엔트로피 기반의 이상징후 탐지 시스템
(An Anomalous Event Detection System based on Information Theory)

한 찬 규 † 최 형 기 ††
(Chan-Kyu Han) (Hyoung-Kee Choi)

요 약 본 논문에서는 엔트로피에 기반한 이상징후 탐지 시스템을 제안한다. 엔트로피는 시스템의 무질서정도를 측정하는 지표로써, 이상징후 출현 시 네트워크의 엔트로피는 급증한다. 네트워크를 IP와 포트 번호를 기준으로 분류하여, 패킷별로 역학을 관찰하고 엔트로피를 각각 측정한다. 분산서비스거부공격이나 웜, 스캐닝 등의 네트워크 공격 출현 시 패킷 교환과정이 정상적일 때와는 다르므로 엔트로피를 통하여 기존기법 보다 높은 탐지율로 이상징후를 탐지할 수 있다. 본 논문에서는 다수의 웜과 서비스거부공격을 포함한 데이터 셋을 수집하여 제안기법을 검증하였다. 또한 지수평활법, Holt-winters 등의 시계열예측 기법과 주성분분석을 이용한 이상징후 탐지 기법과 정확도 측면에서 비교한다. 본 논문에서 제안한 기법으로 웜, 서비스거부공격 등의 이상징후 탐지에 있어 오탐지율을 낮출 수 있다.

키워드 : 엔트로피, 이상징후 탐지, 웜, 서비스거부공격, ROC, 네트워크 보안

Abstract We present a real-time monitoring system for detecting anomalous network events using the entropy. The entropy accounts for the effects of disorder in the system. When an abnormal factor arises to agitate the current system the entropy must show an abrupt change. In this paper we deliberately model the Internet to measure the entropy. Packets flowing between these two networks may incur to sustain the current value. In the proposed system we keep track of the value of entropy in time to pinpoint the sudden changes in the value. The time-series data of entropy are transformed into the two-dimensional domains to help visually inspect the activities on the network. We examine the system using network traffic traces containing notorious worms and DoS attacks on the testbed. Furthermore, we compare our proposed system of time series forecasting method, such as EWMA, holt-winters, and PCA in terms of sensitive. The result suggests that our approach be able to detect anomalies with the fairly high accuracy. Our contributions are two folds: (1) highly sensitive detection of anomalies and (2) visualization of network activities to alert anomalies.

Key words : Entropy, anomaly detection, worm, denial of service, ROC, network security

1. 서 론

인터넷을 통한 웜, 스캐닝, 서비스거부공격 등의 네트워크 공격은 네트워크 자원을 낭비하여 사용자에게 제공되는 서비스의 질과 보안의 강도를 하락시키는 위협적인 존재이다. 2000년에 발생한 Amazon, CNN, Yahoo, Ebay 같은 유명 사이트를 마비시킨 사례나, 2001년 Microsoft Domain Name System(DNS) 서버와 연결된 라우터가 분산서비스거부공격을 받아 라우터의 과부하로 인해 한동안 웹 접속서비스가 마비된 사건이 대표적인 사례로 볼 수 있다[1]. 따라서 이러한 네트워크 공격을 높은 정확도로 탐지하는 기법을 개발하는 것은 시급한 문제이다.

이상징후란 네트워크 공격의 일련의 과정 및 결과에서 추출할 수 있는 트래픽 특성을 지칭한다. 이상징후의 일례로는 높은 대역폭, 정상트래픽과는 다른 IP의 분포

† 학생회원 : 성균관대학교 휴대폰학과
hedwig@ece.skku.ac.kr
†† 정 회 원 : 성균관대학교 정보통신공학부 교수
hkchoi@ece.skku.ac.kr
(Corresponding author임)
논문접수 : 2008년 1월 29일
심사완료 : 2009년 2월 2일

정보과학회논문지: 정보통신 제36권 제3호(2009.6)

혹은 포트번호의 분포 등이 있다. 이상징후는 네트워크 공격을 탐지하는 데에 적합한 단서를 제공하며, 따라서 이러한 이상징후를 높은 정확도로 탐지하는 것은 네트워크 공격에 기민하게 대처하는 데에 일조할 수 있다[1]. 현재까지 이상징후를 효율적으로 탐지하기 위한 기법이 다수 제안되었지만 미탐율과 오탐율이 높거나, 시간복잡도가 높다는 단점이 있다.

이에 본 논문에서는 미탐율과 오탐율을 최소화한 효율적인 이상징후 탐지를 위해 패킷의 이동역학을 고려한 이상징후 탐지기법을 제안한다. 이상징후가 발생하게 되면 패킷의 이동역학이 정상적일 때와는 급변하므로, 패킷의 이동역학을 고려했을 때는 보다 높은 정확도로 이상징후를 탐지할 수 있다. 본 논문에서는 패킷의 이동역학을 관찰하기 위해 열역학의 엔트로피 개념을 차용한다.

엔트로피는 열역학에서 사용되는 개념으로 시스템 내의 무질서한 정도를 측정하는 수치이다. 시스템 내의 현재 상태와 다른 분포를 가진 요소가 유입 또는 유출되면 엔트로피는 급증한다. 이상징후가 포함된 트래픽과 정상 트래픽의 분포는 대역폭, IP, 포트번호, 패킷수 등에서 현저한 차이가 난다. 예를 들어 서비스거부공격은 공격대상 IP로 트래픽이 집중되며 대역폭 사용량이 높아지는 특성이 있다. 이러한 이상징후 발생 시에 엔트로피가 급증하므로 이상징후를 탐지할 수 있다.

본 논문에서는 네트워크를 모델링하고, 모델에서 엔트로피를 측정하여 이상징후를 탐지한다. 네트워크를 모델링 하는 방식은 IP와 포트번호로 분류한다. 먼저 IP를 이용하여 보호해야 할 내부 네트워크는 시스템으로 모델링하고, 외부 네트워크의 공격자는 주변환경으로 모델링한다. 포트번호를 이용한 분류로는 정상 트래픽 서비스 포트를 Interactive, Interactive bulk, Asynchronous bulk 세 개의 트래픽 서비스로 분류하여 시스템으로 모델링하고, User ephemeral 포트번호를 주변환경으로 모델링한다. 시스템과 주변환경 간에 교환하는 정보는 패킷으로 정의되며, 각 시스템과 주변환경 간 패킷 분포를 엔트로피를 통해 측정하여 이상징후를 탐지한다.

제안기법의 검증을 위해 다수의 웜과 서비스거부공격이 포함된 트래픽 데이터 셋을 수집하여 제안 기법을 적용하였다. 또한 정확도 및 오탐률 관점에서 타 알고리즘과의 비교 및 평가를 제시하였다. 비교평가 기준으로는 미탐율과 오탐율을 민감하게 측정할 수 있는 도구를 사용하였다. 비교평가 결과 서비스거부공격 및 분산서비스거부공격, 슬래머 웜이 포함된 트래픽 데이터 셋의 경우 보다 매우 우수한 정확도를 보였다.

시계열예측 기법[2], 주성분분석 기법[3]의 경우 분산서비스거부공격 등의 네트워크 공격은 탐지율이 높은 대신, 포트스캐닝 경우 높은 오탐지율을 보인다. 상대엔트로피 기법[4]의 경우 정상 트래픽에서 목적지 포트번호만이 엔트로피의 관측대상이기 때문에 포트스캐닝 등의 네트워크 공격만이 효율적으로 탐지할 수 있다.

본 논문에서는 먼저 2장에서 이상징후 탐지 기법을 연구한 기존의 관련연구들을 정리한다. 3장에서는 열역학개념을 네트워크에 적용하는 기법을 제안하며, 엔트로피를 통해 이상징후를 탐지하는 원리를 설명한다. 제안기법의 검증을 위해 사용하는 이상징후 트래픽 데이터 셋을 4장에서 설명하고, 제안 기법을 트래픽 데이터 셋에 적용하여 이상징후를 탐지하는 방법 및 결과를 5장에 기술한다. 6장에서는 주성분분석 및 시계열예측 기법과 정확도 면에서 비교평가한 결과를 서술하고, 7장에서 본 논문의 결론을 맺는다.

2. 관련 연구

본 장에서는 기존에 제안된 이상징후 탐지기법과 그에 따른 연구결과를 소개한다. 기존 연구는 일반적으로 트래픽 볼륨에 대한 관찰을 바탕으로, 트래픽 볼륨에 특징적인 변화가 나타날 시에 이를 이상징후라 탐지한다. 트래픽 볼륨이란 단위시간당 패킷수, 단위시간 당 패킷의 크기의 변화량 등을 시간순으로 관측한 데이터를 의미한다. 대표적인 연구로 시계열예측, 주성분분석, 신호분석을 이용한 이상징후 탐지 기법을 소개한다.

Jake D.Brutlag은 트래픽 볼륨에 시계열예측 기법을 적용하여 이상징후를 탐지했다[2]. 시계열예측 기법이란 관측된 시계열 데이터의 과거의 추세를 모델링하여 미래의 값을 예측하는 기법이다. 논문에서는 지수평활법과 Holt-Winters 시계열예측 기법을 사용한다. 지수평활법은 현재 시점의 관측치와 예측치로, 미래 시점의 값을 예측한다. 현재 시점의 관측치에 가중치를 조절하여, 어느 정도로 관측치를 보정할지 결정하게 된다. Holt-Winters는 시계열 데이터가 기준치, 선형추세 및 계절추세로 구성되어 있다고 하고 각 추세마다 가중치를 두어서 보다 정밀하게 예측치를 계산하다. [2]의 시계열예측 기법은 단순한 연산에 기반을 두어 시간복잡도가 낮은 반면 오탐률과 미탐률이 높은 특성이 있다.

Anukool Lakhina와 그의 팀은 주성분분석을 이용하여 이상징후를 탐지할 수 있는 기법을 제안하였다[3]. 주성분분석(Principal Components Analysis:PCA)은 데이터의 특성은 유지하면서 데이터의 차원을 낮추는 기법이다. 다수개의 링크에 나타난 이상징후의 차원을 낮추어 간단한 연산만으로도 이상징후를 추출할 수 있다. 데이터의 차원을 낮추기 위해서는 데이터의 특성을 반영하는 주성분만을 원 데이터에서 추출하여 사용한다. 이 기법의 경우 링크마다 행렬연산과정을 거쳐야 하므

로, 링크의 개수가 n개라면 시간복잡도가 $O(n^2)$에 근접한다.

Yin Zhang와 그의 팀은 다수의 이상징후 탐지 기법을 정리하고, 신호분석을 이용한 기법을 추가 제안하여 이들을 비교 분석하였다[5]. 신호분석을 이용한 이상징후 탐지 기법으로는 푸리에변환[6]과 웨이블릿 분석기법[6]이 있다. 일반적으로 이상징후는 짧은 시간에 높은 대역폭을 차지하므로 고주파로 나타난다. 웨이블릿은 푸리에변환이 주기에 대한 정보만을 제공하고, 해당 주기가 나타난 시간에 대한 정보는 제공하지 못한다는 단점을 보완하여 제안되었다. 푸리에 변환이 고주파 요소만을 찾아내는데 비해 웨이블릿은 고주파 요소 즉 이상징후가 나타난 시간까지도 탐지할 수 있다.

트래픽 볼륨을 대상으로 하는 이상징후 탐지기법 외에 트래픽 특성을 기반으로 하는 이상징후 탐지 기법에 제시 되었다. 트래픽 특성은 패킷헤더에서 측정할 수 있는 특성(예. IP주소 및 포트번호)과 통신특성(예. 시간당 IP주소의 분포)으로 구분된다.

엔트로피는 이러한 트래픽 특성의 변화 정도를 측정할 수 있고, 이를 이용한 이상징후 탐지기법이 다수 제안되었다. [7]에서는 정보이론에서의 엔트로피를 와 시스템 정보를 이용하여 네트워크 공격을 탐지하는 기법을 제시하였다. [8]에서는 인터넷 웜은 IP주소와 포트번호의 엔트로피에 영향을 끼치며, 엔트로피의 특성을 이용하여 웜 뿐 아니라 보다 다양한 공격을 탐지할 수 있다고 주장하였다. Anukool Lakhina는 [9]에서 목적지/출발지 IP주소와 포트번호의 4가지 튜플의 엔트로피를 계산한 뒤, PCA를 적용하는 이상징후 탐지 기법을 제안하였다. Kuai Xu 등은 트래픽 구분을 위해 엔트로피 값을 이용하였다[10]. Yu Gu와 그의 팀은 수집한 정상 트래픽의 엔트로피와 테스트 트래픽의 엔트로피의 차가 일정기준 이상 벗어나면 이상징후라 탐지한다[4]. 일반적으로 트래픽에서 엔트로피 적용은 IP 주소와 포트번호의 목적지-출발지 쌍인 4-튜플에 한정되고 있다. [11]에서는 패킷헤더 특성(4-튜플) 뿐만 아니라 통신특성에 따른 엔트로피도 측정되어야 한다고 주장하였다. 기존 기법과 다르게 본 논문에서의 엔트로피 측정 기법은 패킷 헤더 특성 뿐만 아니라 통신 특성도 고려할 수 있다.

3. 엔트로피에 기반한 이상징후 탐지

네트워크 호스트 간 패킷교환과, 열역학 이론에서 시스템과 주변환경이 물질을 주고 받는 과정은 유사하다. 이에 본 논문에서는 열역학 이론을 응용하여 이상징후 탐지에 적합한 알고리즘을 제시하고자 한다. 열역학의 시스템과 주변환경의 개념을 도입하여 네트워크를 구분하고, 시스템과 주변환경 간에 주고 받는 물질을 패킷으로 규정한다. 시스템과 주변환경의 물질분포에 급격한 변화가 나타나면 엔트로피가 증가한다는 열역학이론에 기반하여 이상징후를 탐지하고자 한다.

3.1 열역학 이론

열역학 이론은 에너지와 열 등의 물질의 교환에 다루는 학문이다[12]. 열역학 이론에서는 물질의 교환을 다루기 위해 관측대상을 **시스템**과 **주변환경**으로 분류한다. 시스템은 관측대상이고 주변환경과 에너지, 데이터 등의 물질을 교환한다. 물질교환에 따라서 시스템의 물질의 분포는 변화하는데, 열역학 이론에서는 이러한 변화를 **엔트로피**로 측정한다[12]. 시스템의 한계를 넘거나 과거의 분포와 현저하게 다른 분포를 야기하는 물질이 유입 또는 유출되는 특정 순간에 시스템의 엔트로피는 급증한다. 예를 들어 종이를 서서히 양쪽에서 잡아당길 때, 종이의 한계를 넘는 순간 종이가 찢어진다. 종이가 찢어지는 그 순간은 종이의 엔트로피가 급증하는 순간이다. 컴퓨터 호스트 간에 패킷을 교환하는 과정은 시스템과 주변환경이 물질을 교환하는 과정과 유사하다. 시스템과 주변환경은 각 컴퓨터 호스트에 대응하고, 시스템과 주변환경이 교환하는 에너지, 데이터 등의 물질은 네트워크에서의 패킷이라 볼 수 있다. 또한 열역학에서는 엔트로피를 통해 시스템의 물질분포의 이상유무를 판단할 수 있다. 즉 이상징후가 나타난 시기에는 엔트로피가 급격히 변하므로 엔트로피 측정을 통해 이상징후 탐지가 가능하다. 이에 본 논문에서는 열역학 이론을 적용하여 이상징후를 탐지하려 한다. 열역학 이론을 적용하기 위해 우선 대상을 시스템과 주변환경으로 구분하고, 시스템과 주변환경과의 물질의 입출입을 엔트로피를 통해 관측한다.

본 논문의 아이디어는 열역학 이론에서의 엔트로피 관찰에서 기인한다. 하지만 열역학에서의 엔트로피 측정은 물질의 분자 특성을 대상으로 하므로, 보다 범용적인 정보이론에서의 엔트로피 측정기법을 따르기로 한다.

3.2 시스템과 주변환경의 구분

열역학 이론을 적용하기 위해 대상을 시스템과 주변환경으로 구분하여야 한다. 시스템과 주변환경의 구분기준으로는 IP와 포트번호를 사용한다.

IP를 구분기준으로 할 때는 그림 1에서 보듯이 시스템은 내부 네트워크로 정하고 주변환경은 외부 네트워크로 정한다. IP를 기준으로 내부, 외부 네트워크를 각각 시스템과 주변환경으로 정한 이유는 네트워크의 활동주체인 호스트를 구분할 수 있는 요인이 IP이기 때문이다. IP를 기준으로 내부, 외부 네트워크로 구분함으로써 이상징후의 근원지 즉 공격자의 위치를 파악할 수 있다. 네트워크에서 시스템과 주변환경 간 물질의 입출입은 패킷을 통해 이루어진다. 시스템과 주변환경 간의

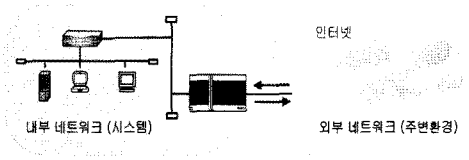


그림 1 IP를 기준으로 한 시스템/주변환경의 구분

패킷교환에 따라 시스템이 가지는 패킷의 양 또는 개수가 변화한다. 변화의 정도를 시스템의 엔트로피로 관찰함으로써 시스템의 정상여부를 탐지할 수 있다.

단순히 발생한 시점을 탐지하는 것보다 플로우 별 이상징후를 탐지해야만 이상징후에 대한 대처를 더욱 효과적으로 할 수 있다. 예를 들어 특정 취약점을 공격하는 웜 출현 시 단순히 발생했다는 시점보다 취약점이 있는 서비스가 어떤 플로우에 속하는지, 어떤 서비스에 취약점이 있는지 파악할 필요가 있다. 이에 이상징후가 나타난 플로우를 파악 및 분석하면 현재 출현한 이상징후가 어떤 이상징후인지 정확하게 판단이 가능하다. 그러나 IP만을 이용하여 시스템과 주변환경을 구분할 시에는 플로우 별 이상징후 판단이 불가하다.

따라서 그림 2와 같이 포트번호를 구분기준으로 하여 시스템과 주변환경을 구분한다. 포트번호를 구분기준으로 할 때 시스템은 Interactive, Interactive bulk, Asynchronous 3개의 서비스로 정하고, User ephemeral 서비스를 주변환경으로 정한다. 포트번호를 구분기준으로 사용한 이유는, 네트워크 공격은 일반적으로 취약점이 있는 서비스를 대상으로 발생한다는 사실에 기반하여 서비스를 구분할 수 있는 요인이 포트번호이기 때문이다. 서비스들은 잘 알려진 포트번호로 정의되어 있어 이들은 시스템으로 구분하고, 나머지 User ephemeral 포트번호는 주변환경으로 구분한다. 플로우 별 이상징후를 탐지하기 위해서는 어플리케이션 서비스의 특성을 반영하고, 각 범주 별로 중복이 없이 구분 지을 수 있어야 한다. 상기 두 항목을 고려하는 RFC1633[23] 서비스 구분기준에 따라 Interactive, Interactive bulk, Asynchronous, User ephemeral로 분류하였다.

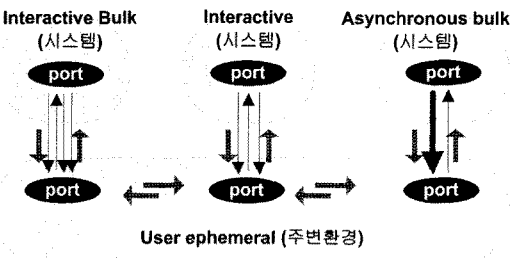


그림 2 포트번호를 기준으로 한 시스템/주변환경의 구분

Interactive는 텔넷, X-windows 등과 같이 세션이이루어진 상태에서 요청-응답 질의가 1:1로 상호 작용하는 트래픽을 말한다. 두 번째로 Interactive bulk는 Interactive 와 마찬가지로 상호작용이 이루어지나 서버로부터의 응답이 그룹 지어서 한꺼번에 온다는 특징이 있다. 대표적인 Interactive bulk로는 FTP, HTTP와 같은 어플리케이션이 존재한다. 마지막으로 Asynchronous 서비스는 요청(Request)에 상관없이 응답(Response) 여부가 빠른 시간 내에 도착하지 않아도 무방한 이메일 등이 이에 속한다.

본 논문에서는 각 시스템과 주변환경 간에 교환하는 물질을 패킷으로 정하였다. 패킷의 이동을 관찰할 수 있는 지표로써 호스트를 나타내는 IP와 포트번호가 대표적이다. IP는 패킷의 출발과 목적지의 위치를 규정하고, 포트번호는 서비스의 특성을 규정한다. 이상징후를 탐지하기 위해서는 이상징후의 출현시간, 지리적 위치 및 공격대상 서비스가 중요 요인이 될 수 있다. 이에 본 논문에서는 이상징후를 야기하는 공격자의 위치를 파악하기 위해 IP를 이용하여 시스템과 주변환경을 구분하였고, 이상징후의 공격대상 서비스를 파악하기 위해 포트번호를 이용하여 시스템과 주변환경을 구분하였다. 또한 이상징후의 출현 시간을 파악하기 위해 IP를 이용한 시스템과, 포트번호를 이용한 시스템에서 엔트로피를 주기적으로 측정하여 관찰한다. 다음 절에서는 시스템과 주변환경 간에 교환하는 패킷의 역학에서 엔트로피를 측정하는 방법을 기술한다.

3.3 엔트로피와 이상징후 탐지

본 논문에서 제안하는 패킷의 역학을 관찰하는 기법과 엔트로피를 측정하는 기법은 다음과 같다. 본 논문에서 제시하는 엔트로피 측정 기법은 1) 패킷헤더 특성(IP와 포트번호), 그리고 2) 패킷 역학을 이용한 통신 특성 모두를 측정할 수 있다. 이와 같은 방식은 보다 세밀하게 이상징후를 탐지 할 수 있다[11].

엔트로피는 일반적으로 목적지와 출발지 IP 주소, 포트번호에 이용되어 왔으며[4,8-10], 엔트로피는 각 튜플당 유지되었다. 본 논문에서는 3.2절에서 구분한 IP, 포트번호 구분법에 따라 패킷역학을 관찰한 뒤에, 해당 역학에서 수집된 정보에 엔트로피를 적용한다. 패킷 역학에서 관찰된 엔트로피는 슬라이딩 윈도우 방식을 적용하여 모니터링된다. 패킷역학을 관찰하는 방법은 다음과 같고, 관찰된 엔트로피에서 이상징후를 측정하는 방법은 5장에 설명한다.

각 시스템과 주변환경에는 토큰이 주어지며 시스템과 주변환경이 가진 토큰의 개수의 순서쌍을 **상태벡터**라 정의한다. 시스템에 i개, 주변환경에 j개의 토큰이 있다면 상태벡터는 (i,j)로 기록된다. 패킷의 이동에 따라 출

발지의 토큰의 개수를 1개 감소시키고, 목적지의 토큰의 개수를 1개 증가시킨 뒤 상태벡터를 기록한다. 예를 들어 시스템에서 주변환경으로 패킷이 1개 이동하였다면 $(i-1, j+1)$인 상태벡터가 기록된다. 만일 시스템에서 시스템으로 또는 주변환경에서 주변환경으로 패킷이 1개 이동한다면 다시 (i,j) 상태벡터가 기록된다. 정해진 시간구간마다 각 상태벡터의 히스토그램을 **상태수**로 정의한다. 만일 정해진 시간 구간에 (i,j) 상태벡터가 k번, (m,n) 상태벡터가 1번 기록되었다면 (i,j)와 (m,n)에 대한 상태수는 각각 k, 1이 된다.

토큰, 상태벡터, 상태수 개념의 이해를 돕기 위한 예는 다음과 같다. 시스템 X와 주변환경 Y가 있다고 가정하자. 시스템과 주변환경이 각 n개의 토큰을 가지고 따라서 상태벡터의 초기값은 (n,n)이다. 시스템과 주변환경 간에 $X \rightarrow Y$, $Y \rightarrow X$, $Y \rightarrow Y$의 패킷이동이 있었다고 가정하면, 상태벡터는 초기값을 포함하여 (n,n), $(n-1, n+1)$, (n,n), (n,n)으로 기록된다. 또한 상태수는 (n,n)은 3번, $(n-1,n+1)$은 1번이 된다.

주어진 시간 내에 상태벡터와 상태수를 구하고, 이를 통해 엔트로피를 측정하여 이상징후를 탐지하고자 한다. 엔트로피는 상태벡터의 개수와 상태수의 분포를 통해 다음 식 (1)과 같이 측정한다.

$$Entropy_t = -\sum_{i=1}^{m_t} p_i \log p_i \quad (1)$$

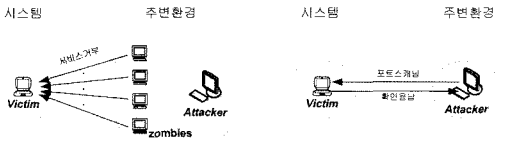
$$p_i = d_i / \sum_{i=1}^{m_t} d_i \quad (2)$$

$$Entropy = -\sum_{i=1}^{2} p_i \log p_i = -\left(3/4 \log 3/4 + 1/4 \log 1/4\right) \quad (3)$$

식 (1)은 단위시간 t번째 구간의 엔트로피인 $Entropy_t$를 측정하며 식 (2)의 d_i는 i번째 상태벡터의 상태수이며, m_t는 단위시간 t동안 기록된 상태벡터의 총 개수를 의미한다. 위의 예에서 (n,n) 상태벡터는 상태수 3번, $(n-1,n+1)$ 상태벡터는 상태수 1번이 된다. 상태벡터의 총 개수 m_t는 2가 되고 각각의 확률은 3/4, 1/4이 된다. 엔트로피는 식 (3)과 같이 계산할 수 있다.

본 논문에서 상태벡터의 개수와 상태수의 분포를 고려하여 엔트로피를 식 (1)과 같이 측정하는 이유는 다음과 같다.

그림 3에서 보듯이 공격대상 호스트을 시스템으로 구분하고, 좀비를 포함한 공격자를 주변환경으로 구분하였을 때 네트워크 공격, 즉 이상징후는 두 가지 범주로 분류가 가능하다. 그림 3(a)의 분산서비스거부공격 등의 네트워크 공격은 공격자로부터 공격대상 호스트로 패킷이 일방적으로 전송이 되는 형태로써 상태벡터의 통계를 구해보면 다양한 상태벡터가 고루 넓게 퍼져있는 형태이다. 즉 분산서비스거부공격은 시스템과 주변환경



(a) (b)

그림 3 이상징후의 두 가지 범주의 특성 (a) 상태벡터 개수의 급증을 초래하는 분산서비스거부공격의 예 (b) 상태수의 급증을 초래하는 포트스캐닝의 예

이 가진 토큰의 변화폭이 커서 상태벡터의 회수인 상태수는 적지만 다수 개의 상태벡터를 만들어낸다. 그림 3(b)의 포트스캐닝 등의 네트워크 공격은 Interactive 형태로 공격자와 공격대상 호스트가 패킷을 주고받으므로 소수의 상태벡터들에 집중되어 있는 형태이다. 즉 포트스캐닝은 시스템과 주변환경이 가진 토큰의 변화폭이 적은 대신 같은 상태벡터가 기록되는 횟수인 상태수는 큰 값을 가진다. 이상징후는 그림 3(a)와 같이 상태벡터의 개수가 증가하거나, 그림 3(b)와 같이 상태수가 증가하는 형태로 발생한다. 식 (1)에서는 상태벡터의 개수와 상태수의 확률을 모두 고려하고 있어 이로써 효율적으로 보다 정확하게 이상징후를 탐지할 수 있다.

2장에서 설명한 기존 제안 기법은 분산서비스거부 등의 네트워크 공격과 포트스캐닝 등의 네트워크 공격 동시에 모두는 탐지가 불가하다. 시계열예측 기법, 주성분분석 기법의 경우 단위 시간당 패킷크기(또는 패킷수)를 기반으로 이상징후를 탐지하기 때문에 단위시간당 패킷크기와 패킷수가 급증하는 분산서비스거부공격 등의 네트워크 공격만이 탐지가 가능하다. 포트 스캐닝 등의 네트워크 공격은 단위 시간당 패킷크기에 영향을 미치는 정도가 적기 때문에 상대엔트로피 기법의 경우 정상 트래픽에서 플래그, 목적지 포트번호만이 엔트로피의 관측대상이기 때문에 포트스캐닝 등의 네트워크 공격만이 탐지가 가능하다.

본 논문에서는 각 시스템과 주변환경이 교환하는 패킷의 이동에 따라 변화하는 토큰 수를 관찰하여 상태벡터로 표현하고, 단위시간 내에 각 상태벡터가 기록된 회수를 상태수로 표현하였다. 상태수의 확률을 엔트로피로 측정하여 이상징후를 탐지하는 제안 기법은 이상징후의 분산서비스거부공격 및 포트스캐닝 등의 두 가지 범주 모두 탐지 가능한 효율적인 이상징후 탐지 기법이다.

4. 트래픽 데이터 셋

이 장에서는 검증을 위해 사용한 트래픽 데이터 셋의

표 1 트래픽 데이터 셋 정보

데이터	이상징후	기간	데이터볼륨	연도	수집정보
MAWI	슬래머웜	7m	66MB	2003	[13]
CAIDA	위티웜	9m	100MB	2001	[14]
NLANR	코드레드웜	9m	96MB	2001	[15]
MIT99	서비스거부	23h50m	382MB	1998-1999	[16]
MIT00	분산서비스거부	2h14m	117MB	2000	[16]

출처와 포함하는 네트워크 공격에 관해 기술한다. 본 논문에서는 제안한 기법의 실효성을 검증하기 위해 실제 트래픽 데이터 셋을 수집하였다. 수집한 트래픽은 5개의 네트워크 트래픽 데이터 셋이며 이상징후는 각 1개씩 포함하고 있다. 이상징후는 각각 슬래머 웜, 위티 웜, 코드레드 웜, 서비스거부공격, 분산서비스거부공격이다.

MAWI Working Group[13]에서 SQL 슬래머 웜이 포함된 트래픽 데이터 셋을 수집하였다. SQL 슬래머 웜은 UDP 포트번호 1434번으로 공격대상 호스트의 정상적인 서비스를 마비시키는 특징이 있다. CAIDA[14]에서는 위티 웜이 포함된 트래픽 데이터 셋을 수집하였다 위티 웜은 취약점이 존재하는 방화벽 프로그램의 결함을 악용하여 UDP 포트번호 4000번으로 전파되는 웜이다. NLANR[15]에서는 코드레드 웜이 포함된 트래픽 데이터 셋을 수집하였다. 코드레드 웜은 Internet Information Services (IIS) 웹서버를 공격하는 웜으로 공격대상 호스트에 서비스거부공격을 시도한다. MIT Lincoln Lab[16]에서는 서비스거부공격과 분산서비스거부공격이 포함된 트래픽 데이터 셋을 수집하였다. 서비스거부공격이 포함된 트래픽 데이터 셋에는 Smurf, IP 스캐닝, 포트 스캐닝 등의 네트워크 공격과 서비스 취약점을 악용하는 네트워크 공격이 다수 포함되어 있다. 분산서비스거부공격이 포함된 트래픽 데이터 셋에는 목적지 IP로 다수의 패킷을 보내는 분산서비스거부공격이 포함되어 있다. MIT Lincoln Lab에서는 연도별로 트래픽 데이터 셋을 제공하는데 1998년과 1999년 트래픽 데이터 셋에는 서비스거부공격이, 2000년 트래픽 데이터 셋에는 분산서비스거부공격이 포함되어 있다.

슬래머 웜과 코드레드 웜, 위티 웜이 포함된 트래픽 데이터 셋은 이상징후 트래픽만을 포함하고 있어 여러 기법들의 실효성을 검증하기 위해서는 공정하지 못하다. 본 논문에서는 정확한 검증을 위해 테스트베드를 구축하고 공격이 없는 정상 트래픽을 배경 트래픽으로 재생한 후 이상징후 트래픽을 재생하여 결합한 데이터 셋을 사용하였다.

5. 구현 및 결과

본 논문의 엔트로피에 기반한 이상징후 탐지 기법은 펄을 사용하여 구현하였다. 네트워크 패킷의 연산처리를 위해서 CPAN[20]에서 제공하는 펄 모듈을 이용하였다. 해당 모듈은 네트워크 패킷 연산처리를 프로토콜 별로 용이하게 하기위해 제공되는 모듈로써, Net::Pcap 모듈[21]과 NetPacket 모듈[22]이다. 본 논문에서는 상기의 모듈을 이용하여 제안 기법을 구현하였다.

본 장에서는 IP의 엔트로피와 포트번호의 엔트로피의 주기적인 관측을 통해 이상징후를 탐지하는 방법에 관해 기술하고, 트래픽 데이터 셋에 적용한 결과에 대해 기술한다. 이상징후를 탐지하기 위해 다음의 식 (4)를 이용한다. 식 (4)에서 E_{1i}는 i번째 단위시간에서 관측한 IP의 엔트로피이며, E_{2i}는 i번째 단위시간에서 관측한 포트번호의 엔트로피이다. 현재 단위시간이 k번째 단위시간이라고 했을 때 다음 식 (4)를 만족하면 이상징후라고 판단한다.

$$\min\left(\left|\frac{\sum_{i=1}^{k}E_{1i}}{k}-\frac{\sum_{i=1}^{k-1}E_{1i}}{k-1}\right|,\left|\frac{\sum_{i=1}^{k}E_{2i}}{k}-\frac{\sum_{i=1}^{k-1}E_{1i}}{k-1}\right|\right)\geq R \quad (4)$$

IP의 엔트로피와 포트번호의 엔트로피에 대해 각각 k번째 단위시간에서의 엔트로피의 평균값과 $k-1$번째 단위시간까지의 엔트로피의 평균값의 차의 절대값을 구한다. 둘 중 더 적은 값이 임계값 R값만큼 넘어서면 이상징후라고 판단한다. 식 (4)의 임계값은 경험적으로 관찰하여 이상징후 오탐과 미탐을 최소화할 수 있는 값인 0.01로 정하였다. 하지만 임계값은 추후 비교에서는 동적인 값으로 할당되며, 0.01이 결정값은 아님을 분명히 한다. 또한, 단위시간인 t는 1초로 분석되었다.

본 논문에서는 보다 시각적인 효과를 위해 관측한 엔트로피를 주기적으로 2차원 그래프에 표기한다. 2차원 그래프의 x축에는 IP의 엔트로피를 표기하고, y축에는 포트번호의 엔트로피를 표기한다. 이상징후를 뚜렷하게 구분할 수 있도록 정상 트래픽만이 포함된 정상구간과 이상 트래픽과 정상 트래픽이 결합된 이상구간을 구분하여 표기하도록 한다. 정상구간과 이상구간을 구분하는 방법은 다음과 같다. 슬래머 웜, 코드레드 웜, 위티 웜이 포함된 트래픽 데이터 셋은 공격만이 포함되어 있으므로 정확한 검증을 위해서는 정상 트래픽과의 결합이 필수적이다. 이에 본 논문에서는 Snort[18]과 Bro[19]등의

대표적인 침입탐지시스템으로 이상징후가 없음이 판명된 정상 트래픽을 배경트래픽으로 트래픽 데이터 셋과 결합하였다. 서비스거부공격과 분산서비스거부공격이 포함된 트래픽 데이터 셋은 정상구간과 이상구간의 구분에 대해 [17]에서 제시하고 있다. 2차원 그래프의 시각적인 구별을 위해 정상구간과 이상구간의 엔트로피를 각 20점씩 샘플링하여 표기하였다.

그림 4는 서비스거부공격이 포함된 트래픽 데이터 셋에 제안한 알고리즘을 적용한 결과이다. 서비스거부의 특성상 특정 공격대상 IP, 특정 공격대상서비스 포트번호로 패킷이 집중되는 경향이 있어 엔트로피가 급증하여 탐지가 용이하다.

그림 5는 분산서비스거부공격이 포함된 데이터 셋에 제안한 알고리즘을 적용한 결과이다. 분산서비스거부공격이 공격대상 IP로 수많은 좀비들이 동시에 공격을 가해 공격대상 IP 및 포트번호의 엔트로피가 급증하고 있다.

이 때문에 분산서비스거부공격이 서비스거부공격 보다 정상구간의 엔트로피와 이상구간의 엔트로피의 구별이 더 뚜렷하다.

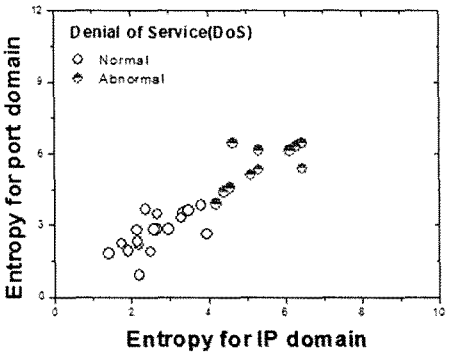


그림 4 서비스거부공격이 포함된 트래픽 데이터 셋에 제안 알고리즘을 적용한 결과

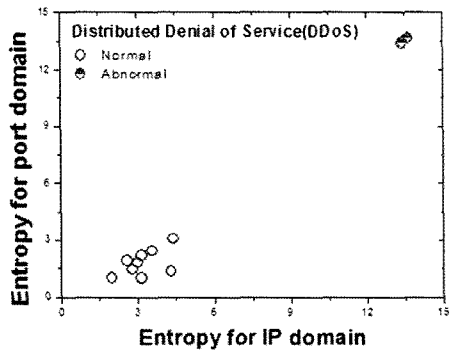


그림 5 분산서비스거부공격이 포함된 트래픽 데이터 셋에 제안 알고리즘을 적용한 결과

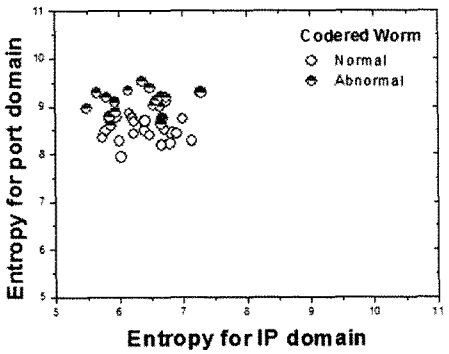


그림 6 코드레드 웜이 포함된 트래픽 데이터 셋에 제안 알고리즘을 적용한 결과

그림 6은 코드레드 웜이 포함된 트래픽 데이터 셋에 제안 알고리즘을 적용한 결과이다. IP의 엔트로피보다 포트번호의 엔트로피 증가가 더 뚜렷하다. 그 이유는 일반적으로 웜이 취약점이 있는 서비스 포트번호를 공격하여, IP 보다는 포트번호의 엔트로피를 증가시키기 때문이다. 코드레드 웜은 웹서버의 취약점을 공격하기 때문에 포트번호 80번으로 공격 패킷이 집중된다. 따라서 포트번호로 구분한 시스템에서 관측한 엔트로피의 증가가 IP의 엔트로피 증가보다 더 뚜렷하다.

그림 7은 슬래머 웜이 포함된 트래픽 데이터 셋에 제안 알고리즘을 적용한 결과이며 코드레드 웜과 마찬가지로 슬래머 웜이 특정 취약한 서비스포트번호 1434번을 대상으로 공격하기 때문에 포트번호의 엔트로피가 IP의 엔트로피보다 증가폭이 더 크다.

그림 8은 위티 웜이 포함된 트래픽 데이터 셋에 제안 알고리즘을 적용한 결과이다. 코드레드 웜, 슬래머 웜과 마찬가지로 위티 웜이 특정 취약한 서비스포트번호 4000번을 대상으로 공격하기 때문에 포트번호의 엔트로피의 증가폭이 IP의 엔트로피의 증가폭보다 크다.

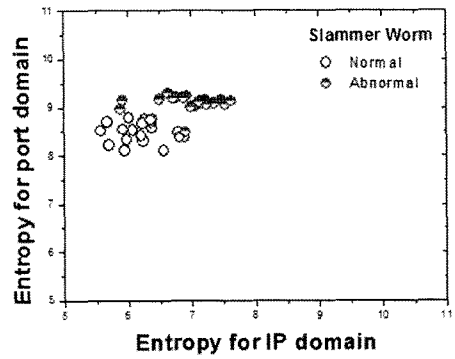


그림 7 슬래머 웜이 포함된 트래픽 데이터 셋에 제안알고리즘을 적용한 결과

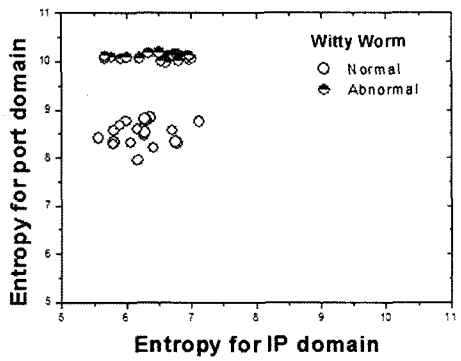


그림 8 위티 웜이 포함된 트래픽 데이터 셋에 제안알고리즘을 적용한 결과

위티 웜의 포트번호 엔트로피의 증가치는 +1.6으로 코드레드 웜과 슬래머 웜의 엔트로피 증가치 +0.6보다 높은데 그 이유는 다음과 같다. 슬래머 웜의 공격대상 포트번호 1434번과 코드레드 웜의 공격대상 포트번호 80번은 SQL 서비스와 HTTP 서비스 등의 정상적인 사용자들의 이용 빈도가 높은 포트번호인데 반해 위티 웜의 공격대상 포트번호 4000번은 정상적인 상황일 때는 거의 사용되지 않는 포트번호이다. 이에 슬래머 웜과 코드레드 웜 출현 시에도 정상적인 사용자들의 활동이 있기 때문에 포트번호의 엔트로피 증가가 위티 웜보다는 뚜렷하지 않다.

6. 비교 및 평가

본 장에서는 제안한 기법을 주성분분석 기법 및 시계열 예측 기법과 정확도 관점에서 비교 평가한다. 여기서 정확도라고 하는 것은 이상징후에 대해서 정확히 탐지를 하는 경우와 이상징후가 아닌 사건에 대해서 이상징후라고 오탐하는 않는 경우를 동시에 고려한다. 전통적으로 사건의 발생여부와 그에 대한 알고리즘의 판단을 평가하는 값으로 true positive(TN), false positive(FP), false negative(FN), true negative(TN)가 있다.

정확도를 측정하기 위해 많이 쓰이는 기준 중에 true positive rate(TPR)과 false positive rate(FPR)이 있다. TPR은 진짜 참인 사건들 중에 참이라고 옳게 판단한 비율이며 TP/(TP+FN)로 계산 할 수 있다. FPR은 진짜 거짓인 사건들 중에 참이라고 잘못 판단한 사건의 비율이며 FP/(FP+TN)으로 계산한다. TPR은 높을수록, FPR은 낮을수록 정확도가 높다고 판단된다. 예를 들어 100개의 사건 중 30개의 이상징후 사건이 있다고 가정하자. 어떤 알고리즘이 모든 100개의 사건을 이상징후로 판단한다면, TPR은 1(30/30)이 된다. 30개의 이상징후 사건들을 모두 이상징후로 판단하였으므로 표면적으로는 정확도가 100%인 것처럼 보인다. 하지만 FPR을 고려해 보면 1(70/70)로 측정되어 FP, 즉 오탐이 많아 정확도가 100%라고 판단할 수 없다. 따라서 TPR과 FPR을 동시에 고려해서 판단한다면 사건에 대한 판단의 정확도를 정밀하게 측정할 수 있게 된다.

높은 적중률을 위해서 TRR을 최대화하고 FRR을 최소화하는 이상징후 탐지 알고리즘에서는 임계치 설정이 중요하다. 임계치 설정이 중요한 만큼 탐지 알고리즘을 비교할 시에는 두 알고리즘의 임계치를 동일하게 할 필요가 있다. 이와 같은 임계치를 절대적으로 동일하고 하는 작업은 쉬운 일이 아니다. 대신에, 임계치의 상대적인 변화에 따른 정확도를 비교하는 방법이 널리 사용되고 있다. 이런 방법 중에 하나가 Receiver Operating Characteristics(ROC) 그래프이다.

ROC 그래프는 임계치 변화에 따른 정확도의 변화를 0과 1 사이 단위공간에 정규화하여 알고리즘을 비교하는 방법이다. ROC 그래프는 x축을 FPR, y축을 TPR로 하고, 임계치 값을 변화시켜 TPR과 FPR의 쌍을 구하고 그 값을 2차원 그래프에 표시하는 방법이다[23]. 모든 알고리즘은 (0,0)에서 시작해서 (1,1)에서 끝나게 되는데 정확도가 높은 알고리즘일수록 ROC 그래프에서 (0,1) 점, 즉 TRR은 1이고 FRR은 0인 점에 근접하게 된다. 다시말해서, 어떤 알고리즘의 ROC 그래프가 그린 면적이 크면 클수록 정확도가 높은 알고리즘이라고 할 수 있다.

다음은 본 논문에서 제안한 엔트로피 기반의 이상징후 탐지 기법을 시계열 예측 기법 중 EWMA, Holt-winters 기법 및 주성분분석 기법와 비교하여 ROC로 도시한 그림이다. 주성분분석은 다중링크 트래픽을 대상으로 하므로, 입력 트래픽 데이터 셋에서 단일링크 트래픽을 복구해 내는 작업이 요구된다. 이를 위해서는 트래픽이 수집된 해당 네트워크 구조가 필수적으로 요구된다.

그림 10은 서비스거부공격이 포함된 트래픽 데이터 셋에 각각의 알고리즘을 적용하였을 때의 ROC 그래프 결과이다. 서비스거부공격이 포함된 트래픽 데이터 셋은 포트스캔, IP 스캔, 스머프, 이메일공격 등 대표적인 서비스거부공격이 포함된 트래픽이나 서비스거부공격은 매우 간헐적으로 포함되어 있다. 이에 시간당 패킷수의 변화가 거의 없어 시계열 알고리즘으로 탐지하기 다소 어렵다. 마찬가지로 IP 및 포트의 엔트로피의 변화가 간헐적으로 발생하여 제안 알고리즘의 정확도가 그다지 뛰어나지 않았다. 분산서비스거부공격은 매우 큰 패킷수, 양의 공격대상 호스트로 집중되어 매우 짧은 시간에 나타난다. 그림 9에서 보듯이 제안기법의 우수성이 뛰어나고, 시계열 예측기법의 우수성이 가장 낮은 것으로 판

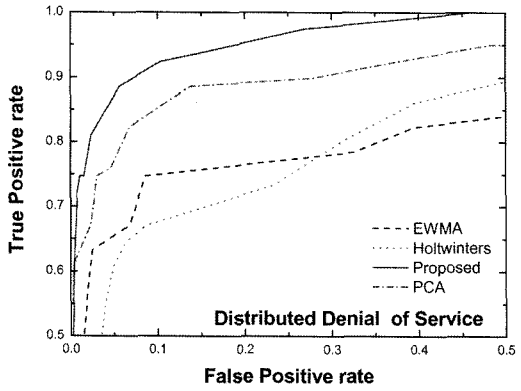


그림 9 분산서비스거부공격이 포함된 트래픽 데이터 셋에 각 알고리즘을 적용한 ROC 그래프

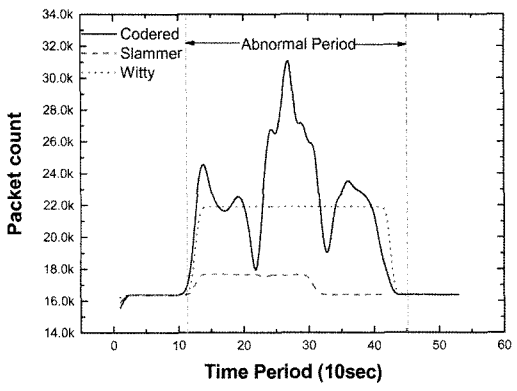


그림 11 슬래머 웜, 코드레드 웜, 위티 웜이 포함된 트래픽 데이터 셋의 패킷수

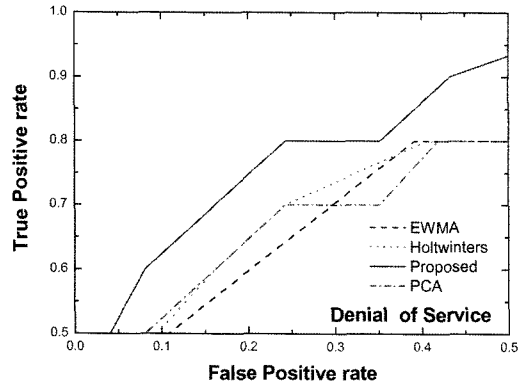


그림 10 서비스거부공격이 포함된 트래픽 데이터 셋에 각 알고리즘을 적용한 ROC 그래프

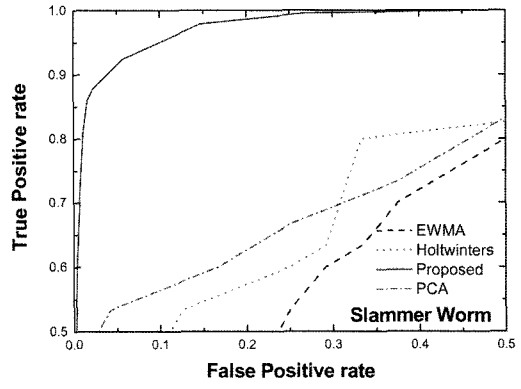


그림 12 슬래머 웜이 포함된 트래픽 데이터 셋에 각 알고리즘을 적용한 ROC 그래프

명되었다. 이유는 분산서비스공격이 포함된 트래픽 데이터 셋은 패킷수가 100개 미만의 극히 미미한 수준의 트래픽으로 매우 짧은 시간에 이상징후가 나타난다. 이 때문에 시계열 알고리즘으로는 민첩하게 탐지하기 어려운 반면, 제안 기법의 경우 IP의 엔트로피가 급격히 증가하기 때문에 기민하게 탐지할 수 있다.

그림 11은 슬래머 웜, 코드레드 웜, 위티 웜이 포함된 트래픽 데이터 셋의 시간단위 패킷수를 분석한 결과이다. 이상구간은 시간단위로 12~45이며 그 정상구간이다. 그림 11에서 보듯이 코드레드 웜이 포함된 트래픽 데이터 셋은 이상구간일 때 패킷수가 급증하며, 그 외 슬래머 웜 및 위티 웜이 포함된 트래픽 데이터 셋은 패킷수의 변화가 미미한 것으로 나타났다. 슬래머 웜은 UDP 포트번호 1434번으로 54byte의 고정된 양의 패킷을 보내는 웜이다. 또한 그림 11에서 보듯이 패킷수가 이상징후가 나타남에 따라 급증하지 않기 때문에 시계열예측 기법으로는 탐지하기 어려우며, 이는 주성분분석 기법에서도 마찬가지이다. 그림 12에서 보듯이 제안 알고리즘에서는 포트번호의 엔트로피가 급격히 증가하여 탐지가 가능하다.

그림 13에서 보듯이 코드레드 웜의 경우 패킷수의 변동폭이 커서 시계열예측 기법 및 주성분분석 기법의 우수성이 제안 알고리즘과 유사하다. 그렇지만 제안 알고리즘은 낮은 임계치 값을 설정하였을 때 보다 높은 우수한 성능을 보이고 있다. 그림 13에서 보듯이 전반적으로 알고리즘의 정확도가 하락하는 이유는 다음과 같다. 코드레드 웜은 일반적인 HTTP의 포트번호인 80번을 대상으로 하기 때문에, 미탐이 증가하였다. 위티 웜의 경우 패킷수가 이상구간 동안 변동이 거의 없다. 그렇지만 UDP 포트번호 4000번으로 집중되기 때문에 제안 기법에서는 포트번호의 엔트로피로 이상징후를 탐지할 수 있다. 그림 14 위티 웜에서는 전반적으로 정확도가 하락한다. 그 이유는 위티 웜을 제공하는 곳에서 정확한 호스트 정보를 제공하지 않기 때문이다.

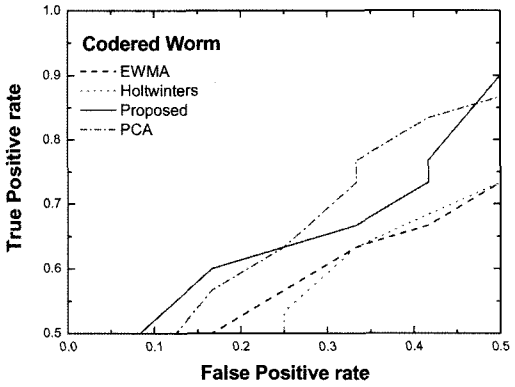


그림 13 코드레드 웜이 포함된 트래픽 데이터 셋에 각 알고리즘을 적용한 ROC 그래프

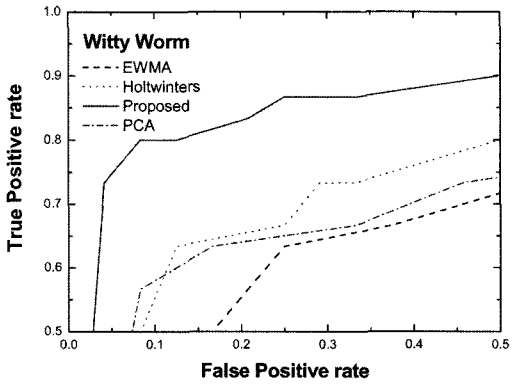


그림 14 위티 웜이 포함된 트래픽 데이터 셋에 각 알고리즘을 적용한 ROC 그래프

7. 결 론

이상징후는 네트워크 자원을 낭비하여 사용자에게 제공되는 서비스의 질과 보안강도를 하락시키는 존재로써, 이러한 이상징후를 높은 정확도로 탐지하는 것은 매우 시급한 문제이다. 이에 본 논문에서는 패킷의 이동에 따른 패킷역학의 변화를 열역학의 엔트로피를 통해 관찰하여 보다 효율적인 이상징후 탐지 알고리즘을 제안하였다. 기존 제안 기법은 분산서비스거부 등의 네트워크 공격과 포트스캐닝 등의 네트워크 공격 동시에 모두 효율적으로 탐지하지 못하였다. 본 논문에서는 각 시스템과 주변환경이 교환하는 패킷의 역학에 따른 엔트로피 관찰을 통한 이상징후 탐지 기법을 제안하였다. 실제 트래픽 데이터 셋에 적용한 결과와 ROC 그래프를 통한 타 알고리즘과의 비교평가결과로 비추어 볼 때 이 기법은 이상징후의 두 가지 범주 모두 높은 탐지율로 탐지 가능한 효율적인 이상징후 탐지 기법이다

참 고 문 헌

[1] Jelena Mirkovic, Sven Dietrich, David Dittrich, Peter Reiher, "Internet Denial of Service: Attack and Defense Mechanisms," Prentice Hall, December 2005.

[2] Jake D.Brutlag, "Aberrant Behavior Detection in Time Series for Network Monitoring," Proceedings of the *14th Systems Administration Conference* (LISA), December 2000.

[3] Anukool Lakhina, Mark Crovella and Christophe Diot, "Diagnosing Network-Wide Traffic Anomalies," *ACM Special Interest Group on Data Communication* (SIGCOMM), August 2004.

[4] Yu Gu, Andrew McCallum and Don Towsley, "Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation," *ACM Internet Measurement Conference* (IMC), 2005.

[5] Yin Zhang, Zihui Ge, Albert Greenberg and Matthew Roughan, "Network Anomography," *ACM Internet Measurement Conference* (IMC), October 2005.

[6] Yerin Yoo, "Tutorial on Fourier Theory," March 2001.

[7] Wenke Lee and Dong Xiang, "Information-Theoretic Measures for Anomaly Detection," *IEEE Symposium on Security and Privacy*, March 2001.

[8] Arno Wagner and Bernhard Plattner, "Entropy based Worm and Anomaly Detection in Fast IP Networks," *IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise* (WETICE), June 2005.

[9] Anukool Lakhina, Mark Crovella and Christophe Diot, "Mining Anomalies Using Traffic Feature Distributions," *ACM Special Interest Group on Data Communication* (SIGCOMM), August 2005.

[10] Kuai Xu, Zhi-Li Zhang, and Supratik Bhattacharyya, "Profiling Internet Backbone Traffic: Behavior Models and Applications," *ACM Special Interest Group on Data Communication*(SIGCOMM), August 2005.

[11] George Nychis, Vyas Sekar, David G.Andersen, Hyong Kim and Hui Zhang, "An Empirical Evaluation of Entropy-based Traffic Anomaly Detection," ACM Internet Measurement Conference (IMC), October 2008.

[12] Roberto Togneri, Christophe J.S.deSilva, "Fundamentals of Information Theory and Coding Design," CHAPMAN & HALL/CRC 2003.

[13] MAWI Working Group Traffic Archive, *available at* http://tracer.csl.sony.co.jp/mawi/

[14] CAIDA - The Dataset on the Witty Worm, *available at http://www.caida.org/data/passive/witty_worm_*dataset.xml

[15] NLANR network traffic packet traces, *available at*

http://pma.nlanr.net/Traces/Traces/long/cred/20010810/
[16] MIT Lincoln Laboratory - DARPA Intrusion Detection Evaluation Data Sets, *available at http://www.ll.mit.edu/IST/ideval/data/data_index.html*
[17] MIT Lincoln Laboratory - 1998 Training Data Attack Schedule, *available at* http://www.ll.mit.edu/IST/ideval/docs/1998/attacks.html
[18] Snort, *available at* http://www.snort.org/
[19] Bro Intrusion Detection System, *available at* http://www.bro-ids.org/
[20] Comprehensive Perl Archive Network (CPAN), *available at* http://www.cpan.org/
[21] Marco Carnut, Tim Potter, Bo Adler, Peter Lister, "Net::Pcap," *available at http://search.cpan.org/~saper/Net-Pcap-0.14/Pcap.pm*
[22] Tim Potter, Stephanie Wehner, "NetPacket," *available at http://search.cpan.org/~atrak/NetPacket-0.04/*NetPacket.pm
[23] Tom Fawcett, "ROC Graphs: Notes and Practical Considerations for Researchers," March 2004.
[24] R.Braden, D.Clark, S.Shenker, "Integrated Services in the Internet Architecture: an Overview," RFC 1633, June 1994.

한 찬 규

2006년 성균관대학교 컴퓨터공학과(공학사). 2008년 성균관대학교 전자전기컴퓨터공학과(공학석사). 현재 성균관대학교 휴대폰학과 박사과정. 관심분야는 인터넷 보안, 이상징후 탐지 등

최 형 기

1992년 성균관대학교 전자공학과(공학사). 1996년 Polytechnique University 전기전자(공학석사). 2001년 Georgia Institute of Technology 전기전자(공학박사). 2001년~2004년 미국 Lancope. Inc. 연구원. 2004년~2006년 성균관대학교 정보통신공학부 전임강사. 2006년~현재 성균관대학교 정보통신공학부 조교수. 관심분야는 인터넷 보안, 모바일 커뮤니케이션 등