If the route includes an illegitimate path, the packet is considered an attack packet. RPF has more practicability and a lower coverage requirement than NIF. However, there are several problems that prevent wide deployment of these approaches, such as BGP modification, router overhead, and the lack of inter-domain cooperation. Moreover, similar to the NIF, the RPF approach cannot filter attack packets with valid source addresses (e.g., reflected packets).

Most of the methods introduced thus far are based on the appearance of DDoS attack, such as spoofed source IP address, bandwidth distribution, attack packet pattern, etc. However, the attacker can hide attack appearance via packet reshaping. DDoS attack can be considered a system anomaly or misuse from which abnormal behavior is imposed on network traffic. Some statistical approaches have been proposed for anomaly detection based on behavior profiling, such as neural networks [Fox et al., 11] and Markov models [Ye, 27]. Behavior profiles for subjects are initially generated. As the system continues running, the anomaly detection can be performed via the variance of the present profile from the original one. In network environments, traffic characterization mechanisms possessing the ability of behavior modeling can also be applied to attack detection against an inscribed anomaly. Barford et al. also develop a network anomaly detection mechanism based on time series and wavelet analysis [Barford et al., 4]. However, a large sampling window size (3 hours) and after-the-fact detection make it less effective, since hackers are capable of attacking most vulnerable targets in well under an hour, possibly less than 15 minutes [Staniford et al., 23]. In this paper, we propose the energy distribution analysis, a characterization mechanism of traffic behavior, to implement attack detection. This mechanism can detect traffic behavior change based on its inherent characteristics. In the following sections, we will present our pro-

posed method in terms of the techniques employed and verification of its effectiveness via simulation.

3. Energy distribution analysis based on wavelets

Measurements and analytical studies have shown that network traffic exhibits self-similarity or long-range dependence. With inherent scaling property, wavelets are well-suited for analyzing self-similar process [Riedi et al., 19; Ma and Ji, 15]. Our proposed energy distribution analysis justifiably develops on the top of the wavelet technique proposed by Abry and Veitch [1]. It is also based on a conjecture that the Internet traffic is long-range dependent or self-similar. Although more complex, perhaps multifractal-like scaling behaviors under sub-second scales have been reported in recent studies [Riedi et al., 19; Zhang et al., 28], we still consider self-similar scaling over large time scales (more than 100 ms) by which we believe traffic behavior change can be presented. We measured the self-similarity of ITA packet trace [24] and found that ITA traces have high Hurst parameter[1] values (>0.7) under large time scales for different detecting points and observation time windows. This is consistent with results found by Paxson and Floyd [18] using the same trace.

3.1. Wavelet analysis and energy distribution

3.1.1. Wavelet analysis

As we explained in section 2, wavelet analysis is an effective tool to provide detailed statistical description of traffic (or traffic dynamics). Our proposed energy distribution analysis justifiably develops on the top of the wavelet technique proposed by Abry and Veitch [1]. A brief introduction to wavelets is in order. For further information about wavelets, please refer to [Daubechies, 9; Abry and Veitch, 1].

Wavelets are a set of functions that decompose data into different frequency components and then study each component according to its resolution (*scale*). With a satisfaction of certain mathematical requirements, wavelets have good localization properties in both time and frequency space. Compared to traditional Fourier series, wavelets have advantages in representing date with discontinuities and sharp spikes. They can be used to analyze nonstationary time series and give a distribution of power in two dimensions (time and frequency) instead of one (frequency in traditional spectral analysis). Also in wavelet analysis, the *scale* we use to study data plays a special role. If we apply a large scale, we would notice gross features of data. Similarly, we obtain detailed features when a small *scale* is applied. So, both the "forest" and the "trees" are under the observation of wavelet analysis.

Wavelet analysis defines a collection of nested subspace V_j corresponding to a collection of scalable and shiftable functions $\Phi_{j,i}(t)$. Time series $x(t)$ is projected into

[1] Hurst parameter (H) presents degree of long-range dependence.

each of the subspaces V_j:

$$\hat{x}_j(t) = (proj_{v_j}x)(t) = \sum_i a_x^j(i)\Phi_{j,i}(t), \quad a_x^j(t) = x\big(2^j t\big). \tag{1}$$

$(proj_{v_j}x)(t)$ is coarser than $(proj_{v_{j-1}}x)(t)$, so the key of wavelet analysis is to examine the loss of information (information difference). We define detail signals [Abry et al., 2; Abry and Veitch, 1]:

$$Details_j(t) = \hat{x}_{j-1}(t) - \hat{x}_j(t) = (proj_{v_{j-1}}x)(t) - (proj_{v_j}x)(t). \tag{2}$$

$Details_j(t)$ can also be obtained from projecting $x(t)$ onto a collection of subspaces W_j (called wavelet subspace). $\Psi_{j,i}(t)$ are wavelet functions used by projecting operation in wavelet space:

$$Details_j(t) = (proj_{W_j}x)(t) = \sum_i d_x(j,i)\Psi_{j,i}(t), \tag{3}$$

where $d_x(j,i)$ is wavelet coefficient. $d_x(j,i)$ can be considered independent and identical distribution variable with zero mean [Abry and Veitch, 1; Ma and Ji, 15]. $|d_x(j,i)|^2$, as variance of $d_x(j,i)$, measures the amount of energy distributed at time instant $2^j i(2^{-j}\nu_0$ in frequency domain) [Abry and Veitch, 1]. Using the average of $|d_x(j,i)|^2$, one can estimate the spectrum of x:

$$\widehat{\Gamma}_x\big(2^{-j}\nu_0\big) = \frac{1}{n_j}\sum_i \big|d_x(j,i)\big|^2, \tag{4}$$

where n_j is the available number of wavelet coefficients at j. $\widehat{\Gamma}_x(2^{-j}\nu_0)$ is then measuring the energy that lies in subband with central frequency of $2^{-j}\nu_0$. We use E_j to represent energy in subband with central frequency of $2^{-j}\nu_0$.

In our study, we utilize a time series $\{x(t)\}$, in which $x(t)$ is defined as the byte counts in a fixed time interval. We set a time interval of 10 milliseconds in our study as in Abry–Veitch wavelet analysis [Roghan et al., 21]. Our study also shows that a time series with a time interval of 10 milliseconds is able to represent busty invariant property of sampled trace, while adequate data can be collected in the available time window(s). In order to have an online implementation, a window-based sequential test is preferred. Therefore, the other two parameters, sliding window W and time step increment T, are also utilized in our study. For every time of T, traffic is sampled with size of W.

3.1.2. Energy distribution

Wavelet analysis actually uncouples the scaled traffic. $|d_x(j,i)|^2$ tells us how much difference (dissimilarity) exists between the two neighboring scaled traffic patterns. On the other hand, having all $|d_x(j,i)|^2$ we can reconstruct the signal series $x(t)$. We notice that the wavelet spectrum provides complete information of the correlation structure of given processes without any loss. In other words, the energy distribution (spectrum) has great potential to characterize traffic behavior.

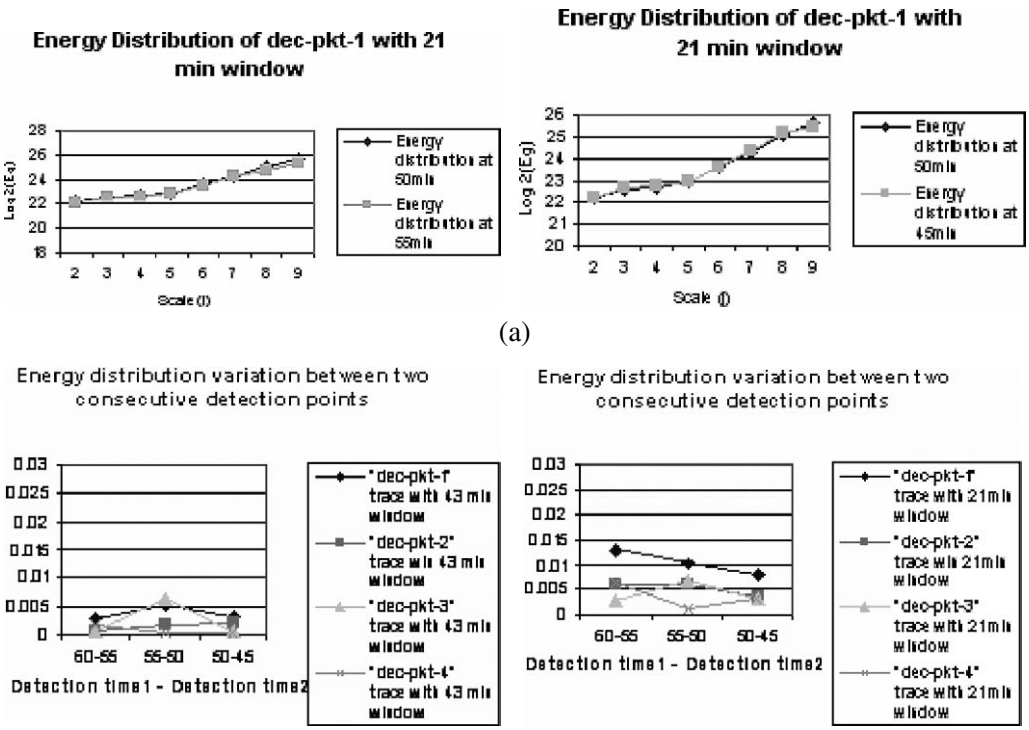


Figure 1. Energy distribution in traffic trace (ITA real trace): (a) energy distribution in ITA trace dec-pkt-1; (b) energy distribution variations in ITA traces.

If we observe the traffic at two consecutive points (we use a sliding sampling window of size W with an incremental time step of T), we have energy distribution E_j^2 at the second point and E_j^1 at the first point. The variation between E_j^2 and E_j^1 may show the characteristic/behavior change in observed traffic. Because of significant autocorrelation in a large time scale (i.e. long-range dependence), the variation of E_j is very limited if the traffic has no characteristics/behavior change. We measured energy distribution and its variation in Internet traffic (using ITA trace [24]) and found that is the case (see figures 1(a), (b), showing an example from our experimental results). Under the given sampling window (21 minutes), energy distribution variation[2] of every trace is quite small ($<$0.15). Since ITA traces were captured at separate time slots during the day, it implies that daily traffic change has little impact on energy distribution property. Throughout this verification procedure, the energy distribution has stayed relatively constant.

Although real trace is preferred in a network traffic study, it has some limitations, such as short length and fixed network context. Simulated trace is then considered an alternative in our study. We can check if our traffic characterization with

[2] The definition of energy distribution variation is described in section 3.2.

energy distribution works in simulated trace. Through the NS simulator [25], we set up a dumbbell-topology (similar topology has been used in [Tian et al., 26; Ma and Ji, 15]) and typical Web workload (similar to SURGE developed at Boston University [Barford and Crovella, 3]). The dumbbell topology (see figure 2(a)) consists of 40 Web server pools, 420 clients and 7 intermediate nodes. One bottleneck exists in the link between the servers and clients. During the simulation, we sweep the number of web sessions from 500 to 3000 and obtain packets trace at bottleneck link. With the same sliding sampling window and incremental time step applied to the ITA trace, we have energy distribution variations of simulation trace (shows in figure 2(b)). Since we extend the simulation time to 180 minutes and only extract the middle section of the trace for analysis, the trace can be considered stable. The object trace presents a good similarity (Hurst parameters > 0.8) and is stationary (we can also say traffic keeps its characteristics/behavior), so the modification of E_j shows very little variation (<0.01). This result matches our findings of the real Internet trace, ITA trace, shown in figures 1(a) and (b).

3.2. Energy distribution analysis

Since energy distribution of Internet traffic changes insufficiently, we speculate that any anomaly in traffic, like attack traffic, will cause a sudden change in energy distribution during a short time span. Based on this, we develop a threshold-based traffic signature as follows:

Suppose the two time series $x(t)$ and $x(t+\tau)$ are monitored successively, and let us define

$$Eg_j^t = \frac{1}{n_j}\sum_k \left|d_x^t(j,k)\right|^2 \quad \text{and} \quad Eg_j^{t+\tau} = \frac{1}{n_j}\sum_k \left|d_x^{t+\tau}(j,k)\right|^2$$

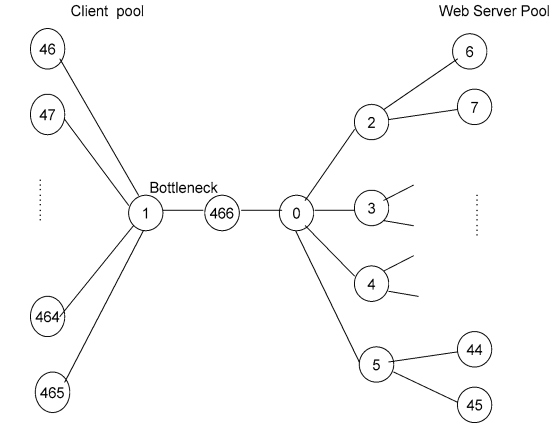
as the energy function of $x(t)$ and $x(t+\tau)$, respectively. Then, the difference of energy distribution in the two time series is

$$\Delta Eg_j = \log Eg_j^t - \log Eg_j^{t+\tau} = \log \frac{Eg_j^t}{Eg_j^{t+\tau}}. \tag{5}$$

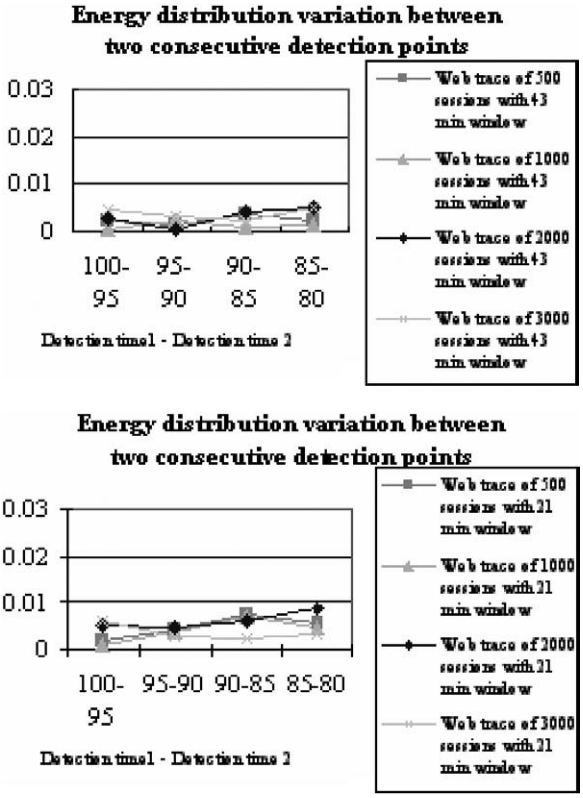
We consider the variance of ΔEg_j, i.e. energy distribution variation in the two time series to be the traffic signature. Thus, we define the normal traffic as time series

$$x(t) \in \left\{x(\tau) \mid \operatorname{var}(\Delta Eg_j) < \delta, \tau > T\right\}, \tag{6}$$

where δ is a threshold and T is a time step increment for the sliding sampling window. For a given value of δ, the traffic behavior is deemed to be normal as long as the traffic signature $\operatorname{var}(\Delta Eg_j)$ is not larger than δ. Since δ and T should reflect the traffic behavior, they may be adaptively adjusted. Also, the sampling window size W may be sensitive to traffic behavior. In our later simulation, DDoS attack traffic is employed as a cause of traffic behavior change, resulting in noticeable changes in the energy distribution variation. However, note that other anomalies, i.e. deviation from "normal" traffic, can be captured in the energy distribution variation.



(a)



(b)

Figure 2. Energy distribution in traffic trace (simulated trace): (a) dumbbell-topology; (b) energy distribution variations in NS-2 simulation trace.

3.3. Method limitations and discussion

3.3.1. Trace size

According to the trace investigation in section 3.1, our method requires a specific size of the sliding sampling window. Since wavelet analysis demands that the input data must be a power of 2, the window size needs to follow this rule. After we tried a series of window sizes, 43 min, and 21 min[3] were selected for our experiments. A smaller window size may not provide enough samples to build up traffic self-similarity while too large a window size may cause unnecessary computation during the analysis and weaken the energy distribution variation. Our method may apply to traffic on-the-fly for real-time network control. Compared with other studies [Tian et al., 26; Ma and Ji, 15], however, a 43 minutes trace seems to be quite long for real time analysis, especially for a high bandwidth link (longer data length). Fortunately, an on-line version of the Abry–Veitch wavelet analysis has been proposed [Roghan et al., 21]. With the filter-banks, it can effectively process sampling data without redundant computation. It also has a low memory requirement and scales naturally to arbitrarily high data rates for real time analysis.

3.3.2. Boundary effect

Boundary effect can exist in wavelet analysis. Given input data, the problem presented in [Abry and Veitch, 1; Abry et al., 2; Roghan et al., 21] was how to select proper range (scale j) of wavelet coefficients. Since our method is based on the Abry–Veitch wavelet analysis, we also needed to carefully choose a wavelet coefficient that would mitigate a boundary effect. Roughan et al. [21] suggested the upper bound of scale ($j'_{\max}$) should be less than $\log_2 n$, the largest scale in sampling data, where n is the length of sampling data. Due to initialization errors in wavelet decomposition, the lower bound of the scale ($j'_{\min}$) is not less than 2. However, there is no rule that determines the best scale range for given sampling data. In our practice, we select a range based on visual inspection of log-scale diagrams for a given network environment. We set ($j'_{\min}$, $j'_{\max}$) to (2, 9) for ITA trace, and (2, 10) for simulated trace.

3.3.3. Load effect

In order to check the load effect on energy distribution analysis, we would need to obtain a complete picture of energy distribution analysis for all load levels (based on average link utilization).

However, due to the difficulty of network measurements for all load levels, we perform our investigation on simulated traffic (see figure 2(b)). In the simulation described in section 3.1, we sweep the number of sessions from 500 to 3000 to build workloads with average link utilization varying from 19% to 95%. Note that our method is applicable to traffic with moderate to high load. We found a significant deviation in very light load (only 200 sessions and 6% link utilization). Figure 3 shows the comparison between energy distribution variation of light load and that of moderate load. Lower

[3] Window sizes are 2621.44 and 1310.72 seconds (a power of 2 times 10 ms).

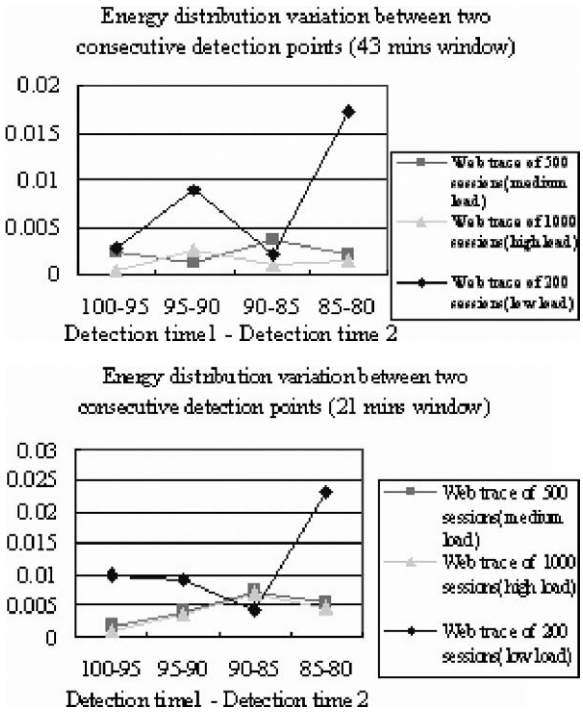


Figure 3. The deviation of energy distribution in light traffic.

traffic base holds a more volatile energy distribution because even a few occasions of modest change is more distinctively reflected to energy distribution than in the moderate or heavy load. Therefore, our energy distribution analysis is limited to traffic with moderate or high load. Since DDoS attack detection does not kick in with low load, this does not limit our approach.

4. Attack detection simulation

Distributed Denial of Service (DDoS) attack detection is one of most typical applications in detection based network control model. DDOS attacks, especially protocol attacks and bandwidth/throughput attacks, produce packets without following regular packet generation mechanism (such as SYN flood) and regular access intensity distribution (such as CodeRed). For example, packet inter-arrival time could be affected by packet length and packet generation pattern. Given a sending rate of packets, large packet size leads to small inter-arrival time. Packet generation patterns (or traffic type), e.g., CBR and VBR, make different distributions of packet inter-arrival time. DDoS attacks have a strong influence on the dynamics of packet inter-arrival time, since they considerably change the distribution of packet length (e.g., Ping of Death) and packet generation patterns (UDP flooding). Attack traffic is capable of making a sudden change, thereby distorting nor-

mal behavior. Stationary energy distribution could be broken since the sudden behavior change distorts temporal correlation over multiple scales.

According to equation (6), our detection method can be considered a threshold-based method. Energy distribution variation caused by irregular behavior (e.g., attack) could violate the threshold δ. We then can detect DDOS attack causing significant variation of energy distribution. In order to detect the attack as early as possible, sequential sampling with a sliding window W is employed. For every step increment T, traffic is sampled with window size of W. For online detection, energy distribution analysis should be completed in time of T.

4.1. Simulation environment

In order to simulate attack flows over background traffic with self-similarity, we constructed a large-scale network simulation test-bed through the NS simulator [25].

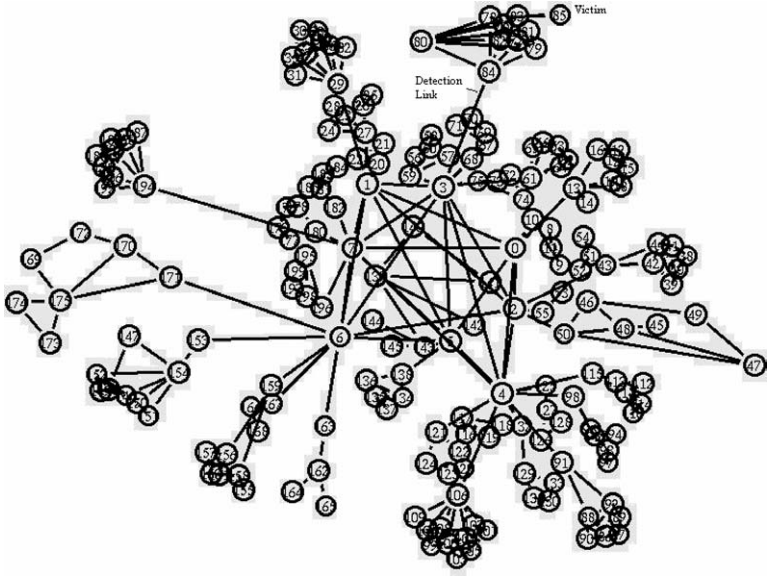
All the network end nodes in the simulation are assumed to be both IP traffic generators and receivers. General nodes and hot spot nodes (victim) would be simulated. In order to avoid the deterministic impact of the statistic-based traffic generator (such as Pareto and exponential), application-based traffic sources are selected in our simulation, e.g., Web, ftp, and CBR. In the simulated network, a number of network nodes are selected to be attacker nodes. Unlike the normal traffic generator, IP traffic from all attackers has the same destination: the victim node. According to the non-responsive feature, the CBR traffic source is chosen for simulating a UDP flooding attack. The attack scenarios simulated are based on attack observations done by [6; Moore et al., 16].

We have two scenarios: 0.05 (scenario 1) and 0.075 (scenario 2) attack coverage (the ratio of attack nodes to whole nodes). In each scenario, cases with attack and without attack are both simulated and the DDoS attack is launched at 3100 s with exponential acceleration having a knee point at 3500 s (network topology and other parameters including attack configuration for the experiments are described in figure 4).

4.2. Simulation result

Similar to the first step, the self-similarity of traffic is extracted by estimating the Hurst parameter. As in the case of ITA trace, the simulated traffic also exhibits a fairly high self-similarity (with and without attack, all scenario cases produced Hurst parameter values in the range of 0.7 and 0.8). We then applied our method to compute the energy distribution variation of different traces.

In the simulation results collected, large difference between consecutive detection points were observed in the traces with attack (see figure 5). As a comparison, the traces without attack have a very limited energy distribution variation, which is very similar to what we obtained from the ITA trace. With a threshold of 0.01, our scheme was able to catch all attack cases; four of four cases. The catch points matches with attack launch timing shown in figure 4(c). Under various scenarios and parameters, catch points are all around 3400 s. One important note here is that energy distribution variation analysis



GT-ITM topology generator is used to generate a three-level network for simulation. Following topology generation parameters are specified:

(1) ratio of end nodes and intermediate nodes,

(2) connection density of the network nodes and

(3) link bandwidth assignment.

(1) Three-level hierarchy: domain, cluster, and nodes.

(2) 10 domains; 4 clusters every domain; 5 nodes every cluster.

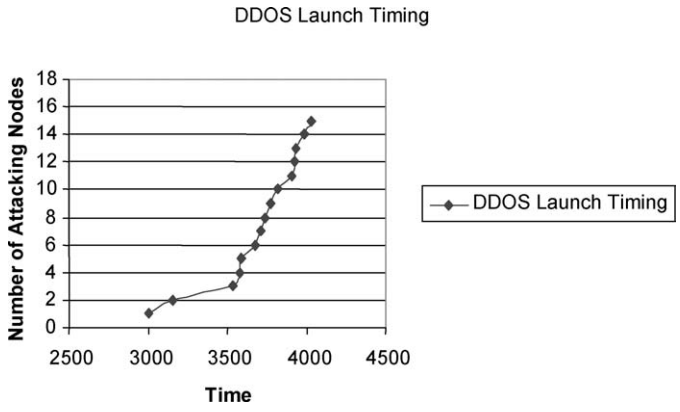
(3) 10 Mps link for domain; 5 Mps for clusters; 2 Mps for nodes

(a)

Total nodes	200
Number of background flows	800 (4 flows every nodes in average)
Attack coverage	0.05/0.075 (10/15 nodes, out of 200 total nodes, to be attackers or zombies)
Attack launch curve	Exponential distribution (see figure 4(c))
Simulation time	6000 seconds
Attack period	From 3000 s to 5000 s
Attack launching period	3100–4000 s
Victim node	85
Detecting path	$3 \rightarrow 84 \rightarrow 85$ (node #3 is the gate way collecting data in the simulation)

(b)

Figure 4. Simulation configurations: (a) network simulation topology; (b) attack configuration; (c) DDoS launch timing.



(c)

Figure 4. (Continued.)

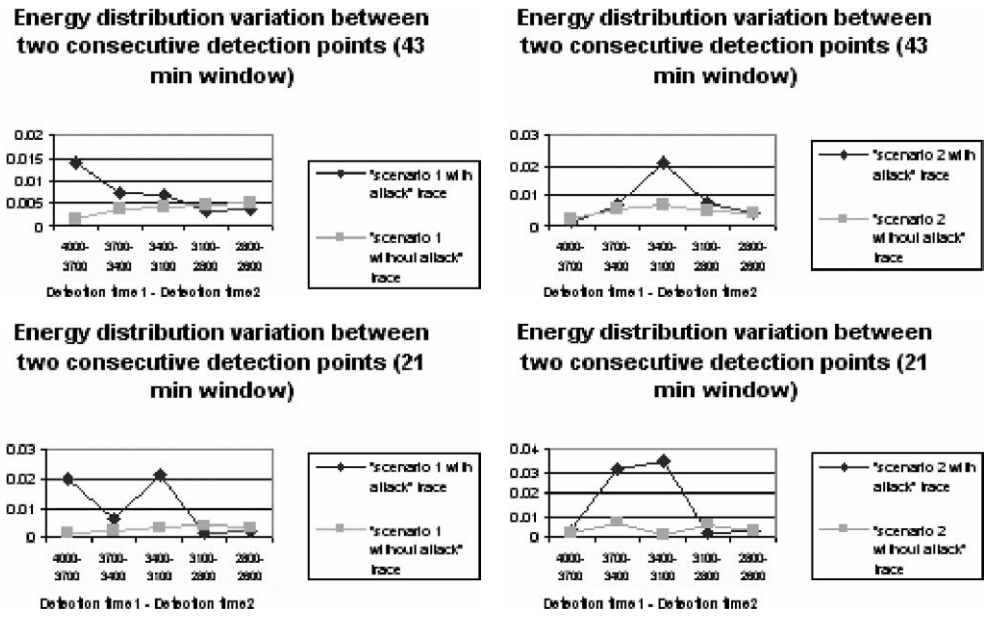


Figure 5. Energy distribution variations in simulated trace.

is able to catch attacks early in the attack launch, far ahead of congestion build-up due to the attacks.

In contrast, we also show the variation of traffic rate[4] in figure 6. It is clear that DDoS attack elicits a significant rate change in the traffic. However, rate variation in the early stage of an attack (before 3800 s in scenario 1) may not be detected by rate

[4] We only present the result of scenario 1 because of space limitation. Also, we have better detection in scenario 2.

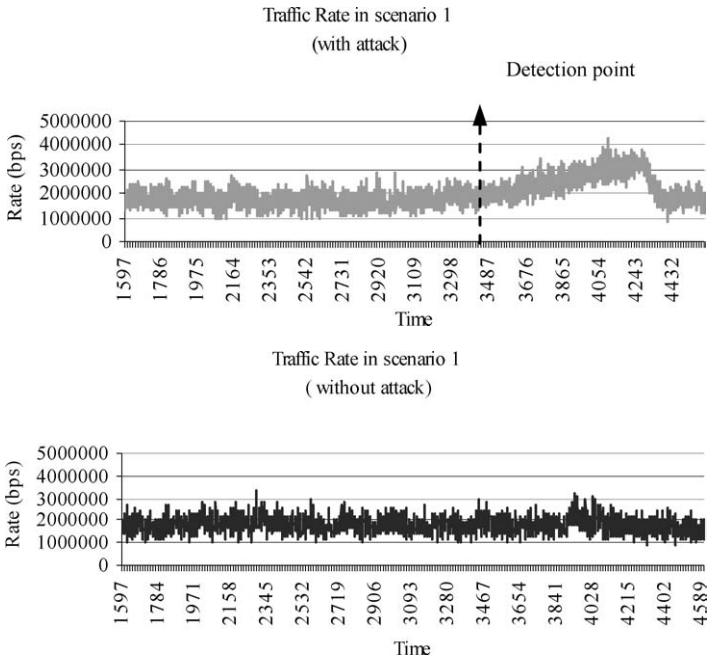


Figure 6. Traffic rate in scenario 1.

watching schemes (e.g., rate threshold) since it is as limited as normal burst. When a significant variation occurs (around 4300 s), congestion has already built up. With respect to the detection point showing in figure 5, energy distribution analysis with a sliding window of 21 minutes can detect attack at 3400 seconds, far ahead of congestion.

4.3. Discussion

We can successfully utilize energy distribution to detect DDOS attack in simulation. The deviation of energy distribution variation caused by attack traffic is significant enough to be detected through a threshold $\delta = 0.01$. System parameters δ, T and W are chosen tentatively. With too large W and/or T, energy distribution variation may be buried under self-similarity, while too small W and/or T will make a less meaningful stochastic sample. Threshold δ and window size T need in-depth investigation with diverse traffic environment. Developing a proper value for δ, T and W in various contexts is a key component of our future research.

Based on our reiterative experiments, what we can learn right now is that the size of W is determined by the appropriate sample size and sampling interval. In order to characterize the normal traffic behavior with fewer faults negative and positive, a proper size of sample data needs to be chosen. In our experiments, we found that only the 21 minutes and 43 minutes window can represent normal traffic behavior properly if we apply a 10 ms sampling interval. Recall that wavelet analysis in use requires that the number of sampling data should be power of 2. One reason of using the time increment

step (T) is that we regard the impact of renewal trace (size of T) on old one (size of $W - T$) which may already be conformed to "normal". In addition, using T increases the responsiveness of detection since the size of the window is too large to catch the behavior change promptly. Obviously, T and W have a strong relationship; a constant ratio probably exists between them. From our experiments, it is feasible making T 25–10% of W. The determination of δ is relatively complicated. Since our scheme is also a threshold-based scheme, the definition of "normal" behavior is not arbitrary. In order to determine δ, an initiation procedure and adaptive adjustment are needed.

Another issue is sampled target (traffic parameter). Time series $x(t)$ can be sampled for any traffic parameter. Instead of using inter-arrival time, other traffic parameters can be considered such as connection amount, packet address distribution, etc. These parameters may represent traffic behavior in different aspects. For diverse applications beyond DDoS detection, we may find a more suitable traffic parameter through which the efficacy of our method can be improved.

Although this method is still a reactive one and catches attacks behind starting point, it can detect attacks before they are launched (cooperate with real-time wavelet analysis) far ahead of congestion build-up. Compared to the resource control scheme (such as rate limitation) associated with SRD characteristics of traffic (such as mean, and variance), our method may have better responsiveness and accuracy. Misuse detection schemes, based on preformed patterns, may recognize known "bad" behavior. However, recognizing a "known" pattern may cause rather serious overhead due to packet decomposition at a high level (such as IP address or TCP/UDP port checking). In addition, knowing how to deal with unknown behavior in a proactive way is a very complicated issue with no known acceptable solution. Therefore, our method may outperform existing schemes in attack detection. In cooperation with the distributed detection mechanism, we can envision better performance for attack detection of the entire network.

5. Conclusion

This work is inspired by the fact that abnormal traffic behavior imposed by DDoS attack can be detected via energy distribution based on wavelet analysis. We have shown the potential of energy distribution analysis for characterizing network traffic behavior. Wavelet analysis is able to capture complex temporal correlation across multiple time scales with very low computational complexity. Wavelet analysis provides energy distribution data for complete information of traffic behavior. With the examination of both real and simulated traffic trace, we have shown that energy distribution remains relatively stationary if the traffic has no characteristics/behavior change. Energy distribution analysis based on wavelet analysis then has been developed.

We have applied the energy distribution analysis to detect DDOS attack as a case study to verify the detection capacity of our method. Parameters of detection method, time step increment T, threshold δ, and sampling window size W, have been studied. Tentative parameter values drawn from our experimental experience have been utilized in simulation. Our results show that energy distribution variance changes markedly when

attack traffic is injected while normal traffic exhibits a remarkably stationary energy distribution. Our experimental results confirm that energy distribution analysis can characterize behavior of network traffic under dynamic condition and outperform other existing schemes.

In our study, only inter-arrival time has been used to construct the time series because it has been widely used in modeling self-similar traffic. However, some other traffic parameters could be considered, such as connection amount, packet address distribution, etc. They represent network traffic behavior in different perspectives. Extending our method to those parameters may improve characterization/detection performance. Energy distribution has great potential to help make better decisions for network control and management. We will also experiment with varying time step increments T and sampling window sizes W in different network environment and develop a method through which parameters can be adaptively adjusted.

Acknowledgement

This work is supported in part by the NSF grant CCR-0242222 and CCR-0209078.

References

[1] P. Abry and D.Veitch, Wavelet analysis of long range dependent traffic, IEEE Transactions on Information Theory 44(1) (1998) 2–15.
[2] P. Abry, D. Veitch and P. Flandrin, Long-range dependence: Revisiting aggregation with wavelets, Journal of Time Series Analysis 19 (May 1998) 253–266.
[3] P. Barford and M.E. Crovella, Generating representative Web workloads for network and server performance evaluation, in: *ACM SigMetrics* (1998) pp. 151–160.
[4] P. Barford, J. Kline, D. Plonka and A. Ron, A signal analysis of network traffic anomalies, in: *Internet Measurement Workshop* (November 2002).
[5] CERT, Overview of attack trends, `http://www.cert.org/archive/pdf/attack_trends.pdf` (8 April 2002).
[6] CERT Advisory CA-2001-23, Continued threat of the Code Red Worm (17 January 2002).
[7] R.K.C. Chang, Defending against flooding-based distributed denial-of-service attack: A tutorial, IEEE Communication Magazine 40(10) (2002) 42–51.
[8] M.E. Crovella and A. Bestavros, Self-similarity in World Wide Web traffic: Evidence and possible causes, IEEE/ACM Transactions on Networking 5(6) (1997) 835–846.
[9] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Transactions on Information Theory 36 (September 1990) 961–1005.
[10] P. Ferguson and D. Senie, Network ingress filtering: Defeating denial of service attacks which employ IP address spoofing, Internet draft (January 1998).
[11] K. Fox, R. Henning, J. Reed and R. Simonian, A neural network approach towards intrusion detection, Technical Report, Harris Corporation (July 1990).
[12] M. LaMonica, Microsoft releases anti-Slammer tools, `http://zdnet.com.com/2100-1105-983603.html` (6 February 2003).
[13] W. Leland, M. Taqqu, W. Willinger and D. Wilson, On the self-similar nature of Ethernet traffic, Proc. of ACM SIGCOMM 23(4) (1993) 183–193.
[14] W. Leland, M. Taqqu, W. Willinger and D. Wilson, On the self-similar nature of Ethernet traffic (extended version), IEEE/ACM Transactions on Networking 2(1) (1994) 1–15.

[15] S. Ma and C. Ji, Modeling heterogeneous network traffic in wavelet domain, IEEE/ACM Transactions on Networking 9(5) (2001) 634–649.
[16] D. Moore, C. Shannon and J. Brown, Code-Red: A case study on the spread and victims of an Internet worm, in: *Proc. of Internet Measurement Workshop* (2002).
[17] K. Park and H. Lee, On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack, in: *Proc. of IEEE INFOCOM'01* (April 2001) pp. 338–347.
[18] V. Paxson and S. Floyd, Wide area traffic: The failure of Poisson modeling, IEEE/ACM Transactions on Networking 3(3) (1995) 226–244.
[19] R. Riedi, M.S. Crouse, V. Ribeiro and R.G. Baraniuk, A multifractal wavelet model with application to TCP network traffic, Special Issue on Multiscale Statistical Signal a Nalysis and Its Applications of IEEE Transactions on Information Theory 45 (April 1999) 992–1018.
[20] R. Ritke, X. Hong and M. Gerla, Contradictory relationship between Hurst parameter and queueing performance (extended version), Telecommunication Systems 16 (February 2001) 159–175.
[21] M. Roghan, D. Veitch and P. Abry, Real-time estimation of the parameters of long-range dependence, IEEE/ACM Transactions on Networking 8 (August 2000) 467–478.
[22] S. Sarvotham, R. Riedi and R. Baranuik, Connection-level analysis and modeling of network traffic, in: *Proc. of the ACM SIGCOMM IMW* (November 2001).
[23] S. Staniford, V. Paxson and N. Weaver, How to own the Internet in your spare time, in: *Proc. of the 11th USENIX Security Symposium* (August 2002).
[24] The Internet traffic archive, `http://ita.ee.lbl.gov/` (April 2000).
[25] The network simulator-ns-2, `http://www.isi.edu/nsnam/ns/` (July 2003).
[26] X. Tian, J. Wu and C. Ji, A unified framework for understanding network traffic using independent wavelet models, in: *Proc. of IEEE INFOCOM* (June 2002).
[27] N. Ye, A Markov chain model of temporal behavior for anomaly detection, in: *Workshop on Information Assurance and Security* (June 2000).
[28] Z. Zhang, V.J. Ribeiro, S. Moon and C. Diot, Small-time scaling behaviors of Internet backbone traffic: An empirical study, in: *Proc. of IEEE INFOCOM* (April 2003).