DDoS 공격 탐지를 위한 패킷 샘플링 기법들의 성능 분석

강 길 수[†] · 이 준 희[†] · 최 경 희[††] · 정 기 현[†††] · 심 재 홍[††††]

요 약

일반적으로 패킷 샘플링은 네트워크의 관리 및 보안 등의 목적으로 네트워크를 통과하는 패킷들 중 일부만을 수집하여 분석하는 기법이다. 본 논문에서는 분산 서비스 거부(DDoS) 공격을 효율적으로 탐지하고 트래픽 분석 성능을 높이기 위해, DDoS 공격 탐지 기술에 다양한 패킷 샘플링 기법들을 적용하고 이들이 DDoS 탐지 성능에 어떠한 영향을 미치는지 비교 분석한다. 트래픽 특성 분석을 위해 DDoS 탐지 기법의 하나인 기존의 트래픽 비율 분석법을 사용하여 실험하였다. 실험 결과 DDoS 공격 탐지를 위한 패킷 샘플링 기법의 사용은 모든 패킷을 조사하지 않고도 기존의 트래픽 비율 분석법과 비슷한 성능을 보이는 것을 확인할 수 있었으며, DDoS 공격 탐지의 정확성을 유지하면서도 분석 트래픽 양을 현저히 줄일 수 있었다.

Performance Analysis of Packet Sampling Mechanisms for DDoS Attack Detection

Kil-Soo Kang[†] · Joon-Hee Lee[†] · Kyung-Hee Choi[††]
Gi-Hyun Jung[†††] · Jae-Hong Shim[††††]

ABSTRACT

Packet sampling is the techniques to collect a part of the packets through network and analyze the characteristicsof the traffic for managing the network and keeping security. This paper presents a study on the sampling techniques applied to DDoS traffic and on the characteristics of the sampled traffic to detect DDoS attack efficiently and improve traffic analysis capacity. Three famous sampling techniques are evaluated with different sampling rates on various DDoS traffics. To analyze traffic characteristics, one of the DDoS attack detection method, Traffic Rate Analysis (TRA) is used. Simulation results verify that using sampling techniques preserve the traffic characteristics of DDoS and do not significantly reduce the detection accuracy.

키워드 : 분산 서비스 거부 공격 탐지(DDoS Attack Detection), 패킷-기반 샘플링(Packet-based Sampling), 트래픽 비율 분석(Traffic Rate Analysis)

1. 서 론

DDoS(distributed denial of service) 방어에서 탐지(detection)란 DDoS 공격이 발생했을 때 이를 감지하여 방어자에게 알려주는 것을 의미하며 DDoS Reactive[1] 탐지 전략을 통해 DDoS 공격이라고 판단될 경우, 이에 대해 적절한 대응을 수행하는 DDoS 공격 방어 기법이다. 방어기법의 필수적인 요소이다[4]. 또한 IDS(intrusion detection system)의 중요한 탐지영역 중의 하나에 속하기도 한다. DDoS 탐지에서 중요한 것은 탐지의 정확성과 탐지의 신속성이다. 탐지자는 DDoS공격의 발생을 정확하게 판단하고 신속하게 방어자에게 알릴 수 있어야 한다. 만약 DDoS가 아닌 트래픽에 대해 DDoS 발생을 알리거나, 이미 공격이 진행되어 시스템이 사용 불가된 상태에서 경고한다면 아무 의미가 없을 것이다.

DDoS 탐지는 기본적으로 트래픽의 흐름과 패킷 헤더 필드의 내용을 조사함으로써 이루어진다. 따라서 패킷을 분석하는 작업이 필수적으로 동반된다. 그러나 현재 네트워크 망의 속도는 점차 가속화되는 추세이다. 이러한 환경에서 패킷 분실이나 지연 없이 실시간으로 패킷을 분석하기에는 많은 어려움이 따르고 추가경비를 요한다. 일반적으로 DDoS 공격의 탐지는 IDS와 같은 시스템을 다른 네트워크 망에 별도로 설치하여 패킷 분석 작업을 수행하지만, 고속의 방대한 네트

† 준 회 원 : 아주대학교 정보통신전문대학원 정보통신과
†† 정 회 원 : 아주대학교 정보통신전문대학원 교수
††† 정 회 원 : 아주대학교 전자공학부 교수
†††† 정 회 원 : 조선대학교 인터넷소프트웨어공학부 교수
논문접수 : 2004년 2월 10일, 심사완료 : 2004년 8월 17일

1) 탐지 전략을 통해 DDoS 공격이라고 판단될 경우, 이에 대해 적절한 대응을 수행하는 DDoS 공격 방어 기법이다

워크 망에서는 이것마저도 힘들어지게 된다. 따라서 네트워크를 통과하는 모든 패킷을 캡쳐하여 분석하는 것이 아니라, 일부 패킷만을 캡쳐하는 샘플링 방법이 유일한 대안이 되고 있다.

본 논문에서는 이러한 패킷 샘플링 기법을 DDoS 탐지 기법에 적용하여 그 효과를 분석해 보고자 한다. 이를 위해 세 가지 샘플링 방법인 단순 무작위 표본 샘플링(simple random sampling), 패킷 기반 체계적 샘플링 기법(packet-based systematic sampling), 패킷 기반 층화 무작위 샘플링 기법(packet-based stratified random sampling) 등을 DDoS 탐지기법 중의 하나인 트래픽 비율 분석법에 적용한 후, 원본 트래픽과 샘플된 트래픽의 특성을 비교 분석함으로써 샘플링 기법들의 성능을 분석하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 샘플링 기법에 관한 기존 연구를 분석한다. 3장에서는 DDoS 공격 탐지 기법으로 사용된 트래픽 비율 분석법을 소개하고, 이를 활용한 실험 방법을 기술한다. 4장에서는 본 논문에서 수행된 실험 결과를 제시한 후 이를 비교 분석한다. 마지막 5장에서 결론을 맺는다.

2. 관련 연구

2.1 패킷 샘플링

패킷 샘플링이란 전체 패킷의 일부분을 샘플링함을 말한다. 예를 들어, 네트워크에 1,000,000 패킷이 흘렀다고 가정하고, 그 중 0.25%(2,500개) 패킷을 무작위로 선택하여 샘플링하였다고 하자. 만약 샘플링된 패킷의 1,000개가 음성 패킷이었다면, 원본 패킷에서는 어느 정도가 음성 패킷이었을까? 최소 1,000개에서 최대 998,500개가 된다. 그러나 이는 모든 가능성에 대한 수치이며 샘플링된 패킷의 40%가 음성 패킷이었으므로, 원본 트래픽의 경우도 마찬가지로 40% 내외로 예측하는 것이 가장 타당할 것이다. 이를 일반적인 수식으로 표현할 경우, 총 패킷(N)에 대한 특정 클래스에 속하는 패킷의 수(N_c)는 다음 수식으로 구해진다.

$$N_c = \frac{c}{n} \cdot N$$

여기서 N은 총 패킷의 수, n은 샘플링된 패킷의 수, c는 특정 클래스에 속한 패킷의 샘플링된 수를 의미한다. 물론 샘플링된 음성 패킷이 정확하게 400,000개의 패킷이 될 확률은 극히 적다. 대신에 95%의 확률로 존재할 범위를 계산한다면, 패킷의 수는 381,000에서 419,000개 사이에 존재할 것으로 추측된다. 다음 식은 N_c의 편차를 나타낸다.

$$\sigma^2 = N^2 \cdot \frac{c \cdot \left(1 - \frac{c}{n}\right)}{n \cdot (n-1)}$$

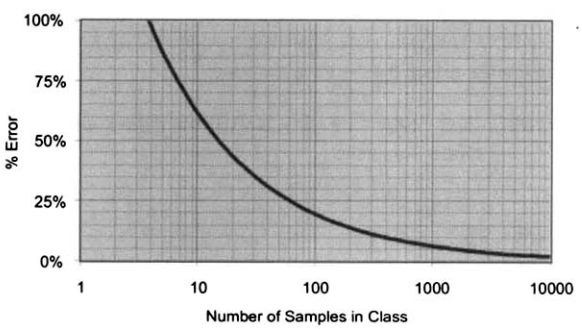
따라서 N_c가 95%의 신뢰구간을 갖는 범위는 다음과 같다.

$$[N_c - 1.96\sigma,\ N_c + 1.96\sigma]$$

보다 간단히 표현하여 음성 패킷이 존재할 수는 400,000 패킷의 4.8%의 오차 한계 범위 내에 존재 한다 라고 말할 수 있다. 다음 공식은 이를 오차율로 표현한 것이다.

$$\%error \le 196 \cdot \sqrt{\frac{1}{c}}$$

위 공식을 도식화한 (그림 1)은 c의 증가에 따른 샘플링 오차율을 보여준다. 그래프를 통해 알 수 있듯이 샘플링의 정확성은 샘플의 수에 의존한다는 성질을 알 수 있다. 따라서 초당 400개의 패킷이 발생하는 네트워크에 0.25% 비율로 샘플을 얻는다면, 초당 40,000개의 패킷이 발생하는 네트워크에서 동일한 오차범위를 갖기 위해선 0.0025%의 샘플링 비율로 패킷을 샘플링하면 된다. 따라서 이론적으로 샘플링 트래픽의 오차범위를 줄이기 위해선 샘플링 비율을 높이거나 오랜 시간 동안 샘플 패킷 데이터를 얻는 것이다.



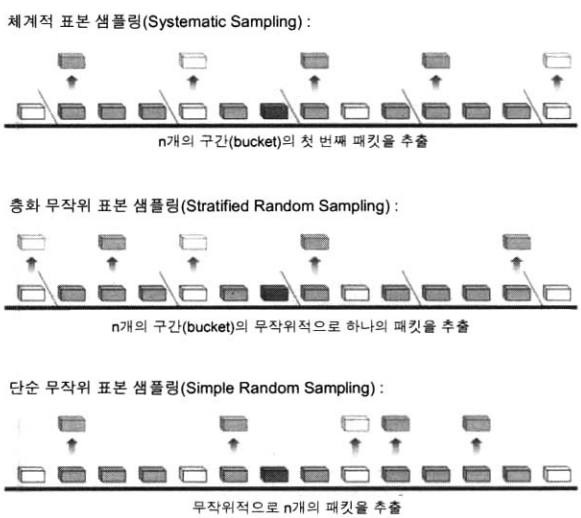
(그림 1) 상대적 샘플링 오차율

2.2 샘플링 기법

패킷 샘플링 기법과 관련한 연구는 네트워크 트래픽 분석을 목적으로 1993년도부터 시작되었다. 주로 네트워크 관리를 위해 필요한 트래픽 통계를 작성하는 작업에 활용되었으며, 방대한 트래픽 양으로 인한 계산 오버헤드를 줄이려는 노력의 일환으로 샘플링 기법을 사용하였다[6]. 최근에는 네트워크 관리나 보안을 위한 실시간 네트워크 트래픽 모니터링이나 또는 계정관리/청구서작성 (accounting/billing) 서비스를 위한 트래픽 플로우 별 트래픽 양 측정을 위해 고속의

네트워크 망에서도 샘플링 기법을 사용하고 있다[5].

이러한 샘플링 기법은 전체 네트워크 트래픽 양의 $1/N$을 확률적으로 샘플링 하는 것을 기본으로 한다[7]. 여기서 N은 표본 집단의 개수를 모집단의 개수로 나눈 것을 말한다. 즉 평균적으로 매 N개의 패킷마다 하나씩 샘플링을 수행한다. 샘플링 방법은 패킷을 샘플링하는 방법에 따라 구간을 나눠 각 구간의 첫 번째 패킷을 샘플링하는 체계적 샘플링(systematic sampling), 각각의 구간 안에서 무작위적으로 패킷을 샘플링하는 층화 무작위 샘플링(stratified random sampling), 그리고 패킷을 단순한 확률에 따라 무작위로 선별하는 단순 무작위 샘플링(simple random sampling) 등이 있다. (그림 2)는 대표적인 이들 세 가지 샘플링 기법을 보여준다[6].



(그림 2) 대표적인 샘플링 기법

그림에서 구간은 하나의 패킷을 선택하는 간격을 의미한다. 일반적으로 구간은 일정한 패킷의 수나 일정한 시간간격으로 정의할 수 있으며, 전자를 패킷 기반 방법(event based), 후자를 시간 기반 방법(time based)이라 한다[6]. 실제 구현에 있어 패킷 기반 방법은 매번 도달하는 패킷에 대해 그 수를 계산하고 그 수가 구간 크기에 도달하면 새로운 구간이 시작하게 된다. 이때 체계적 샘플링 기법의 경우 그 다음 패킷을 샘플링하고, 층화 무작위 샘플링 기법의 경우 구간 안에 속하는 무작위 수를 발생해 그 수에 도달하는 패킷을 샘플링하게 된다. 그러나 시간 기반 방법의 경우 일정한 시간을 계산해야 하므로 하드웨어적 성능 문제로 인해 실제로 구현하여 사용한 예는 없다. 따라서 본 논문의 실험에서는 시간 기반 샘플링 기법은 평가에서 제외하였다.

기존의 연구 결과에 따르면 위의 샘플링 기법들 중 패킷 기반의 층화 무작위 샘플링 기법이 가장 원본 트래픽의 특성을 잘 반영하고 있다고 보고되어 있다[6].이 샘플링 기법은 현재 sFlow[3]나 Sampled NetFlow[5] 등의 상용 모니터링 도구에서 기본적으로 사용하는 방법이기도 하다.

그 외 기법으로 전체 트래픽의 플로우 중 가장 많은 비율을 차지하는 플로우의 패킷만을 선별해 트래픽 양을 측정하는 샘플링 기법도 있다[2]. 이 기법은 오차율을 줄일 수 있는 장점이 있으나 가장 많은 비율의 플로우 크기가 나머지 플로우들에 비해 매우 클 경우에만 유용하다.

3. DDoS 공격 탐지기법

본 논문에서는DDoS 공격시의 패킷들을 여러 샘플링 기법에 적용해서 샘플링하고, 샘플링된 패킷들의 특성을 다시 트래픽 비율 분석법을 통해 분석함으로써, 여러 샘플링 기법들이 DDoS 탐지에 어떠한 영향을 미치는지 비교 분석하고자 한다. 따라서 본 장에서는 기존 트래픽 비율 분석법을 활용한 DDoS 공격 탐지에 패킷 샘플링 기법을 적용하는 방안에 대해 논의하고자 한다.

3.1 트래픽 비율 분석법을 이용한 DDoS 탐지기법

본 논문에서는 DDoS 공격 탐지를 위해 트래픽 비율 분석법(traffic rate analysis)[1]을 사용한다. 트래픽 비율 분석법이란 전체 트래픽에서 특정한 형태를 가진 트래픽의 비율(rate)을 이용하여 트래픽을 분석하고 DDoS 공격의 탐지를 수행하는 기법이다. 트래픽 비율 분석법은 TCP 플래그 비율(TCP flag rate)과 프로토콜 비율(protocol rate)로 나누어진다. 본 논문의 실험에서는 TCP 플래그 비율을 사용하였으며, 이는 다음의 수식으로 정의된다.

$$R_{td}[K_i] = \frac{total\ number\ of\ a\ flag(K)\ in\ a\ TCP\ header}{total\ number\ of\ TCP\ packets}\ (inbound)$$

$$R_{td}[K_o] = \frac{total\ number\ of\ a\ flag(K)\ in\ a\ TCP\ header}{total\ number\ of\ TCP\ packets}\ (outbound)$$

여기서 TCP 플래그 비율은 전체 TCP 패킷에 대한 특정 플래그를 가진 TCP 패킷의 비율이며, td는 트래픽 비율을 측정한 시간 간격[2]을 의미한다. 또한 K는 SYN, FIN, RST, ACK, PSH, URG, NULL 등의 TCP 플래그를 나타내며, 이들 플래그는 각각 S, F, R, A, P, U, N 등의 약어로 표현된다.

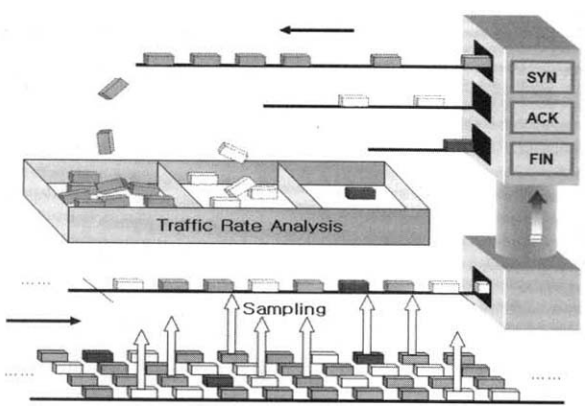
특정플래그의 비율 변화를이용하여 DDoS 공격을 감지하는 것이 트래픽 비율 분석법의 기본적인 작동 원리이다. 예를 들어, SYN Flooding 공격이 시작될 경우 SYN 플래그를

2) 본 논문에서의 td 값은 1초로 통일하였다.

가진 TCP 패킷이 급격히 증가하게 되고 이는 $R_{td}[S_i]$의 값이 증가하게 되는 원인이 된다. 이것을 이용하면 SYN Flooding 공격을 탐지할 수 있다. 트래픽 비율 분석법은 TCP의 모든 플래그를 포함해서 TCP, UDP, ICMP 프로토콜의 비율도 함께 고려하므로, SYN Flooding 공격뿐만 아니라 다양한 DDoS 공격의 탐지에 이용될 수 있다는 장점이 있다[1]. 또한 통계 기반의 탐지 기법이므로 모집단이 되는 본래의 트래픽으로부터 샘플링한 표본 집단으로서의 패킷들에 대해 그대로 적용할 수 있어 본 실험의 DDoS 탐지 기법으로서 채택하였다.

3.2 패킷 샘플링을 적용한 DDoS 공격 탐지

트래픽 비율 분석법을 사용하여 DDoS를 탐지하려 할 경우 네트워크를 통해 지나가는 모든 패킷에 대한 분석 작업을 수행해야 한다. 그러나 고속의 네트워크에서는 빠른 속도의 방대한 트래픽 양으로 인해 모든 패킷을 분석할 수 없는 경우가 발생하게 되며, 이는 곧 DDoS 공격 탐지를 위해 기존 트래픽 비율 분석법을 바로 적용할 수 없다는 것을 의미한다. 또한 이러한 고속의 네트워크 환경에서는 실시간 패킷 캡쳐, 분석, 계산 등의 오버헤드도 지나치게 커지게 된다. 이러한 문제를 해결하기 위한 방법의 하나로 분석할 패킷의 양을 줄이는 방법이 있다. 분석할 패킷의 양을 줄이기 위한 적절한 대처 방안으로 패킷 샘플링을 고려해 볼 수 있다.



(그림 3) 패킷 샘플링과 트래픽 비율 분석법을 활용한 DDoS 공격 탐지

본 논문에서는 샘플링 기법으로 단순 무작위 샘플링, 체계적 샘플링, 층화 무작위 샘플링 등을 사용한다. 고속의 네트워크를 통해 들어온 패킷들은 먼저 샘플링을 거쳐 그 양을 줄이게 된다((그림 3) 참조). DDoS를 탐지할 대상 패킷의 수를 줄임으로써 더욱 큰 트래픽에 대한 DDoS의 탐지를 가능하게 한다. 이렇게 트래픽으로부터 샘플링 된 패킷들은 트래픽 비율 분석을 하게 된다. 트래픽 비율 분석은 패킷들에 대한 특정 플래그들의 각각의 비율을 계산한다. 본 논문에서는 TCP 플래그 비율 분석법을 사용하므로 SYN, FIN, ACK, RST, PSH 등의 플래그를 기준으로 각각의 플래그 개수를 세고 전체 TCP 패킷 수에 대한 각 플래그의 비율, 즉 $R[S_i]$, $R[F_i]$, $R[A_i]$, $R[R_i]$, $R[P_i]$ 등을 구한다. 이때 특정 플래그의 비율이 어떤 특성을 보인다면 DDoS 공격으로 간주된다. 플래그 비율 분석에서는 탐지 룰을 적용하여 플래그들이 보이는 비율의 특성을 탐지해내는데, 탐지 룰에 따라 DDoS를 판단하는 기준이 각각 다르다. 예를 들어, 어떤 탐지 룰은 $R[S_i]$의 비율이 0.4를 넘어가면 DDoS로 판단한다. 만약 SYN Flooding 공격이 이루어 진다면 $R[S_i]$의 비율이 비정상적으로 높아질 것이다. $R[S_i]$의 비율이 탐지 룰에서 정해 놓은 값을 넘어서면 이것은 DDoS 공격이 시작되었음을 뜻하게 되고 탐지 룰은 이를DDoS 공격으로 판단한다.

앞에서 언급한 바와 같이 본 논문에서 사용할 트래픽 비율 분석법은 패킷의 특성을 분석하여 DDoS를 탐지해 내는 기법이다. 그러나 샘플링을 통해 트래픽 비율 분석법에 적용될 패킷들의 수를 제한했다면, 여러 패킷들을 모아 일정 시간마다 모은 패킷들의 플래그 비율을 분석하도록 고안된 트래픽 비율 분석법의 특성상 DDoS 탐지 결과에 영향을 미치게 된다.

따라서 본 논문에서는 기본적인 패킷 샘플링 기법들을 사용해 DDoS가 발생한 트래픽의 특성을 트래픽 비율 분석법을 사용하여 분석한다. 또한 패킷 샘플링 기법과 트래픽 비율 분석법 등이 DDoS 탐지 기법과 결합했을 때 이들이 탐지 결과에 미치는 영향을 분석하고 각 샘플링 기법의 성능상의 특징들을 논의하고자 한다.

4. 모의 실험

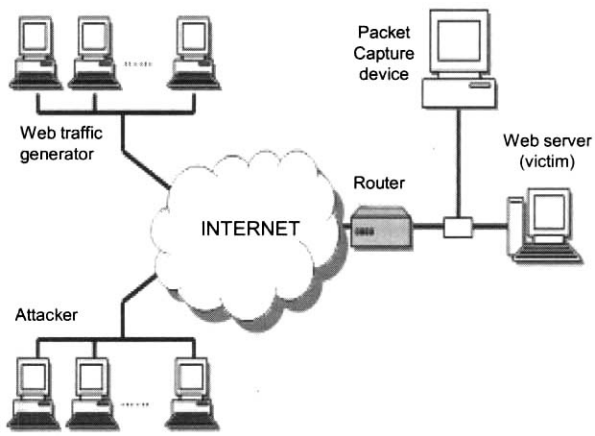
본 장에서는 생성된 트래픽 데이터를 각각의 샘플링 기법들을 채택한 시뮬레이터에 적용하여 실험한 실험결과를 보인다. 또한 샘플링 기법 별 DDoS 트래픽의 특성 변화를 관찰하고, DDoS 탐지의 정확도를 비교 분석한다.

실험 결과 먼저 DDoS 공격 시 부분적으로 샘플링 하여 얻은 트래픽은 본래의 트래픽 특성을 쉽게 잃지 않음을 확

인하였고, 또한 DDoS 탐지의 정확성과 신속성이라는 측면에서 샘플링 기법의 종류는 크게 영향을 미치지 않음을 확인할 수 있었다.

4.1 실험 환경

샘플링 기법들을 공평하게 비교하기 위해 미리 수집한 패킷 데이터를 바탕으로 동일한 환경에서 시뮬레이션을 수행하였다. 패킷 데이터는 (그림 4)와 같은 실험환경을 구성해 총 1,800초(30분) 동안 웹 트래픽과 DDoS 트래픽들을 발생시켰다.



(그림 4) 패킷 샘플링 실험환경

웹 트래픽은 Microsoft Windows XP 시스템에서 Microsoft Application Stress Tool을 이용하여 10개의 쓰레드(thread)를 생성한 후, 각 쓰레드 당 10개의 소켓을 생성하여 접속하였다. DDoS 트래픽의 발생은 TFN2K[8]를사용하였다. DDoS 공격의 경우 탐지의 정확도를 측정하기 위해 임의의 시간을 두어 여러 번 발생시켜 공격 트래픽이 웹 트래픽과 적절히 섞이도록 했다. <표 1>은 DDoS 공격을 발생시킨 구간을 보여준다.

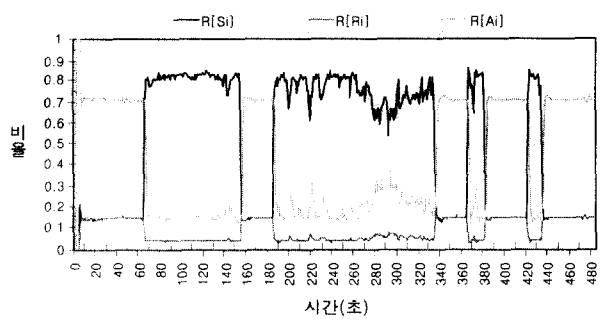
생성된 트래픽은 tcpdump를 이용해 샘플링하여 실험에 사용할 약 1기가 바이트 분량의 패킷 데이터를 수집하였다. 수집된 패킷 데이터는 샘플링 모듈과 트래픽 비율 분석 DDoS 탐지 모듈로 구성된 시뮬레이션 프로그램에 입력되어 결과를 측정하였다. 샘플링 기법으로 단순 무작위 샘플링, 체계적 샘플링, 층화 무작위 샘플링 등을 이용하였고, 탐지 모듈로서 트래픽 비율 분석법을 이용한 DDoS 탐지 코드를 이용해서 실험하였다. 이 모든 프로그램은 리눅스 플랫폼에서 C코드로 작성되었다.

〈표 1〉 DDoS 발생 구간

구분		DDoS		DDoS		DDoS		DDoS		DDoS		DDoS
시간(초)	60	90	30	150	30	20	40	10	50	5	55	3
누적	60	150	180	330	360	380	420	430	480	485	540	543
시간(초)	57	600	30	150	30	60	30	60	60	30	150	
누적	600	1200	1230	1380	1410	1470	1500	1560	1620	1650	1800	

4.2 DDoS 트래픽의 특성 변화

(그림 5)의 트래픽 특성 그래프[3]는 트래픽 비율 분석법에 의한 그래프로서 시간대별 TCP 패킷의 SYN(S_i), RST(R_i), ACK(A_i) 등의 플래그 비율 변화를 나타낸 것이다. 그림에서 SYN 플래그의 비율이 많아지는 구간이 DDoS SYN Flooding 공격이 들어오는 구간이며, [60, 150], [180, 330], [360, 380], [420, 430] 등의 네 개 구간에서 DDoS 공격을 인위적으로 발생시켰다. 나머지 구간은 Web 트래픽을 나타내고 있다.



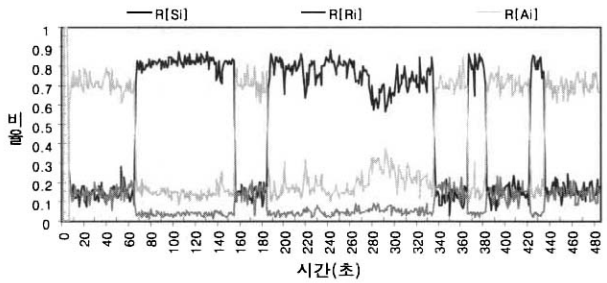
(그림 5) 시간 별 플래그 비율 변화

(그림 6)은 (그림 5)의 원본 네트워크 트래픽을 기반으로 하여 샘플링 된 트래픽을 보여준다. 그림에서 세 그래프는 각각 1/20, 1/100, 1/1000 등의 구간으로 샘플링 되었다. 다음 절에서 논의하겠지만 DDoS 공격 때의 트래픽은 샘플링 기법에 크게 영향을 받지 않기 때문에 샘플링 기법에 따른 플래그 비율 변화는 생략하였다.

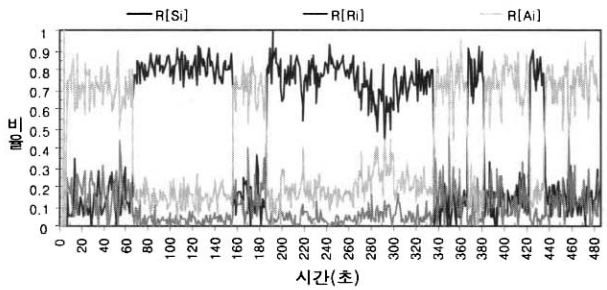
그림에서 알 수 있듯이 1/100 이하의 샘플링에서는 원본 DDoS 공격 때의 트래픽 특성을 쉽게 잃지 않아 DDoS 탐지 성능이 거의 떨어지지 않음을 알 수 있다. 사실 1/1000 샘플링 비율에서도 DDoS 탐지 성능은 90%를 상회한다. 따라서 DDoS 공격을 받고 있는 트래픽에서 샘플링 기법

3) 비율 변화를 상세하게 표현하기 위해 0초~485초 사이의 그래프만을 표시하였다.

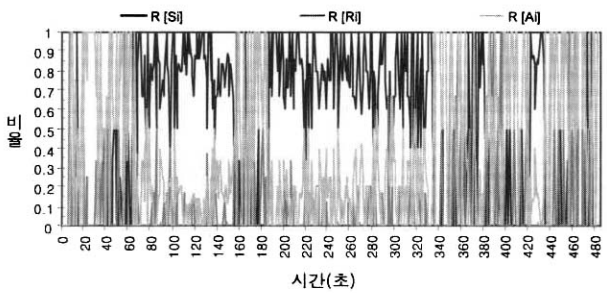
의 적용은 DDoS 트래픽의 특성을 그대로 반영하고 있음을 알 수 있다.



(a) 1/20 샘플링 비율



(b) 1/100 샘플링 비율



(c) 1/1000 샘플링 비율

(그림 6) 샘플링 적용시간 별 플래그 비율 변화

4.3 DDoS 탐지의 성능 비교

본 장에서는 각 패킷 샘플링 기법들의 DDoS 탐지 성능을 비교해 본다. DDoS 탐지 성능이라 함은 DDoS 발생시 이를 얼마나 정확한 시간에 올바르게 알려주는가를 의미한다. 실험 결과를 통해 각 패킷 샘플링 기법들이 트래픽 비율 분석을 채택한 DDoS 탐지 성능에 어떠한 영향을 미치는지를 확인할 수 있다.

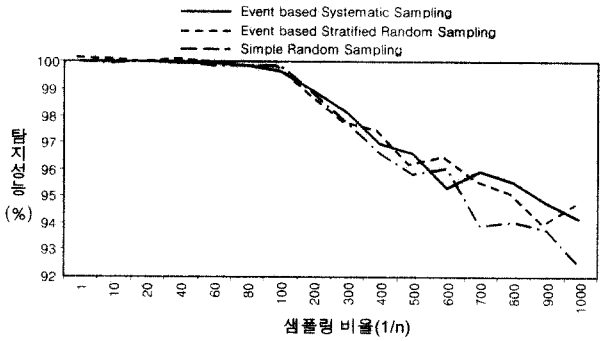
(그림 7)은 패킷이 도착하는 이벤트를 기반으로 한 체계적 샘플링 기법(event based systematic sampling)과 층화 무작위 샘플링 기법(event based stratified random sampling), 그리고 단순 무작위 샘플링 기법(simple random sampling) 등의 샘플링 기법들의 샘플링 비율 별 DDoS 탐지 성능을 보여준다. 이 때 탐지 성능이란 실제로 DDoS가 일어났을 때와 그렇지 않은 경우를 정확히 구분해 내는 정도를 말한다. 즉 100%의 확률에서 오류부정(false negative)의 확률과 오류긍정(false positive)의 확률을 뺀 결과이다. 오류부정이란 DDoS 공격임에도 불구하고 그렇지 않다고 판단하는 경우이고, 오류긍정이란 DDoS 공격이 아님에도 불구하고 DDoS 공격으로 잘못 판단하는 경우를 뜻한다. 예를 들어 100회의 DDoS 탐지를 시도했을 때, 실제로 DDoS 공격이 이루어지고 있을 경우와 DDoS 공격이 이루어지고 있지 않음을 정확히 판단한 경우가 90회이고, 나머지 10회는 실제의 상황과 반대로 판단했다고 가정하면, 이 때의 탐지 성능은 $\frac{100-10}{100} \cdot 100(\%)$, 즉 90%가 된다.

〈표 2〉 TRA 룰

Number	Rule Definition
1	if $R_i[S_i] \geq 0.4$ *or* $R_i[A_i] \geq 0.97$ *then DDoS attack*
2	if $R_1[S_i] > 0.4$ *then DDoS attack*
3	if $R_i[A_i] \leq 0.442884$ *then DDoS attack*
:	:
30	if $R_i[S_i] > 0.48$ *then DDoS attack* *else* if $R_i[A_i] \geq 0.98$ and $R_i[S_i] \geq 0.0.2$ *then DDoS attack*

(그림 7)의 실험결과는 트래픽 비율 분석(TRA) DDoS 탐지에 있어서 각 패킷 샘플링 기법 별로 정확성이 얼마만큼 차이가 나는지를 보여준다. 정확성 비교를 위한 탐지는 기계학습을 통해 생성된 30개의 탐지 룰의 평균치와 탐지 확률이 높은 룰의 수치를 비교 하였다. TRA 룰은 <표 2>와 같이 정의 되었고, 이 중 두 번째 탐지 룰이 가장 높은 탐지 확률을 보였다[1]. 따라서 샘플링을 통한 DDoS의 탐지에 대한 본 논문의 실험에서는 실험의 효율성을 위해 DDoS 탐지율이 가장 높은 두 번째 룰을 적용하기로 하였다. 이 룰은 전체 트래픽 비율 중 SYN 플래그의 비율이 0.4 즉 40% 이상을 차지한다면 이를 DDoS 공격이 발생한 것으로 간주한다.

(그림 7)에서 샘플링 비율을 1/1000 이상으로 하면 샘플링 기법에 관계없이 성능이 92% 이상 나옴을 확인 할 수 있다. 또한 샘플링 기법들은 각각의 구간별로 우열을 가릴 수 있겠지만 그 차이는 두드러지지 않음을 알 수 있다. 따라서 샘플링 기법의 종류는 트래픽 비율 분석법을 채택한 DDoS 탐지성능에 크게 영향을 주지 않는다는 사실을 확인할 수 있다.



(그림 7) 각 샘플링 기법의 샘플링 비율 별 탐지 성능

5. 결 론

본 논문에서는 패킷 샘플링 기법을 트래픽 비율 분석을 채택한 DDoS 공격 탐지 기술에 적용하여 트래픽의 특성 변화와 탐지 성능의 차이를 분석해 보았다. 결론적으로 연구에 사용한 트래픽 비율 분석법은 샘플링 기법을 적용시켰을 때 1/100의 샘플링 비율 이상일 경우 원본 데이터와 DDoS 탐지의 성능차이가 1~2% 이내였고, 1/1000의 샘플링 비율의 경우에도 효용성을 충분히 입증하고 있다. 또한 트래픽 비율 분석 DDoS 탐지에 있어 각각의 샘플링 기법들 사이의 성능 차이가 크게 없음을 확인할 수 있었다. 그러나 1/1000 이하의 샘플링 비율에서는 성능이 90% 이하로 저하됨에 따라 DDoS 탐지 기능의 역할을 제대로 수행하지 못한다.

향후 DDoS 공격에 의한 트래픽 변화를 좀더 다양한 환경과 측정기기를 통해 수집하고 이를 패킷 샘플링에 적용하여 실험한 결과를 통해, 패킷 샘플링 기법이 트래픽 비율 분석법뿐만 아니라 다른 DDoS 공격 탐지 기법에도 그 효용성이 있음을 확인할 계획이다.

참 고 문 헌

[1] Cheolho Lee, Sanguk Noh, Kyunghee Choi and Gihyun Jung, "Characterizing DDoS Attacks with Traffic Rate Analysis." In Proceedings of the International Conference e-Society, Vol.1, pp.81-88, Lisbon, Portugal, June, 2003.

[2] Cristian Estan and George Varghese, "New Directions in Traffic Measurement and Accounting," ACM SIGCOMM Internet Measurement Workshop 2001, San Francisco, CA, Nov., 2001.

[3] InMon Corp., "Using sFlow and InMon Traffic Server for Intrusion Detection and other Security Applications," 2001, see : http://www.sflow.org/SamplingforSecurity.pdf.

[4] Jelena Mirkovic, Janice Martin and Peter Reiher. "A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms." Computer Science Department, University of California, Los Angeles, Technical Report No.020018, 2002.

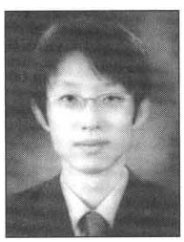
[5] Joseph Reves and Sonia Panchen, "Traffic Monitoring with Packet-Based Sampling for Defense against Security Threats," Passive & Active Measurement Workshop, Colorado, USA, Mar., 2002.

[6] Kimberly C. Claffy, George C. Polyzox, and Hans-Werner Braun, "Application of Sampling Methodologies to Network Traffic Characterization," Computer Communication Review, Vol.23, No.4, pp.194-203, Oct., 1993, appeared in Proceedings ACM SIGCOMM '93, San Fracisco, CA, pp. 13-17, Sep., 1993.

[7] Nick Duffield, Carsten Lund and Mikkel Thorup, "Properties and Prediction of Flow Statistics from Sampled Packet Streams," ACM SIGCOMM Internet Measurement Workshop 2002, Marseille, France, Nov., 2002.

[8] Symantec Security Response TFN2K, see : http://security response.symantec.com/avcenter/venc/data/tfn2k.html.

강 길 수

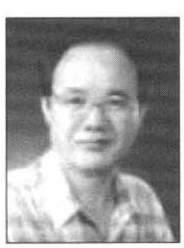


e-mail : monsterdarlf@hotmail.com
2002년 아주대학교 정보 및 컴퓨터공학부 (학사)
2004년 아주대학교 정보통신전문대학원 정보통신공학과(공학석사)
2004년~현재 텔코웨어 기술부문
관심분야 : 분산 시스템, 망 관리, 실시간 시스템

이 준 희

e-mail : rynex@hanmail.net
2002년 아주대학교 정보 및 컴퓨터공학부 (학사)
2004년 아주대학교 정보통신전문대학원 정보통신공학과(공학석사)
2004년~현재 알폰스테크(주) 소프트웨어 개발부
관심분야 : 망 관리, 임베디드 시스템, VoIP

최 경 희

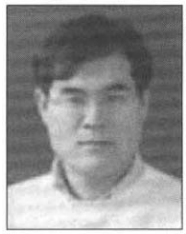


e-mail : khchoi@madang.ajou.ac.kr
1976년 서울대학교 수학교육과(학사)
1979년 프랑스 그랑데꼴 Enseeiht 대학 (석사)
1982년 프랑스 Paul Sabatier 대학 정보공학부(박사)
1982년~현재 아주대학교 정보통신전문대학원 교수
관심분야 : 운영 체제, 분산시스템, 실시간 및 멀티미디어시스템 등

정 기 현

e-mail : khchung@madang.ajou.ac.kr
1984년 서강대학교 전자공학과(학사)
1988년 미국 Ilinois 주립대 EECS(석사)
1990년 미국 Purdue 대학 전기전자공학부 (박사)
1991~1992년 현대반도체 연구소
1993년~현재 아주대학교 전자공학부 교수
관심분야 : 컴퓨터구조, VLSI 설계, 멀티미디어 및 실시간 시스템 등

심 재 홍



e-mail : jhshim@chosun.ac.kr
1987년 서울대학교 전산과학과(학사)
1989년 아주대학교 컴퓨터공학과(석사)
2001년 아주대학교 컴퓨터공학과(박사)
1989년~1994년 서울시스템(주) 공학연구소
2001년~2001년 아주대학교 정보통신전문원 BK21 전임연구원
2001년~현재 조선대학교 인터넷소프트웨어공학부 전임강사
관심분야 : 운영 체제, 분산시스템, 실시간 및 멀티미디어시스템