네트워크 트래픽 특성 분석을 통한 스캐닝 웜 탐지 기법

(Scanning Worm Detection Algorithm Using Network Traffic Analysis)

강 신 헌 † 김 재 현 ††
(Shin-Hun Kang) (Jae-Hyun Kim)

요 약 스캐닝 웜은 자기 스스로 복제가 가능하며 네트워크를 통해서 짧은 시간 안에 아주 넓은 범위에 걸쳐 전파되므로 네트워크의 부하를 증가시켜 심각한 네트워크 혼잡현상을 일으킨다. 따라서 실시간으로 스캐닝 웜을 탐지하기 위해 많은 연구가 진행되고 있으나 대부분의 연구가 패킷 헤더 정보를 이용하는 방법에 중점을 두고 있으며, 이 방법은 네트워크의 모든 패킷을 검사해야 하므로 비효율적이며 탐지시간이 오래 걸린다는 단점이 있다. 따라서 본 논문에서는 네트워크 트래픽량, 트래픽량의 미분값, 트래픽량의 평균 미분값, 트래픽량의 평균 미분값과 평균 트래픽량의 곱에 대한 variance를 통해 스캐닝 웜을 탐지하는 기법을 제안한다. 실제 네트워크에서 측정한 정상 트래픽과 시뮬레이터로 생성한 웜 트래픽에 대해 성능을 분석한 결과, 기존의 탐지기법으로는 탐지되지 않는 코드레드와 슬래머를 제안한 탐지기법으로 탐지할 수 있었다. 또한 탐지속도를 측정한 결과 웜 발생초기에 모두 탐지가 되었는데, 슬래머는 발생 후 4초만에 탐지되었으며, 코드레드와 위티는 발생한지 11초만에 탐지되었다.

키워드 : 스캐닝 웜, 트래픽 특성, 탐지 기법, 네트워크 보안

Abstract Scanning worm increases network traffic load and result in severe network congestion because it is a self-replicating worm and send copies of itself to a number of hosts through the Internet. So an early detection system which can automatically detect scanning worms is needed to protect network from those attacks. Although many studies are conducted to detect scanning worms, most of them are focusing on the method using packet header information. The method using packet header information has long detection delay since it must examine the header information of all packets entering or leaving the network. Therefore we propose an algorithm to detect scanning worms using network traffic characteristics such as variance of traffic volume, differentiated traffic volume, mean of differentiated traffic volume, and product of mean traffic volume and mean of differentiated traffic volume. We verified the proposed algorithm by analyzing the normal traffic captured in the real network and the worm traffic generated by simulator. The proposed algorithm can detect CodeRed and Slammer which are not detected by existing algorithm. In addition, all worms were detected in early stage: Slammer was detected in 4 seconds and CodeRed and Witty were detected in 11 seconds.

Key words : Scanning worm, Traffic characteristics, Detection algorithm, Network security

· 본 연구는 지식경제부 및 정보통신연구진흥원의 IT신성장동력핵심기술개발사업[2008-F-015-01, 서비스 가용성을 위한 이동성 관리 기술 연구]과 대학IT연구센터 지원사업의 일환으로 수행하였음

† 학생회원 : 아주대학교 전자공학과
cnonyk@ajou.ac.kr
†† 정 회 원 : 아주대학교 전자공학과 교수
jkim@ajou.ac.kr

논문접수 : 2007년 12월 10일
심사완료 : 2008년 8월 8일

정보과학회논문지: 정보통신 제35권 제6호(2008.12)

1. 서 론

최근 고속 광대역 네트워크를 통하여 멀티미디어 서비스를 비롯한 다양한 서비스가 가능하게 된 반면에, 바이러스, 웜, DoS(Denial of Service) 공격, 스파이웨어, 애드웨어 등 네트워크를 통한 악의적인 공격 또한 과거에 비해 많이 증가하고 있다[1,2]. 네트워크를 통한 여러 가지 공격 중 가장 심각한 피해를 초래하는 것은 웜이다. 웜이란 사용자의 개입 없이 네트워크를 통해 스스로 자신의 복사본 또는 변형체를 퍼뜨려 취약점이 있는 서비스를 공격하는 악성 코드를 말한다. 웜은 감염된 시스템의 파일을 지우거나 변형시켜 중요한 정보를 훼손시키거나 중요한 정보를 이메일을 통해 유출시키는 등의 피해를 준다. 특히 감염된 시스템을 원격에서 제어하여 스팸 메일을 보내는 데 사용하거나 DDoS(Distributed Denial of Service) 공격에 사용할 수도 있다. 또한 자동으로 대량 복제가 가능하여 시스템의 프로세서나 메모리 등의 자원뿐만 아니라 네트워크 링크의 대역폭을 소비하여 다른 정상적인 서비스를 불가능하게 하는 등 피해를 줄 수 있다.

웜은 이동식 디스크, 파일 공유 시스템, 이메일, 메신저 등 여러 가지 경로를 통해 전파된다. 기존의 웜은 사용자가 감염된 파일을 실행시키거나, 이메일의 첨부파일을 실행시키거나, 감염된 매체를 시스템에 연결하는 등 사용자의 개입에 의해 전파되었으므로 전파속도에 한계가 있었고 피해 범위가 넓지 않았다. 이에 반해 스캐닝 웜은 임의의 IP 주소를 가진 시스템에 계속적으로 자신의 복사본을 전파하여 짧은 시간 안에 네트워크를 통해 아주 넓은 범위에 걸쳐서 피해를 준다.

스캐닝 웜에 의한 피해사례를 살펴보면, 2001년 7월에 발생하여 14시간 만에 약 359,000 대의 컴퓨터를 감염시켰던 코드레드[3], 2003년 1월에 발생하여 10분만에 약 75,000 대의 컴퓨터를 감염시켰던 슬래머[4], 2003년 8월에 발생하여 수 시간 만에 500,000 대의 컴퓨터를 감염시켰던 블래스터[5], 그리고 2004년 3월에 발생하여 45분만에 12,000 대의 컴퓨터를 감염시켰던 위티[6] 등이 있다. 이와 같이 스캐닝 웜은 사용자가 미처 대응하기 전에 짧은 시간 안에 네트워크에 전파되므로 자동으로 스캐닝 웜을 탐지하고 대응할 수 있는 시스템이 필요하다.

스캐닝 웜의 탐지를 위해 많은 연구[7-13]가 진행되고 있지만 대부분의 연구가 패킷 헤더 정보를 이용하는 방법에 중점을 두고 있으며 스캐닝 웜의 일반적인 트래픽 특성을 이용한 탐지에 관한 연구는 미비한 실정이다. 패킷 헤더 정보를 이용하는 방법은 트래픽 특성을 이용한 탐지 방법에 비하여 더 정확한 탐지가 가능하지만 탐지를 위해서 모든 패킷을 다 검사해야 하므로 탐지에 소요되는 시간이 길다. 반면에 트래픽 특성을 이용한 탐지 방법[12,13]은 패킷 헤더 정보를 이용하는 방법에 비하여 정확도가 떨어지지만 모든 패킷을 다 검사할 필요가 없으므로 빠른 시간에 효율적인 탐지가 가능하다. 따라서 본 논문에서는 스캐닝 웜의 패킷 헤더 정보에 중점을 둔 기존의 탐지 방법대신 트래픽 특성을 이용하여 탐지하는 방법을 제안한다. 그리고 실제 네트워크를 측정한 트래픽과 시뮬레이터로 발생시킨 웜 트래픽을 통해 제안하는 탐지기법과 기존의 탐지기법의 성능을 비교 분석하고 타당성을 검증한다.

2. 관련 연구

스캐닝 웜을 탐지하기 위한 기존 연구는 검사대상에 따라 크게 세 종류로 나눌 수 있다. 첫째는 패킷 내부의 내용을 모두 검사하여 기존에 알려진 웜들과 같은 패턴이 있는지 비교하는 방법[7]이다. 이 방법은 기존에 알려진 웜들만 탐지가 가능하고, 변종 웜이나 새로운 웜은 탐지가 불가능하다. 또한 기존 웜의 패턴 데이타를 모두 저장하고 있어야 하며 모든 패킷을 이 데이타와 비교해야 하므로 비효율적이고 탐지시간이 오래 걸린다.

둘째는 패킷 헤더에 있는 IP주소, 포트번호 등의 정보를 검사하는 방법이다. 일반적으로 스캐닝 웜은 가능한 한 많은 호스트를 감염시키기 위해 수신 IP 주소가 랜덤하며, 특정 서비스의 취약점을 공격하기 때문에 수신 포트 번호가 고정되어 있다. 따라서 발신 IP 주소, 발신 포트 번호, 수신 IP 주소, 수신 포트 번호 및 프로토콜 등 패킷 헤더 정보를 이용하여 스캐닝 웜을 탐지할 수 있다. 관련 연구로는 패킷의 발신 IP 주소, 수신 IP 주소 및 수신 포트 번호를 3차원 그래프로 나타내 특정한 형태를 띠게 되면 스캐닝 공격이라고 판단하는 방법[8]이 있으며, IP 주소, 포트 번호에 대한 엔트로피를 측정하여 수신 IP 주소가 랜덤하게 분포되어 있거나 특정 포트에 트래픽이 집중되어 있는 것을 탐지하는 방법[9]이 있다. 그밖에 IP 주소, 포트 번호가 같은 패킷들을 플로우 단위로 묶어 플로우 헤더 정보와 트래픽 패턴 정보를 생성하여 공격 유형별로 정의된 기준에 따라 공격을 탐지하는 방법[10]과 각 네트워크 경계마다 설치된 스캔 모니터에서 수집한 데이타를 종합하여 칼만 필터를 통해 웜의 감염율을 추정하는 방법[11] 등이 있다. 이 방법은 IP 주소가 랜덤하고 포트번호가 일정하다는 스캐닝 웜의 특성을 이용하기 때문에 변종 웜이나 새로운 웜의 탐지가 가능하다. 하지만 이 방법 역시 모든 패킷의 헤더를 다 검사해야 되기 때문에 시스템이 복잡하고 탐지시간이 오래 걸린다.

셋째는 개별 패킷이 아닌 집합 트래픽의 특성을 분석

하여 탐지하는 방법이다. 이 방법은 스캐닝 웜의 일반적인 트래픽 특성을 이용하기 때문에 변종 웜이나 새로운 웜의 탐지가 가능하고, 패킷의 크기와 개수만 분석하므로 시스템이 간단하고 짧은 시간안에 효율적인 탐지가 가능하다. 관련 연구로는 단위 시간당 패킷의 수와 비트 수 정보만을 이용해 트래픽량에 대한 패킷 수의 비율을 구하여 웜을 탐지하는 방법[12]과 트래픽량의 variance, VMR, correlation coefficient를 분석하여 탐지하는 방법[13]이 제안된 바 있다. [13]에 따르면 variance에 의한 탐지가 가장 성능이 좋기 때문에 본 논문에서는 트래픽량 뿐만 아니라 트래픽량의 미분값과 평균값 등 다양한 데이타에 대한 variance를 분석하여 효율적이면서도 높은 정확도로 스캐닝 웜을 탐지할 수 있는 기법을 제안한다.

3. 트래픽 특성 분석을 통한 스캐닝 웜 탐지기법

3.1 스캐닝 웜 트래픽 특성

코드레드는 최초의 스캐닝 웜으로서 패킷 크기가 크고 TCP를 이용하여 웜 패킷을 전송하기 때문에 세션을 열기 위해 한번 SYN 패킷을 보낸 후 SYN/ACK 패킷을 받을 때까지 또는 타임아웃 시간이 지날 때까지 기다려야 하므로 전파속도가 느리다. 그림 1은 두 곳의 Class B 네트워크에서 측정된 코드레드의 시간당 스캔 시도 횟수[14]를 나타내고 있다. 코드레드 발생 후 약 14:00부터 스캔 시도 횟수가 급격히 증가하여 약 5시간만에 최대값에 도달하였다. 슬래머는 현재까지 가장 빠른 전파속도를 가진 스캐닝 웜으로서 패킷 크기가 작고 UDP를 이용하여 웜 패킷을 전송하기 때문에 ACK 패킷을 기다리지 않고 계속해서 웜 패킷을 전송할 수 있다. 그림 2는 슬래머가 발생했을 때 위스콘신 대학에서 측정한 슬래머의 패킷 수[4]를 나타내고 있는데 발생 후 약 3분만에 최대 속도에 도달한 것을 볼 수 있다. 이와 같이 스캐닝 웜 트래픽은 아주 짧은 시간 안에 급격하게 증가하는 특성을 가지고 있다.

스캐닝 웜의 특성을 분석하여 표 1에 정리하였다. 코

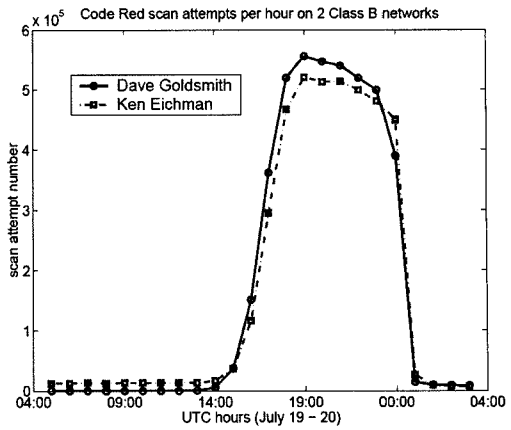

그림 1 코드레드의 시간당 스캔 시도 횟수

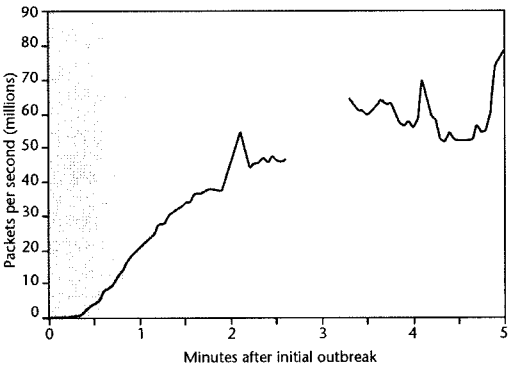

그림 2 슬래머의 초당 패킷 수

드레드Ⅱ는 코드레드Ⅰ이 먼저 발생하여 전파되고 있는 기간에 발생했기 때문에 정확한 스캔속도를 알 수 없어 표에서 생략하였다. 스캐닝 웜은 수신 포트 번호 또는 발신 포트 번호가 고정되어 있으며 수신 IP 주소는 랜덤하거나 부분적으로 랜덤한 특성을 보인다. 스캐닝 웜은 코드레드 웜, 블래스터 웜과 같이 패킷 크기가 비교적 크고 스캔 속도가 느린 것과 슬래머 웜처럼 패킷 크

표 1 스캐닝 웜의 특성

구 분	수신 IP 주소	수신 포트 번호	패킷 크기 (byte)	스캔속도 (packets/sec)
Code RedIv2	Random	80	3569	11
Code RedII	12.5%: random 50%: in the same class B 37.5%: in the same class C	80	3818	.
Slammer	Random	1434	404	4000
Blaster	40%: in the same class C 60%: random	135	6176	15
Witty	Random	Random (src port #:4000)	796~1307	357

기가 작고 스캔 속도가 빠른 것이 있다. 따라서 패킷 크기와 스캔 속도를 개별적으로 고려해서는 스캐닝 웜의 특성을 파악할 수가 없다. 하지만 스캐닝 웜의 패킷 크기와 스캔 속도를 곱한 트래픽량은 항상 크다는 것을 알 수 있다. 따라서 본 논문에서는 트래픽량의 분석을 통해 스캐닝 웜의 트래픽 특성을 이용한 탐지 방법을 제안한다.

제안하는 탐지 시스템의 구조는 그림 3과 같으며, A와 같은 로컬 네트워크의 경계라우터나 B와 같이 몇 개의 네트워크 트래픽이 합쳐지는 지점 또는 C와 같은 백본라우터에서 단위시간당 트래픽량을 모니터링한다. 본 논문에서는 트래픽량의 variance 뿐만 아니라 트래픽량의 미분값에 대한 variance, 트래픽량의 평균미분값의 variance, 트래픽량의 평균미분값에 평균 트래픽량을 곱한 값의 variance를 고려하여 가장 성능이 좋은 방법을 선택한다.

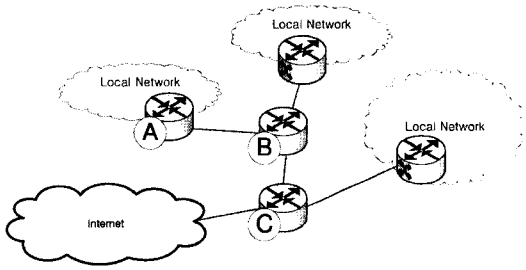

그림 3 스캐닝 웜 탐지 시스템 구조

3.2 트래픽량의 variance

트래픽의 특성을 분석하기 위한 가장 간단한 방법으로 먼저 트래픽량의 variance를 이용하는 방법을 고려하였다. Variance는 구간 내의 각 샘플값과 평균값의 차이를 나타내는데 스캐닝 웜이 발생하면 네트워크의 트래픽량이 급증하므로 정상 트래픽에 비해 단위시간당 트래픽량의 variance가 증가한다. t초에 측정한 variance, VAR(t)를 식 (1)과 같이 정의한다.

$$\mathrm{VAR}(t)=\begin{cases}\left[\sum_{k=t-W+1}^{t/t_s}\left[X(k)-\mu_X(t)\right]^2\right]\Big/W, & if\ t\geq W\\ 0, & elsewhere\end{cases} \quad (1)$$

여기서 $X(t)$는 t-t_s 초에서 t초까지의 시간 동안 링크를 통과한 트래픽의 비트 수, W는 variance계산구간을 정하는 윈도우 크기, $\mu_X(t)$는 t-W+1초에서 t초까지 W초간 $X(t)$ 값의 평균이다. 만약 VAR(t)가 이전 W초 동안의 평균값보다 α배 이상 크면 Alarm_VAR(t)를 1로 설정하고, 나머지 경우는 0으로 설정한다. 매초마다 이전 W_d초 동안의 Alarm_VAR(t) 값을 합하여 $\tau \times W_d$ 보다 크면 스캐닝 웜이 발생했다고 판단한다. $X(t)$의 정의에서 알 수 있듯이 트래픽량의 측정 시간단위는 t_s초이며, t_s의 크기에 따라 탐지 속도 및 정확도가 많이 달라지므로 링크 속도 및 네트워크 구성을 고려하여 정확한 설정이 필요하다.

3.3 트래픽량의 미분값의 variance

변화량을 분석하는데 일반적으로 많이 사용하는 미분을 통하여 급격하게 증가하는 트래픽량을 탐지한다. 트래픽량이 급증하는 구간에서는 트래픽량의 미분값이 커지기 때문에 탐지가 가능하다. t초에 측정한 트래픽량의 미분값 $d(t)$는 다음과 같이 정의된다.

$$d(t)=X(t)-X(t-1) \quad (2)$$

트래픽량의 미분값 $d(t)$에 대한 variance, DVAR(t)는 다음과 같이 정의된다.

$$\mathrm{DVAR}(t)=\begin{cases}\left[\sum_{k=t-W+1}^{t/t_s}\left[d(k)-\mu_d(t)\right]^2\right]\Big/W, & if\ t\geq W\\ 0, & elsewhere\end{cases} \quad (3)$$

여기서 $\mu_d(t)$는 t-W+1초에서 t초까지 W초간 $d(t)$ 값의 평균이다. 만약 DVAR(t)가 이전 W초 동안의 평균값보다 β배 이상 크면 Alarm_DVAR(t)를 1로 설정하고, 나머지 경우는 0으로 설정한다. 매초마다 이전 W_d초 동안의 Alarm_DVAR(t) 값을 합하여 $\tau \times W_d$ 보다 크면 스캐닝 웜이 발생했다고 판단한다.

3.4 트래픽량의 평균 미분값의 variance

네트워크 트래픽량이 급격히 증가하게 되면 트래픽량의 미분값이 전반적으로 증가하지만, 불규칙한 특성이 있기 때문에 미분값의 평균을 취하여 variance를 구하는 방법을 제안한다. t초에 측정한 미분값의 평균 $\mu_d(t)$에 대한 variance, DMVAR(t)는 다음과 같이 정의된다.

$$\mathrm{DMVAR}(t)=\begin{cases}\left[\sum_{k=t-W+1}^{t/t_s}\left[\mu_d(k)-\mu_{\mu_d}(t)\right]^2\right]\Big/W, & if\ t\geq 2W\\ 0, & elsewhere\end{cases} \quad (4)$$

여기서 $\mu_{\mu_d}(t)$는 t W+1초에서 t초까지 W초간 $\mu_d(t)$ 값의 평균이다. 만약 DMVAR(t)가 이전 W초 동안의 평균값보다 γ배 이상 크면 Alarm_DMVAR(t)를 1로 설정하고, 나머지 경우는 0으로 설정한다. 매초마다 이전 W_d초 동안의 Alarm_DMVAR(t) 값을 합하여 $\tau \times W_d$ 보다 크면 스캐닝 웜이 발생했다고 판단한다.

3.5 트래픽량의 평균미분값과 평균트래픽량의 곱의 variance

네트워크 트래픽량이 급격히 증가하게 되면 트래픽량

미분의 평균값이 증가한다. 이때 트래픽량도 증가하고 미분의 평균값도 증가하므로 트래픽량과 미분의 평균값을 곱하면 증가량이 더 커져서 더 쉽게 탐지할 수 있다. 따라서 트래픽량 미분의 평균값에 평균 트래픽량을 곱한 값의 variance를 구하는 방법을 제안한다. t초에 측정한 미분값의 평균 $\mu_d(t)$와 트래픽량의 평균 $\mu_X(t)$의 곱에 대한 variance, DMMVAR(t)는 다음과 같이 정의된다.

$$\mathrm{DMMVAR}(t)=\begin{cases}\left[\sum_{k=t-W+1}^{\lfloor t/t_s \rfloor}\left[\mu_d(k)\mu_X(k)-\mu_{\mu_d\mu_X}(t)\right]^2\right]\Big/W, & \text{if } t\geq 2W\\ 0, & \text{elsewhere}\end{cases} \quad (5)$$

여기서 $\mu_{\mu_d\mu_X}(t)$는 t-W+1초에서 t초까지 W초간 $\mu_d(t)$와 $\mu_X(t)$의 곱의 평균이다. 만약 DMMVAR(t)가 이전 W초 동안의 평균값보다 δ배 이상 크면 Alarm_DMMVAR(t)를 1로 설정하고, 나머지 경우는 0으로 설정한다. 매초마다 이전 W_d초 동안의 Alarm_DMMVAR(t) 값을 합하여 $\tau \times W_d$보다 크면 스캐닝 웜이 발생했다고 판단한다.

4. 성능 분석

정상 트래픽만 발생하는 경우와 스캐닝 웜이 같이 발생하는 경우의 네트워크 트래픽에 대하여 제안한 탐지기법과 기존의 탐지기법을 적용하여 성능을 비교분석한다. 정상 트래픽은 WIDE Project의 MAWI Working Group[15]에 의해 2007년 12월 23일에 대륙간 링크에서 실제로 측정한 트래픽을 사용하였으며, 스캐닝 웜 트래픽은 OPNET을 이용한 시뮬레이터로 발생시켰다. 스캐닝 웜 트래픽으로는 코드레드, 슬래머와 위티를 선택했는데 그 이유는 스캐닝 웜 중 가장 패킷 크기가 크고 스캔 속도가 낮은 코드레드와 가장 패킷 크기가 작고 스캔 속도가 높은 슬래머, 그리고 중간정도의 패킷 크기와 스캔 속도를 가진 위티를 탐지할 수 있다면 다른 웜의 탐지 역시 가능할 것이기 때문이다.

본 논문에서 고려한 참조망 구조는 그림 4와 같다. 스캐닝 웜 트래픽을 발생시키는 로컬 네트워크 열다섯개와 백본 라우터가 IP백본망을 통해 연결되어 있으며 IP백본망과 백본 라우터를 연결하는 링크에서 트래픽량을 측정하였다. 각 네트워크는 시나리오에 따라 3가지의 스캐닝 웜 트래픽을 400초부터 순차적으로 발생시켜900초까지 지속시킨다. 스캐닝 웜 트래픽은 [3-6]을 참조하여 설정하였으며, 트래픽 별 자세한 설정은 표 2에 정리하였다.

4.1 성능 분석 결과

성능 분석은 정상 트래픽만 발생하는 경우, 정상 트래픽과 코드레드가 발생하는 경우, 정상 트래픽과 슬래머가 발생하는 경우, 정상 트래픽과 위티가 발생하는 경우의 4가지 시나리오에 대하여 수행하였다.

그림 5는 백본 링크에서 측정한 트래픽량(초당 비트

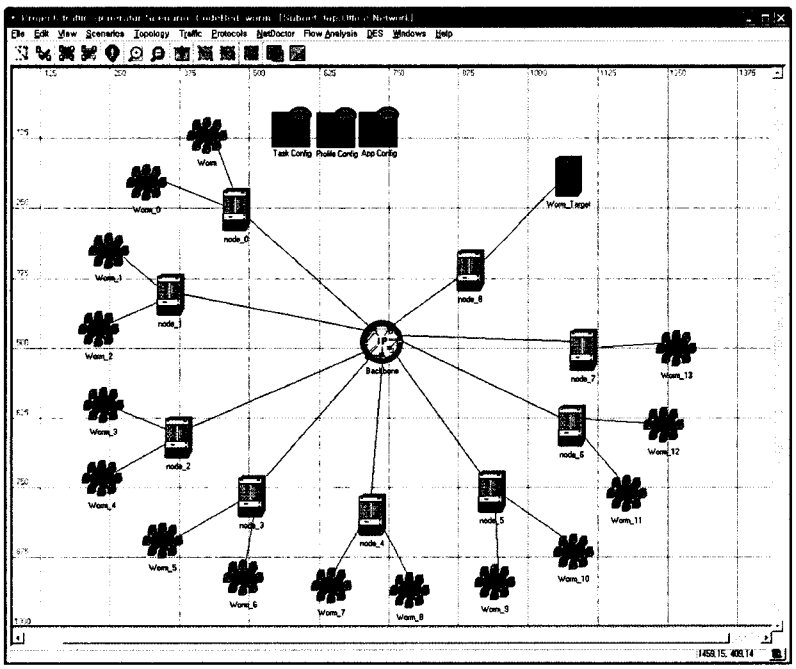

그림 4 참조망 구조

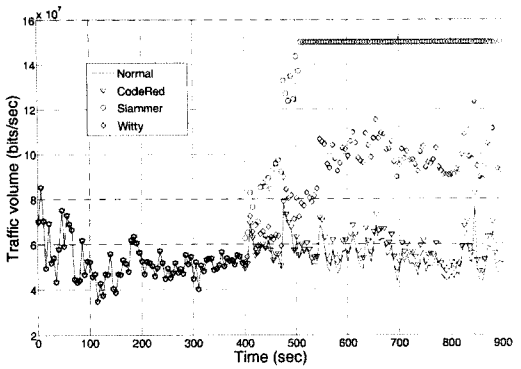

그림 5 네트워크 트래픽량

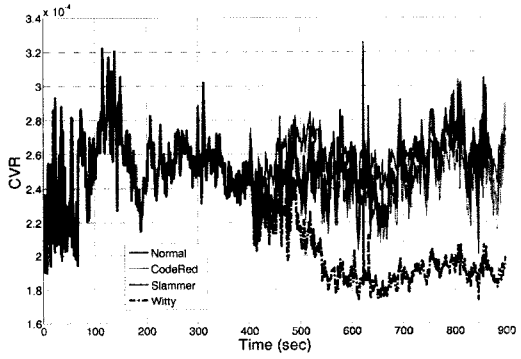

그림 6 네트워크 트래픽의 CVR

수)을 시간에 따라 나타낸 것이다. 스캐닝 웜 트래픽이 400초부터 554초까지 순차적으로 발생하여 900초까지 지속된다. 400초에 스캐닝 웜이 발생하기 시작하면 트래픽량이 증가하는데 코드레드에 의한 트래픽량의 증가는 미미한 반면 위티와 슬래머에 의해서는 트래픽량이 크게 증가함을 볼 수 있다. 측정 대상 링크가 150Mbps의 용량을 가지고 있으므로 최대값을 150Mbps로 제한하고 그 이상 발생하는 트래픽은 무시하였다.

그림 6은 기존에 제안된 트래픽 특성을 이용한 탐지 방법[12]에 따라 네트워크 트래픽의 패킷 개수와 트래픽

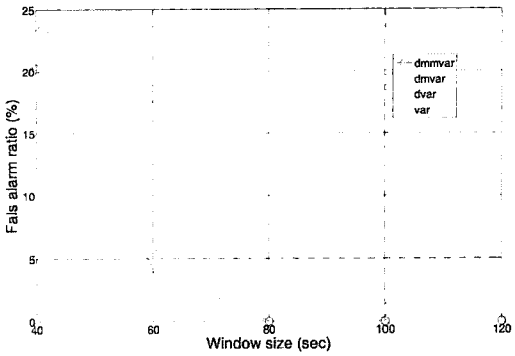

그림 7 윈도우 크기에 따른 false alarm비율

표 3 스캐닝 웜 탐지시간

구분	VAR	DVAR	DMVAR	DMMVAR
코드레드	11초	19 초	76 초	75 초
슬래머	4 초	16 초	14 초	13 초
위티	11 초	19 초	75 초	45 초

량의 비율인 CVR(packet count to volume ratio)을 측정한 결과이다. 위티가 발생한 경우는 약 540초부터 정상 트래픽의 CVR보다 낮은 값을 보여 구분이 가능하다. 하지만 코드레드와 슬래머가 발생하는 경우의 CVR은 정상 트래픽의 CVR과 구별되지 않아서 탐지가 불가능함을 알 수 있다.

다음으로 3장에서 제안한 탐지 기법을 이용한 탐지 결과를 살펴보겠다. 탐지 파라미터의 설정값을 바꿔가며 여러 번 반복실험하여 정상 트래픽 구간에서 웜이 발생한 것으로 잘못 판단하는 경우(false positive)가 발생하지 않도록 다음과 같이 설정하였다. $W=100, W_d=10, t_s=1,$ $\alpha=\beta=\gamma=\delta=1.2, \tau=0.1$. 그림 7은 윈도우 크기에 따른 false positive확률의 변화를 나타내고 있다. 윈도우 크기가 작을 때는 VAR을 이용한 탐지가 다른 방법에 비해 false positive확률이 더 높지만 100초 이상이 되면 false positive가 더 이상 발생하지 않는 것을 볼 수 있

표 2 웜 트래픽 세부 설정

Attribute		Value
코드 레드	Inter-Request Time (sec)	Exponential (0.091)
	Request Packet Size (bytes)	Constant(3569)
	Start Time (sec)	Constant (400, 430, 454, 473, 488, 500, 510, 519, 527, 534, 540, 545, 549, 552, 554)
슬래머	Inter Request Time (sec)	Exponential (0.00025)
	Request Packet Size (bytes)	Constant (404)
	Start Time (sec)	Constant (400, 430, 454, 473, 488, 500, 510, 519, 527, 534, 540, 545, 549, 552, 554)
위티	Inter-Request Time (sec)	Exponential (0.0028)
	Request Packet Size (bytes)	Uniform_int (796,1307)
	Start Time (sec)	Constant (400, 430, 454, 473, 488, 500, 510, 519, 527, 534, 540, 545, 549, 552, 554)

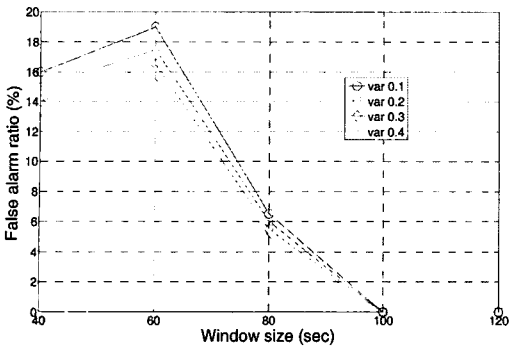

그림 8 윈도우 크기와 τ에 따른 false alarm비율 (VAR)

표 4 τ에 따른 탐지시간 (VAR)

구분	τ = 0.1	τ = 0.2	τ = 0.3	τ = 0.4
코드레드	11 초	12 초	13 초	14 초
슬래머	4 초	5 초	6 초	7 초
위티	11 초	12 초	13 초	14 초

다. 표 3은 윈도우 크기가 100초일 때, 웜이 발생한 후 탐지될 때까지 소요되는 시간을 나타낸다. 제안한 탐지기법 중 VAR을 이용한 탐지가 가장 탐지속도가 빨랐으며, 계산도 가장 간단하기 때문에 가장 효율적인 탐지방법이라고 할 수 있다. 나머지 탐지방법들은 모두 미분값을 이용하는데, 미분값은 측정시점의 측정값과 이전 측정값의 차이므로 음의 값과 양의 값이 모두 발생하므로 평균값이 작아지고 불규칙해지기 때문에 VAR에 비해 탐지속도가 느려지는 것이라 생각된다.

그림 8은 위와 같은 조건에서 VAR을 이용하여 탐지했을 때 윈도우 크기와 τ의 값에 따른 false alarm비율의 변화를 나타낸다. τ가 클수록 false alarm비율이 낮아지며, 윈도우 크기가 100초 이상이 되면 τ의 값에 상관없이 false alarm이 일어나지 않는 것을 볼 수 있다. 표 4는 VAR을 이용하여 탐지하였을 때 τ의 값에 따른 탐지시간을 나타내고 있는데, τ가 클수록 탐지시간이 길어지는 것을 알 수 있다. 이와 같이 τ의 값에 따라 정확도와 탐지시간 사이에 trade-off가 있으며, 윈도우 크기가 100초 이상일 때는 τ를 0.1로 설정했을 때 정확도와 탐지속도를 가장 높일 수 있었다.

5. 결 론

본 논문에서는 스캐닝 웜의 일반적인 트래픽 특성을 분석하여, 네트워크 트래픽량의 VAR, DVAR, DMVAR, DMMVAR을 정의하고, 트래픽 측정 시스템 구조와 탐지기법을 제안하였다.

제안한 스캐닝 웜 탐지 기법은 네트워크 트래픽량만을 분석하므로 네트워크의 모든 패킷 헤더를 검사하는 방식에 비해 계산이 간단하고 많은 메모리가 요구되지 않기 때문에 더 효율적인 탐지가 가능하여 탐지 속도를 높일 수 있다. 또한 패킷의 개수와 트래픽량의 비율만 분석하는 기존의 탐지기법에 비해서 일정 구간동안 트래픽량의 통계적 특성을 분석하므로 탐지 정확도를 높일 수 있다.

컴퓨터 시뮬레이션으로 생성한 웜 트래픽과 실제 네트워크를 측정한 트래픽을 통해 제안한 탐지기법의 성능을 분석한 결과 기존의 탐지기법으로 탐지되지 않는 웜까지 모두 발생 초기에 탐지되었다.

참 고 문 헌

[1] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, "A Taxonomy of Computer Worms," in *Proc. ACM workshop on rapid malcode*, 2003, pp. 11-18.

[2] D.M. Kienzie and M.C. Elder, "Recent Worms: A Survey and Trends," in *Proc. ACM workshop on rapid malcode*, 2003, pp. 1-10.

[3] D. Moore, C. Shannon, and J.Brown, "Code-Red: a case study on the spread and victims of an Internet worm," in *Proc. Second Internet Measurement Workshop*, 2002, pp. 273-284.

[4] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "Inside the Slammer Worm," *IEEE Security & Privacy*, pp. 33-39, Jul./Aug. 2003.

[5] "W32.Blaster.Worm," [Online]. Available: http://securityresponse.symantec.com/avcenter/venc/data/w32.blaster.worm.html

[6] C. Shannon and D. Moore, "The Spread of the Witty Worm," *IEEE Security & Privacy*, pp. 46-50, Jul./Aug. 2004.

[7] K. Wang, G. Cretu, and S. Stolfo, "Anomalous Payload-Based Worm Detection and Signature Generation," *Lecture Notes in Computer Science*, 3858, pp. 227-246, 2006.

[8] H. Kim, I. Kang, and S. Bahk, "Real-Time Visualization of Network Attacks on High-Speed Links," *IEEE Network*, pp. 30-39, Sep./Oct. 2004.

[9] S. Noh, C. Lee, K. Ryu, K. Choi, and G. Jung, "Detecting Worm Propagation Using Traffic Concentration Analysis and Inductive Learning," *Lecture Notes in Computer Science, 3177(1)*, pp. 402-408, 2004.

[10] M. Kim, H. Kang, S. Hong, S. Chung, and W. Hong, "A Flow-based Method for Abnormal Network Traffic Detection," in *Proc. IEEE/IFIP NOMS*, 2004, pp. 599-612.

[11] C. Zou, W. Gong, D. Towsley, and L. Gao, "The Monitoring and Early Detection of Internet Worms," in *Proc. 10th ACM conference on Computer and communication security*, 2003, pp. 190-199.

[12] B. Roh and S. Yoo, "A Novel Detection Methodology of Network Attack Symptoms at Aggregate Traffic Level on Highspeed Internet Backbone Links," *Lecture Notes in Computer Science, 3124*, pp. 1226-1235, Aug. 2004.
[13] 김재현, 강신헌, "네트워크 트래픽 특성을 이용한 스캐닝 웜 탐지기법", *한국정보보호학회논문지, 제 17권, 제 1호*, pp. 57-66, 2007년 2월.
[14] C. C. Zou, W. Gong, and D. Towsley, "Code Red Worm Propagation Modeling and Analysis," in *Proc. 9th ACM Conference on Computer and Communications Security*, 2002, pp. 138-147.
[15] "MAWI Working Group Traffic Archive," [Online]. Available: http://tracer.csl.sony.co.jp/mawi/

강 신 헌

2006년 아주대학교 전자공학부 학사. 2008년 아주대학교 전자공학과 석사. 2008년~현재 아주대학교 전자공학과 박사과정. 관심분야는 네트워크 보안, 무선 네트워크 등

김 재 현

1991년 한양대학교 전산과 학사. 1993년 한양대학교 전산과 석사. 1996년 한양대학교 전산과 박사. 1997년~1998년 미국 UCLA 전기전자과 박사후 연수. 1998년~2003년 Bell Labs, Performance Modeling and QoS Management Group, 연구원. 2003년~현재 아주대학교 전자공학부 부교수. 관심분야는 무선 인터넷 QoS, MAC 프로토콜, IEEE 802.11/15/16/20 등