네트워크 트래픽 분석을 통한 웜 탐지방법에 관한 연구

노대종*, 노태열*, 박승섭**
*부경대학교 대학원
**부경대학교 전자컴퓨터정보통신공학부 교수
E-mail : woory@mail1.pknu.ac.kr, parkss@pknu.ac.kr

A Study on Worm Detection Algorithm Using Network Traffic Analysis

DaeJong Noh*, TeaYol Noh*, SeungSeob Park**
*Dept. of Computer Science Education, Pukyong National University
**Division of Electronic, Computer and Telecommunication Engineering, Pukyong National University

요 약

인터넷 사용의 급증과 함께 Code-Red나 Slammer와 같은 웜이 급격히 확산 되고 있으며 네트워크를 통해 스스로 전파되면서 네트워크 자원을 고갈시킴으로써 문제가 더욱 심각해지고 있다. 이에 따라 웜을 탐지하기 위한 많은 방법들이 제시되었다. 본 논문에서는 DoS 탐지를 위해 고안 된 트래픽 비율 분석법을 이용하여 정상 네트워크와 웜이 발생 시키는 스캐닝 관련 행위에 대한 패킷 비율을 비교하였다. 이 방법을 통해 네트워크 내에서 웜에 감염된 호스트를 찾아내고 오탐지율을 최소화하는 방법과 웜 전파 행위를 탐지해 내는 방법에 대해서 제안한다. 또한 실제 네트워크에서 수집된 트래픽으로부터 웜의 특성을 분석해 본 결과 최근 웜들의 전파방식을 분석 할 수 있었다.

1. 서론

인터넷의 급속한 발달과 사용자 수의 계속적인 증가로 인해 다양한 네트워크 서비스가 만들어져 제공되고 있다. 이러한 인터넷의 성장과 더불어 다양한 네트워크 보안 공격에 대한 위협들도 증가 하였다. 다양한 서비스들에 내재된 많은 보안 취약점들이 공개되면서 이러한 취약점들을 해커들이 접근 권한을 얻기 위해 이용하고 있다.

특히 웜의 경우 네트워크상에 커다란 피해를 유발할 가능성을 가지고 있다. 1988년 Morris 웜[2]의 출현 이래로 웜에 의한 보안 공격들은 계속 증가되어 왔고, 바이러스와는 다른 형태의 성질을 갖는 웜에 대응하기 위한 방법들이 계속해서 개발 되어져 왔다. 웜의 특성은 사용자의 개입이 없이 자체적으로 빠른 속도로 전파되면서 로컬 시스템이나 파일을 파괴하는 것뿐만 아니라 네트워크 자원을 고갈시킴으로써 관리자로 하여금 제때 대응하기 어렵게 만든다.

그렇기 때문에 웜을 탐지하는 시스템에 대한 필요성이 강조되고, 이런 시스템은 웜의 전파 행위를 자동적으로 대응하는데 사용될 수 있다. 특히 네트워크 관리자에게 있어서 네트워크 모니터링을 통해 웜 트래픽을 분류하고, 네트워크 내부의 감염된 호스트를 찾아내어 관리하는 것은 웜의 지속적인 감염을 저지하고 전파 행위를 미연에 방지하는데 필요한 수단이다.[3],[4].

웜의 초기 전파 단계에서 웜을 탐지하기 위해 많은 웜 탐지 구조와 시스템들이 개발되었다. 그리고 웜 행위의 탐지에 있어서 오탐지를 줄이는 것은 중요한 요소이다. 이를 위해 웜과 비슷한 행위를 하는 정상적인 응용 프로그램들이 발생시키는 행위에 대해서도 고려할 필요가 있다. 웜의 트래픽을 분석하는 것은 웜을 효과적으로 탐지하고, 웜의 트래픽을 분류하는데 근거를 제시하여 준다. 본 논문에서는 네트워크에서 수집된 웜 트래픽의 특성을 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 웜의 대표적인 특성인 스캔 행위와 트래픽 비율 분석법에 대해서 알아보고, 3장에서는 트래픽 비율 분석법을 이용하여 포트별 특징을 이용한 웜 탐지 방법에 대해서 기술하며, 4장에서는 인터넷 트래픽의 상당수를 차지하는 포트별 트래픽 특징을 상세하게 분석 평가한다. 마지막으로 5장에서는 결론 및 향후 연구를 제시한다.

2. 관련연구

2.1웜의 스캐닝 행위

웜은 자신을 전파시키기 위해 많은 연결 요청 패킷을 뿌림으로써 인터넷 상에 취약점을 가진 호스트를 찾아 감염시킨다. 대표적으로 Code-Red 와 Slammer 의 경우 랜

덤(Random) 스캐닝을 사용하였고, Blaster는 순차적(sequential) 스캐닝 방식을 사용한 웜이다. 최근 들어 지능적인 웜도 발생하고 있으나 타겟 주소의 범위가 전체 인터넷 크기인 2^{32} 에서 10^9 으로 줄어들게 되는 것이고, 그 행위 자체는 랜덤 스캐닝과 같다[1].

그러나 사용되지 않는 IP 주소들에 대한 정보를 모두 파악하는 것은 매우 힘든 일일 뿐만 아니라, 서버처럼 항상 가동되는 호스트들만 존재하는 것도 아니고, IP 주소라는 것은 상당히 유동적이기 때문에 완전한 전파 리스트를 확보하고 웜을 전파시키는 것은 생각하기 어렵다. 웜은 타겟 호스트를 감염시키기 위해서 취약점이 노출된 서비스를 노리게 된다. 따라서 해당 서비스와 관련된 포트 번호는 웜에 대한 중요한 판단 기준이 된다.

2.2 트래픽 비율 분석법

트래픽 비율 분석법이란 전체 트래픽에서 특정한 형태를 가진 트래픽의 비율(rate)을 이용하여 트래픽을 분석하고 DoS 공격의 탐지를 수행하는 기법이다. 트래픽 비율 분석법은 TCP 플래그 비율(TCP flag rate)과 프로토콜 비율(protocol rate)로 구분된다[6]. 다음 수식은 TCP 플래그 비율을 정의하고 있다.

$$R_{td}[K\hat{i}|o] = \frac{\sum flag(K)}{\sum TCP\ packets} \quad (inbound/outbound) \ldots(1)$$

TCP 플래그 비율은 TCP 패킷만을 대상으로 하며, 특정한 TCP 플래그를 가진 패킷의 개수를 전체 TCP 패킷의 개수로 나눠서 구할 수 있다. 여기서 td는 트래픽 비율을 측정한 시간 간격을 의미하고, k는 SYN, FIN, RST, ACK, PSH, URG, NULL등의 TCP 플래그를 나타낸다.

다음 수식은 프로토콜 비율을 정의하고 있다. 프로토콜 비율은 특정한 프로토콜(4계층 프로토콜:TCP, UDP, 또는 ICMP)을 갖는 패킷의 개수를 전체 IP 패킷의 개수로 나눠서 구할 수 있다[6].

$$R_{td}[(TCP|UDP|ICMP)i|o] = \frac{\sum (TCP|UDP|ICMP)\ packets}{\sum IP\ packets} \quad \ldots\ldots(2)$$

위의 두 수식은 DDoS 공격을 탐지하기 위해 제안되어졌으나 이와 비슷한 공격유형을 가지는 웜을 탐지하고 특성을 분석하기에는 충분하지 않다.

3. 트래픽 비율 분석법을 이용한 웜 탐지방법

3.1 Outbound 트래픽을 이용한 내부 호스트 감염 탐지

웜의 전파 행위가 빠르게 이루어진다면 감염된 호스트가 뿌리는 패킷의 수가 많게 된다. 또한 각각의 패킷은 서로 다른 타겟 호스트로 향하기 때문에 이 부분에서 트래픽의 비정상성이 두드러지고, 이는 웜 탐지에 충분히 활용될 수 있다. 감염된 호스트의 포트별 스캐닝 패킷량을 평상시의 패킷량과 비교하여 그 양이 현저히 증가하였고 시간의 경과에 상관없이 동일한 목적지 포트로 스캐닝 패킷을 거의 일정한 속도로 계속해서 보내는 행위가 관찰될 때에는 웜에 의한 현상이 아닌가를 의심해 볼 필요가 있다. 다음은 트래픽 비율 분석법을 이용하여 수식으로 정의한 것이다[6].

$$R_{td}[number] = \frac{\sum port[number]\ packets}{\sum TCP\ SYN\ packets} \ (Outbound) \ldots(3)$$

여기서 td는 트래픽 비율을 측정하기 위한 시간 간격을 의미하고, number는 각 포트번호를 의미한다. 이 방법을 이용한 분석법으로 웜이 사용하는 이미 잘 알려진 포트번호는 물론 새로 생성된 웜이 사용하는 포트번호도 알아낼 수 있을 것이다. 또한, Outbound된 TCP SYN 패킷에 대하여 응답한 패킷의 비율을 분석함으로써 내부네트워크의 감염 여부를 짐작할 수 있다. 왜냐하면 감염 호스트의 패킷에 대한 외부네트워크의 정책이나 라우터의 설정에 의해서 응답패킷들이 폐기되었다고 추측해 볼 수 있기 때문이다. 다음은 트래픽 비율분석법을 이용하여 수식으로 정의한 것이다[6].

$$R_{td} = \frac{\sum Inbound(ACK+RST+ICMP)\ packets}{\sum Outbound\ packets\ about\ Worm} \quad \ldots..(4)$$

위의 두 방법을 이용하여 각 네트워크에 알맞은 임계치를 설정하여 트래픽을 측정하면 내부네트워크의 웜의 감염 여부를 감지할 수 있을 것이다.

3.2 Inbound 트래픽을 이용한 외부 공격 웜 탐지

레코드의 해당 포트번호에 대해서 내부로 들어오는 트래픽을 모니터링 하여 연결 요청을 받는 서로 다른 호스트 수가 비정상적으로 크다면 이는 웜 트래픽으로 볼 수 있다. 정상적인 경우라면 살아있는 호스트만으로 연결 요청이 이루어지기 때문에 이 수치가 일정 수준을 넘어서지 않는다. 이 단계를 통해 웜 트래픽과 정상 트래픽을 보다 확실히 구별해 냄으로써 오탐지를 줄일 수 있다. 포트 별 서로 다른 목적지 주소 수에 대한 정보는 Inbound 트래픽 모니터링을 통해 생성된 포트 사용 레코드를 이용한다. 만약 해당 포트에 대한 Inbound 트래픽에서 스캐닝 패킷이 내부 네트워크의 상당한 수의 IP 주소에 도달한다면, 이는 웜 트래픽으로 간주한다. 이를 정리한 알고리즘은 다음과 같다.

$$R_{td}[Pn_a] = \frac{\sum IP(Pn)}{\sum TCP\ SYN\ packets} \quad (Inbound),$$

$$R_{td}[Pn_e] = \frac{\sum IP(Pn)}{\sum TCP\ SYN\ packets} \quad (Inbound),$$

$$If \quad R_{td}[Pn_a] > R_{td}[Pn_e] \quad Than \quad Worm \quad \ldots\ldots\ldots(5)$$

여기서 $R_{td}[Pn_a]$는 의심스러운 레코드로 감지된 트래픽의 각 포트번호별 IP주소수의 비율이며 $R_{td}[Pn_e]$는 평상시

의 각 포트번호별 IP주소수의 비율이다.

3.3 웜 탐지에 필요한 임계값

각 포트 번호별 특성에 맞는 각각의 임계값이 필요하다. 여기서 임계값은 평상시의 전체 패킷량의 평균값과 각 포트별 평균 비율을 이용하여 다음과 같이 구할 수 있다[7].

$$T_{port} = \mu_{Tport} + k\sigma_{Tport} \quad (k = 1,2,3,4,\ldots) \quad \ldots\ldots(6)$$

이 때, k 값이 증가하면 신뢰도가 높아지는데 본 논문에서는 3으로 설정하여 신뢰도를 99.5%로 가정한다. 또한 각 포트별 트래픽의 평균 패킷량 μ_{Tport}는 다음과 같이 구할 수 있다.

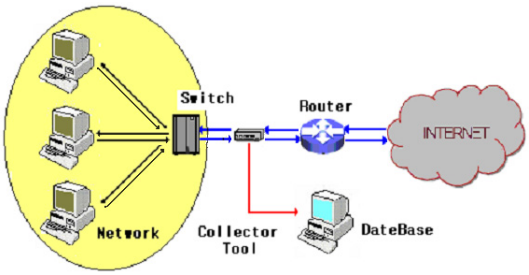
$$\mu_{Tport} = (1-\alpha)\mu_{old} + \alpha\mu_{new} \quad (\text{단 } \alpha = 0.9) \ldots\ldots(7)$$

이 때, α는 0.9로 설정하여 트래픽의 평균에 유동성을 부여한다.

4. 웜 트래픽 특성 분석 및 평가

4.1 웜 트래픽 분석을 위한 레코더 생성 방법

웜의 트래픽을 수집하기 위하여 네트워크의 인터넷 접속점을 모니터링 하였다. (그림1)과 같이 수집 도구는 네트워크로 오가는 실시간 트래픽을 수집한다.



(그림1) 인터넷 트래픽 수집 방법

트래픽을 모니터링 할 때는 두 가지로 나누어 패킷 정보를 수집한다. 하나는 네트워크에서 인터넷으로 나가는 Outbound 트래픽을 모니터링 하는 것으로 네트워크 내부에서 웜에 의해 감염된 호스트들이 발생시키는 스캐닝 행위를 감시한다. 다른 하나는 인터넷에서 웜이 내부 네트워크를 감염시키기 위해 들어오는 Inbound 트래픽을 모니터링 한다. Inbound 트래픽에서는 포트 번호 별로 스캐닝 패킷이 내부 네트워크의 얼마나 많은 IP 주소에 도달하는지를 감시 한다. 이렇게 수집한 정보들을 이용하여 각각의 레코드를 생성한다.

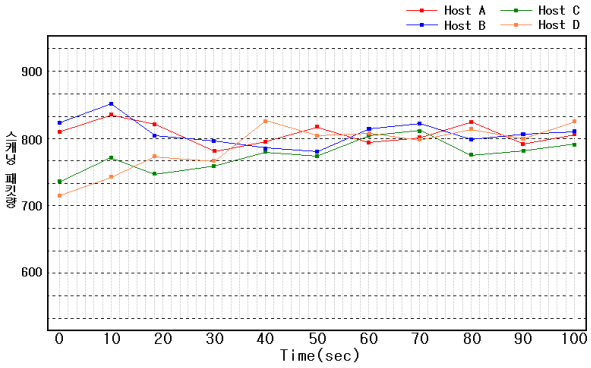
여기서 포트번호(port number)의 경우, TCP는 목적지 포트 번호만을 살펴보지만, UDP는 근원지와 목적지 포트 번호에 대해 각각 레코드를 생성한다. 실제로 근원지 포트 번호로 서비스를 구별하는 UDP 웜이 존재하였다[5]. 앞서 설명했듯이 TCP 는 SYN 패킷만을 수집하면서 목적지 포트 번호로만, UDP는 모든 패킷을 수집하면서 근원지와 목적지 포트 둘 다 레코드를 생성한다.

4.2 트래픽 비율 분석법을 이용한 웜 트래픽 분석

4.2.1 스캐닝 패킷량 비율

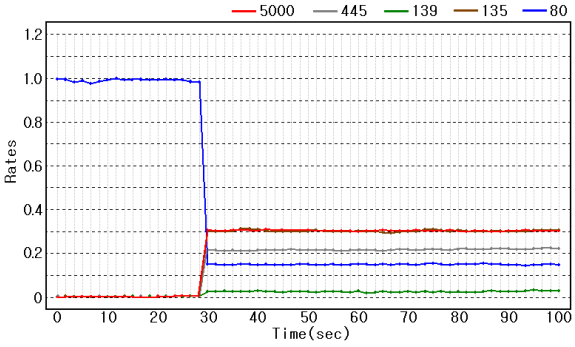
웜의 스캐닝 패킷량은 인터넷 웜이 얼마나 많은 트래픽을 발생시키는지, 또 얼마나 빠르게 전파를 시도하는지 파악할 수 있는 중요한 요소이다.

(그림2)에서 웜이 네트워크의 호스트들을 감염시켰을 때, 각 감염된 호스트가 네트워크 밖으로 스캐닝하는 속도는 거의 일정함을 알 수 있다. 다른 포트번호에 대해서도 비슷한 결과를 확인할 수 있다. 본 논문에서는 관찰된 감염 호스트의 Outbound 되는 TCP SYN 패킷 관련 트래픽 중 웜 관련 트래픽이 차지하는 비율은 각 포트별로 98% 이상을 차지했으며 이 비율은 시간의 경과에 상관없이 지속 되었다.



(그림2) 80번 포트를 이용하는 웜의 스캐닝 패킷량

(그림3)은 웜이 사용하는 각 포트 별로 감염된 호스트들의 스캐닝 패킷량을 비율로 나타낸 것이다. 한 호스트에서 동일한 목적지 포트로 스캐닝 패킷을 거의 일정한 양으로 계속해서 보낼 때, 이러한 트래픽에 대해서는 웜에 의한 행위가 아닌가를 의심해 볼 필요가 있다.



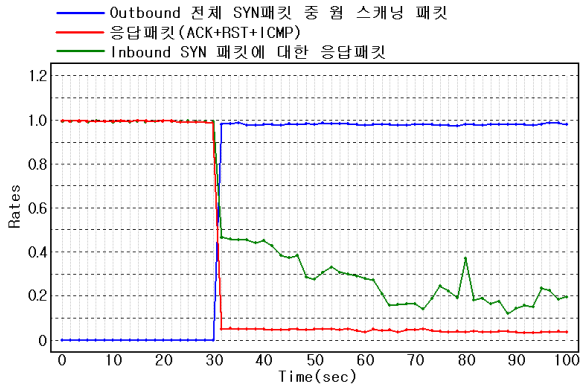
(그림3) Outbound 웜의 포트별 스캐닝 패킷량 비율

또한 스캐닝 트래픽의 목적지 주소 분포를 분석함으로써 실제로 웜이 어떤 식으로 타겟 호스트를 선정하는지 알 수 있다. 그리고 최근의 웜들이 타겟을 선정함에 있어서 어떤 방식을 가지고 있는지 파악할 수 있다.

4.2.2 응답 패킷 비율

웜 스캐닝에 대한 응답패킷을 수집하여 탐지에 사용되는 웜의 스캐닝 효율을 살펴보았다. (그림4)를 보면 Outbound 트래픽의 스캐닝에 대한 응답패킷이 극히 적

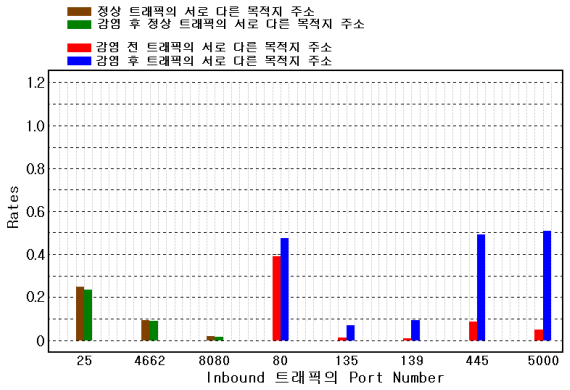
음을 볼 수 있다. 만약 응답패킷을 탐지 방법에 적용할 때에는 응답패킷들이 100% 도착하지 않을 수도 있다는 특성을 고려해야 할 것이다. 본 논문에서 관찰된 응답패킷의 비율은 시간당 웜이 내보낸 패킷량의 6%정도에 불과하였다.



(그림4) Outbound 웜 스캐닝패킷과 응답패킷의 비율

4.2.3 스캐닝 트래픽의 목적지 주소 분포 분석

웜은 하나의 주소를 시작으로 그 주소로부터 일정한 블록의 순차적인 주소에 대해서 스캐닝을 시도하기도 하고, 대부분의 웜은 랜덤 스캐닝 방식을 사용한다. 이런 방식의 스캐닝은 할당되지 않은 IP 대역으로 패킷이 전송될 수밖에 없다. (그림5)에서 나온 포트번호들은 앞 장의 탐지 알고리즘을 적용하였을 때 의심스러운 레코드로 선정된 것들이다. 여기서 TCP 25, 4662, 8080은 정상 트래픽들이다. 이 세 개의 포트번호는 의심스러운 레코드 내의 목적지 주소 분포 관찰로는 웜으로 간주되지 않지만, 정상 트래픽이라는 보다 확실한 근거를 얻기 위해 Inbound 트래픽에서 서로 다른 목적지 주소의 수를 관찰한다.



(그림5) Inbound 트래픽의 포트번호별 서로 다른 목적지 주소 수

웜이 활동하는 80, 445, 5000 포트와는 다르게 적은 수치를 기록하고 있음을 볼 수 있다. (그림5)은 이와 같은 특성이 웜 트래픽과 정상 트래픽을 분류하는 기준이 될 수 있음을 보여준다.

5. 결론 및 향후연구

기존의 트래픽 비율 분석은 정상적인 웹 트래픽과 다양한 DoS 공격이 발생 하였을 때의 웹 트래픽을 구분함에 있어서 명확하지가 않고, Dos 공격 시 나타나는 트래픽 특성과 응용 프로그램(P2P, 스트리밍 서비스 등) 사용 시 나타나는 트래픽 특성이 유사 할 수 있다. 이러한 DoS 공격 탐지를 위한 트래픽 비율 분석법으로는 웜을 탐지하는데 있어서 부족한 점이 많다.

본 논문에서 제안한 알고리즘은 웜의 스캐닝 특성을 이용하여 설계되었고, 네트워크 내부의 감염된 호스트를 알아낼 뿐만 아니라 정상 트래픽으로 인한 오탐지를 줄이는 방법도 고려하였다. 본 논문에서 제안한 웜 탐지를 위한 트래픽 비율 분석법을 통해서는 웜이 사용하는 이미 잘 알려진 포트 번호는 물론 새로 생성된 웜이 사용하는 포트 번호를 알아 낼 수 있으며 웜이 어떤 식으로 타겟 호스트를 선정하는지도 알 수 있다. 제안된 알고리즘을 통해 네트워크 관리와 웜의 특성을 분석하는데 있어서 보다 다양하고 구체적인 접근이 가능하다.

또한, 웜의 스캐닝 패킷량, 스캐닝 패킷의 목적지주소의 분포, 스캐닝 패킷이 차지하는 트래픽의 양과 그에 대한 응답 패킷들의 비율을 비교함으로써 웜의 탐지에 활용할 수 있는 트래픽 특성들을 알아보았다. 그러나 보다 다양한 관점에서의 웜 트래픽 분석이 수행되어야 하겠다. 그리고 이를 활용하여 웜 탐지 알고리즘을 향상시킬 수 있는 방안도 고려되어야 할 것이다.

참고문헌

[1] 홍성철, 조룡권, 주홍택, 홍원기, “An Internet Worm Detection Algorithm for Enterprise Networks”, KNOM Review, Vol.7, No.2, December 2004.

[2] Charles Schmidt, and Tom Darby, "The What, Why, and How of the 1988 Internet Worms,"
http:// www.snowplow.org/tom/worm/worm.html,July 2001.

[3] V. H. Berk, R. S. Gray, and G. Bakos, "Using Sensor Networks and Data Fusion for Early Detection of Active Worms," In Proceedings of the SPIE Aerosense, 2003, pp. 92-104.

[4] C. C. Zou, L. Gao, W. Gong, and D. Towsley, "Monitoring and Early Warning for Internet Worms," In Proceedings of 10th ACM Conference on Computer and Communications Security (CCS'03), October 2003, pp. 190-199.

[5] Colleen Shannon, and David Moore, "The Spread of the Witty Worm"
http://www.caida.org/analysis/security/witty/,Match 2004

[6] Cheolho Lee, Kyunghee Choi, Gihyun Jung, Sanguk Noh, "An Analysis of Network Traffic on DDoS Attacks against Web Servers", 정보처리학회논문지 C 제10-C권 제3호, June, 2003.

[7] Jongyeup Lee, Misun Yoon, Hoon Lee, "Monitoring and Investigation of DoS Attack", Changwon National Universoty, 2004