情報保護學會論文誌
第14券 第4號, 2004. 8

트래픽 분석에 의한 광대역 네트워크 조기 경보 기법*

권 기 훈[a)†‡], 한 영 구[a)], 정 석 봉[a)], 김 세 헌[a)], 이 수 형[b)], 나 중 찬[b)]
한국과학기술원[a)], 한국전자통신연구원[b)]

Fast Detection Scheme for Broadband Network Using Traffic Analysis

Ki Hoon Kwon[a)†‡], Young-Goo Han[a)], Seok-Bong Jeong[a)], Sehun Kim[a)],
Soo Hyung Lee[b)], Jung Chan Na[b)]
KAIST[a)], ETRI[b)]

요 약

인터넷의 급속한 발달과 더불어 네트워크 환경에서의 침입은 빠르게 증가하고 있으며, 그 피해 또한 급격히 증가하고 있다. 최근의 인터넷 공격은 특정 호스트나 네트워크에 대한 피해를 초래할 뿐만 아니라, 네트워크 전반의 성능 저하를 유발한다. 기존의 침입 탐지 시스템은 각 지역망 및 특정한 대상 시스템을 보호하기 위한 솔루션들로, 기간망 수준의 실시간 공격 탐지에 적용하기 힘든 문제점을 가지고 있다. 본 논문에서는 네트워크 수준의 실시간 공격탐지를 위하여 각 포트별 트래픽을 대상으로 지수평활법을 적용하는 광대역 네트워크 침입 탐지 기법 제안하였다. 8일간의 기간망의 트래픽 데이터를 대상으로 한 실험에서, 제안한 기법은 공격으로 추정되는 급격한 트래픽의 증가를 적절히 탐지함을 보여주었다.

ABSTRACT

With rapid growth of the Internet, network intrusions have greatly increased and damage of attacks has become more serious. Recently some kinds of Internet attacks cause significant damage to overall network performance. Current Intrusion Detection Systems are not capable of performing the real-time detection on the backbone network. In this paper, we propose the broadband network intrusion detection system using the exponential smoothing method. We made an experiment with real backbone traffic data for 8 days. The results show that our proposed system detects big jumps of traffic volume well.

Keywords : *Traffic Analysis, Fast Detection, Broadband Network*

Ⅰ. 서 론

인터넷의 사용이 급증하면서 국가 경제와 산업에 막대한 가치가 새롭게 창출되었으며, 기존 경제활동 또한 활력과 효율성이 제고되었다. 그러나 정보화가 급속히 진행됨에 따라 해킹, 바이러스, 웜 등의 사이버 상의 위협은 매우 빠른 속도로 증가하고 있다[19,20]. 또한 그 피해의 정도 역시 급격히 증가하여, 개별 시스템에 대한 공격뿐만 아니라 네트워크 전반에 대한 성능 저하를 유발하고 있다.

이러한 인터넷 침해에 대응하기 위한 대표적인 보안 솔루션으로 침입차단시스템과 침입탐지시스템을 들 수 있다. 침입차단시스템은 침입이 발생하지 않도록 네트워크의 출입구를 제어하는 기능을 수행하므로 인증을 받지 않은 외부의 접근 시도는 차단할 수 있

접수일 : 2004년 4월 26일 ; 채택일 : 2004년 7월 20일
* 본 연구는 정보통신부 대학 IT 연구센터 육성·지원사업의 연구결과로 수행되었습니다.
† 주저자, ‡ 교신저자 : khkwon@tmlab.kaist.ac.kr

으나, 이미 인증된 사용자나 이를 가장한 침입자에 의한 공격에는 취약하다. 특히 내부인 혹은 허가된 외부인에 의해 발생하는 시스템 및 네트워크 침입을 막기 위해서는 침입을 즉각적으로 탐지, 대처하는 기술이 필요하다. 침입탐지시스템은 이러한 요구에 부응하는 보안도구로서 정보시스템 또는 네트워크로부터 보안 관련 정보들을 수집, 분석하여 침입 또는 오용을 탐지할 뿐 아니라 침입에 대한 적절한 대응기능을 포함하는 시스템이다[1].

그러나 기존의 침입탐지시스템은 특정 호스트나 네트워크를 대상으로 침입을 탐지하기 때문에, 네트워크 전반을 통해 급속히 전파되는 침해를 조기에 적극적으로 대응하는 것에는 한계가 있다. 2003년 1월 25일에 발생한 MS-SQL 슬래머 웜의 경우, 웜 발생 후 10분만에 전세계의 취약 호스트의 90% 이상을 감염시킨 것으로 나타났다[2]. 또한 1434 포트를 통하여 방대한 양의 패킷을 무작위로 발생시켜 국가 전체적인 네트워크 불능사태를 초래하였다[3]. 이와 같이 네트워크 상에서 급속히 전파하는 인터넷 침해를 막기 위해서는 특정 호스트나 네트워크를 대상으로 하는 침입탐지기법 뿐만 아니라, 대규모의 광대역 네트워크를 대상으로 하는 조기 이상탐지기법이 필요하다.

본 논문에서는 네트워크 전반에 대한 침해에 대응하기 위하여, 지수평활법을 활용한 광대역 네트워크 침입 탐지 기법을 제안한다. 광대역 기간망에서의 포트별 트래픽 양을 지수평활법을 사용하여 분석하여, 네트워크 상에 발생한 침해를 탐지해 낼 수 있다.

II. 관련 연구

본 논문에서는 포트별 트래픽의 양을 트래픽 특성 파라미터로 선택하였다. 네트워크 트래픽 양의 정상상태를 정형화하기 위한 연구는 다양한 방법을 통하여 이루어져 왔다. Groschwitz와 Polyzos는 NSFNET 기간망의 주간 트래픽 양을 예측하기 위해서, Seasonal ARIMA 모형을 적용하였다[6]. 또한 Seasonal ARIMA 모형을 활용한 무선통신의 트래픽 예측에 관한 연구도 수행되었다[7]. 또한 웨이블릿 분석(Wavelet Analysis)을 사용하여 원래의 트래픽을 각각의 시간 주기에 따라 분해하여 각각의 필터링된 신호를 살펴봄으로써 트래픽 상황에 대한 장기적인 이상과 단기적인 이상을 쉽게 구분할 수 있다[8]. 퍼지-자기회귀 모형(Fuzzy-AR model)은 비선형적이고 비정류적인 고속 네트워크 트래픽을 예측하는 데 적합하다[9].

트래픽 양과 같이 시간에 따라 변동하는 데이터의 이상을 탐지하는 기법으로 시계열 분석이 있다. 서비스 관리의 측면에서 웹 서버의 초당 http 명령의 횟수에 대한 분석을 수행하기 위해서 분산분석(ANOVA, Analysis of Variance)을 이용하여, 월별 요인, 요일 요인, 시간 요인을 파악하여 데이터를 정류화한 후, 자기회귀모형에 적용한 연구가 수행되었다[4]. 그러나 분산분석은 방대한 양의 훈련 데이터가 필요하다는 단점이 존재한다. Kalman filter를 사용하면 이러한 문제점을 해결할 수 있다. 또한 우도비(Likelihood Ratio)를 활용하여, 장기간에 걸쳐 발생하는 네트워크 상황의 변화를 탐지할 수 있다[5]. 또한 통계적 품질 관리 기법 중 하나인, 지수가중이동평균(EWMA : Exponentially Weighted Moving Average) 기법을 MIT-LL 데이터에 적용하여 침입을 탐지하는 연구가 수행되었다[10]. 그러나 이러한 기법들은 필요한 정보의 양이 많고 연산에 필요한 시간이 길기 때문에 광대역 네트워크 상의 실시간 침입 탐지에 적용하기에는 많은 문제점을 지니고 있다.

인터넷 상의 공격을 조기에 탐지하기 위한 연구도 널리 수행되었다. Kalman filter를 사용한 인터넷 웜 감시 및 조기 경보 기법이 연구되었다[14]. 또한 SNMP를 기반으로 트래픽의 MIB 변수를 활용한 DDoS 조기 탐지 기법에 관한 연구도 수행되었다[15,16]. [17]에서는 Dempster-Shafer 증거이론을 활용하여 네트워크 침해를 조기에 경보하는 기법을 연구하였다. 또한 MIB 변수를 대상으로 통계적 분석을 수행하여 DoS 공격을 조기에 탐지하는 기법에 관한 연구도 수행되었다[18]. [19]에서는 mobile agent를 활용한 네트워크 조기 이상 탐지 구조를 제시하였다. 그러나 이 연구들은 특정 네트워크를 대상으로 하는 이상 탐지 기법을 제안하였기 때문에, 광대역 네트워크를 대상으로 적용하기 힘들다.

III. 시스템 요구 사항

1. 트래픽 특성 파라미터의 선택

네트워크 상의 침입을 탐지하기 위해서 우리는 네트워크를 통과하는 IP 패킷에 대한 정보를 IP 계층에서 쉽게 얻을 수 있다. 이러한 정보는 프로토콜의 종류, source IP 주소, source 포트, destination

IP 주소, destination 포트, 트래픽의 양 등이 있다. 또한 Cisco의 Netflow나 SNMP(Simple Network Management Protocol)의 MIB 정보를 활용할 수도 있다.

그러나 광대역 기간망이나 주요 게이트웨이를 통과하는 트래픽은 매우 방대하기 때문에, 다양한 트래픽 특성을 실시간으로 분석하는 것은 쉬운 일이 아니다[11]. 본 논문에서는 광대역 네트워크 상에서의 침해를 실시간으로 탐지하기 위해서, 비교적 쉽게 분석할 수 있는 목적지 포트별 트래픽의 양을 트래픽 특성 파라미터로 선택하였다. DDoS 또는 급속한 웜의 전파와 같은 네트워크 전반에 걸쳐 피해를 주는 유형의 침해가 발생하면, 해당 공격이 사용하는 포트의 트래픽 양은 급격히 증가한다. MS-SQL 슬래머 웜의 경우, 1434 포트를 통해서 급속히 전파되었으며, 해당 포트의 트래픽 양이 급속히 증가하였다[3].

일반적으로, DDoS 공격이나 Flooding 공격이 발생하면 그 공격과 연관된 트래픽의 양은 급격히 증가한다. 그러므로 해당 공격이 사용하는 포트의 트래픽을 분석하면 전체 포트의 통합된 트래픽 양을 분석하는 것보다 더욱 쉽게 공격의 발생 유무를 파악할 수 있다. [그림 1]은 1주일 간의 전체 트래픽 양을 [그림 2]는 같은 기간 동안의 1434 포트의 트래픽 양을 나타낸다. [그림 1]에서는 마지막 날에 발생한 공격을 탐지할 수 없으나 [그림 2]를 통해서는 쉽게 탐지할 수 있다.

2. 모델 요구 사항

광대역 네트워크 상의 조기 이상 탐지를 위하여 목적지 포트별 트래픽의 양을 대상 트래픽 파라미터로 선택하였다. 그러나 현재 인터넷에서 사용되는 포트의 수는 매우 많기 때문에, 각 포트별 트래픽의 정상 상태를 규명하고 이상 상태를 파악하는 방법은 시스템에 많은 부담을 주지 않는 계산 복잡도가 낮은 기법일 필요가 있다. 또한 신속한 이상 탐지를 위해서는 실시간에 가까운 빠른 속도로 적용이 가능해야 한다. 그리고 기간망에서는 비정상이 존재하지 않는 순수한 훈련 데이터를 얻기가 현실적으로 불가능하기 때문에 훈련 데이터 없이 적용할 수 있는 기법일 필

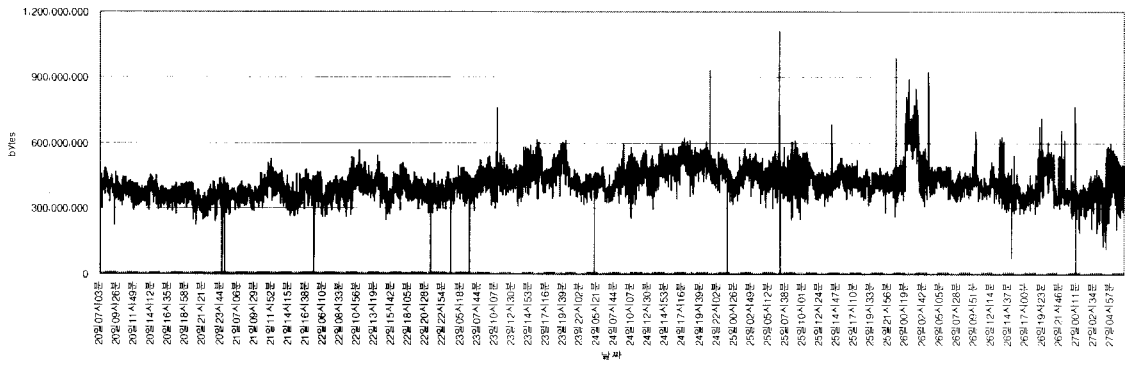


그림 1. 전체 트래픽 볼륨

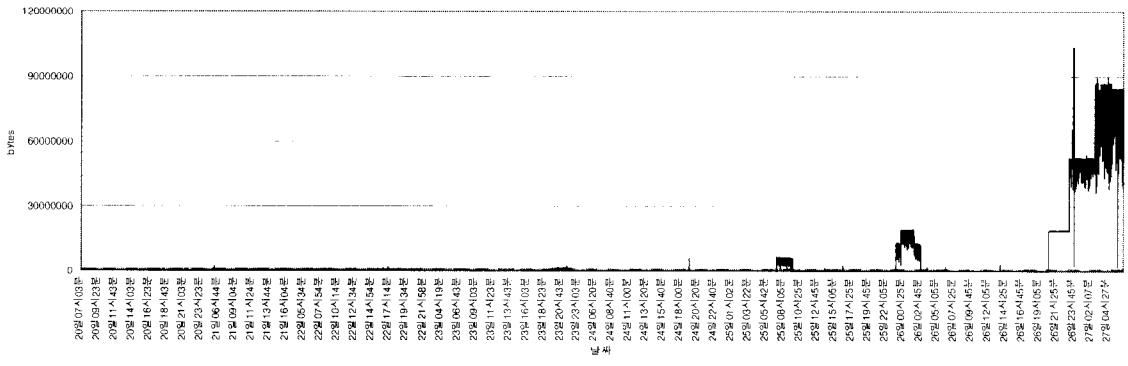


그림 2. 1434 포트의 트래픽 볼륨

요가 있다.

Ⅳ. 제안하는 이상 탐지 기법

광대역 네트워크의 이상상태를 조기에 파악하기 위해서는 복잡한 사전처리과정을 거치지 않고 얻을 수 있는 트래픽 파라미터를 대상으로 계산의 복잡도가 낮고 훈련 기간이 필요 없는 기법을 적용할 필요가 있다. 본 논문에서는 이러한 특성을 만족하는 지수평활법을 바탕으로 한 이상탐지기법을 제안하였다.

1. 지수평활법

지수평활법은 일정 시간 간격에 따라 관찰된 과거 데이터의 가중 평균을 다음 시점의 예측치로 사용하는 방법으로, 가중 평균을 계산할 때 최근의 데이터에 더 많은 가중치를 부과하는 발전된 형태의 이동평균법이다. 이 방법은 계산이 쉽고 필요한 자료가 적다는 장점을 가지고 있어 광대역 네트워크의 포트 별 트래픽에 대한 분석 기법으로 적합하다[12].

지수평활법은 t시점의 실제치 X_t가 입력되면, 아래의 식을 이용하여 $t+1$시점에 대한 예측치 Y_{t+1}을 계산한다.

$$Y_{t+1}=\alpha X_t+(1-\alpha),\ Y_t\ 0<\alpha\leq 1 \tag{1}$$

(1)에서 α는 평활상수이고, 일반적으로 0.05~0.3의 범위에서 정해진다. 평활상수는 실제치의 다음 시점의 예측치에 대한 가중치(weight)이다. 평활상수가 크면 최근 시점의 실제치가 큰 비중을 가지게 되고, 예측치는 실제치의 변동에 민감하게 반응한다. 반대로 평활상수의 값이 작을 경우에는 실제치의 비중이 작아져서 예측치의 변동이 작아지게 된다. 초기 시점에서의 예측치는 실제치와 같다.

예측치의 분산은 평균제곱오차(MSE : mean squared error)로 나타난다. 평균제곱오차는 다음과 같다.

$$\sigma_t^2=MSE=\sum\frac{E_t^2}{n},\ E_t=X_t-Y_t \tag{2}$$

그러나, 평균제곱오차를 이용하여 분산을 계산하기 위해서는 과거 n 시점 동안의 예측 오차에 대한 정보가 필요하다. 이러한 정보의 보관에 대한 부담을 줄이기 위해서, 평균제곱오차 대신에 지수평활법을 사용하여 분산을 추정할 수도 있다. 이 경우의 예측치의 분산은 다음과 같다.

$$\sigma_t^2=\gamma\cdot E_t^2+(1-\gamma)\cdot\sigma_{t-1}^2,0<\gamma\leq 1 \tag{3}$$

(3)에서 γ는 분산에 관한 평활상수이다. 분산의 초기값은 적절한 값으로 정해야 한다. 초기값이 너무 작으면 예측 초기의 오탐율이 높아지고, 초기값이 너무 크면예측 초기의 탐지율이 낮아진다.

그리고, 분산을 계산하는 대신 평균절대오차(MAD : Mean Absolute Deviation)를 사용할 수도 있다. MAD는 다음의 식으로 나타낼 수 있다.

$$MAD_t=\sum\frac{|E_t|}{n} \tag{4}$$

평균절대오차를 계산하기 위해서는 매 시점의 n시점의 예측 오차에 대한 정보가 필요하다. 분산의 경우와 마찬가지로, MAD의 경우에도 지수평활법을 이용하여 다음과 같은 방식으로 업데이트할 수 있다.

$$MAD_t=\gamma|E_t|+(1-\gamma)MAD_{t-1} \tag{5}$$

이 방법은 과거자료가 적게 필요하고, 최근의 자료가 과거자료보다 강조되는 장점이 있다. 여기서 MAD의 초기값은 적절한 값으로 결정되어야 한다. MAD이 너무 작으면 예측 초기의 오탐율이 높아지고, 초기값이 너무 크면 예측 초기의 탐지율이 낮아진다.

예측오차가 기대값이 0인 정규분포를 따르면, 예측오차의 표준편차 σ와 MAD사이에는 일반적으로 $MAD=0.8\sigma$의 관계가 성립한다[12].

앞에서 구한 예측치, 분산, MAD를 바탕으로 다음 시점의 실제치의 이상유무를 판단하는 기준인 관리한계를 결정할 수 있다. 관리한계란 통계적 품질관리에서 사용되는 개념으로, 제품의 특성치가 정상적인지 비정상인지 판단하는 기준이 되며, 상한과 하한으로 구성되어 있다. 제품의 특성치가 상한과 하한 사이에 존재할 경우, 제품은 정상적인 상태라고 판단하고, 이 범위를 벗어날 경우 비정상적인 상태라고 판단하여 이상의 원인을 파악하게 된다[21]. 이를 네트워크 트래픽에 적용하면 트래픽이 급격히 증가하여

관리한계를 벗어날 경우에 비정상적인 트래픽의 증가가 발생했다고 판단할 수 있다. 트래픽의 이상상태는 일반적으로 트래픽 양의 급격한 증가로 나타나므로, 관리한계의 상한만을 고려한다. 관리한계의 상한(UCL)은 다음과 같다.

$$UCL = Y_{t+1} + z_{\alpha}\sigma_t \approx Y_{t+1} + 1.25 \cdot z_{\alpha} MAD_t \quad (6)$$

z_{α}는 관리한계의 범위와 관리한계 내에 속할 확률을 결정하는 값으로, 표준정규분포표에서 찾을 수 있고, 대표적인 값은 [표 1]과 같다.

표 1. 관리한계 범위와 확률

z_{α}	관리한계내의 확률 $(1-\alpha)$
1	0.8159
1.65	0.9505
2	0.9772
2.58	0.9951
3	0.9987

2. 이상 상태 처리

실제 트래픽이 관리한계를 벗어날 경우, 그 시점의 트래픽은 정상상태가 아니라고 판단할 수 있다. 이러한 이상상태의 트래픽을 바탕으로 다음 시점의 트래픽을 예측하게 되면, 예측의 정확도는 떨어지게 된다. 그러므로 트래픽의 이상상태가 발견되면, 다음 시점에 대한 예측을 중지해야 한다.

트래픽의 이상에 대한 문제 해결이 끝나게 되면, 트래픽은 다시 정상적인 상태가 되었다고 말할 수 있다. 트래픽이 정상 상태로 돌아오면, 지수평활법을 다시 시작(reset)하여 트래픽을 예측한다.

트래픽의 이상상태가 종료했음을 판단하는 방법은 다음과 같다.

- □ 관리자의 직관 : 이상상태가 발생할 경우, 해당 포트의 트래픽의 급격한 증가가 발생한다. 정상상태의 트래픽과 이상상태의 트래픽의 차이가 크기 때문에 관리자는 쉽게 이상상태가 끝났음을 알 수 있다.
- □ 과거의 트래픽 양과 비교 : 과거의 현재 시간대의 평균 트래픽과 실제 트래픽을 비교하여, 실제 트래픽이 일정 범위에 속했을 경우, 이상상태가 종료했다고 판단한다. 그러나 이 방법을 사용하기 위해서는 많은 양의 정보가 필요하다.
- □ 정상상태의 트래픽의 최대값과 비교 : 과거의 일정 기간 동안의 정상상태의 최대 트래픽 양과 실제 트래픽을 비교하여 이상상태의 종료를 판단하는 방법이다.
- □ 트래픽의 안정성을 판단 : 이상상태가 발생하면, 해당 포트의 트래픽 양은 급격히 증가하기 때문에, 트래픽 양의 변화량 또한 증가하게 된다. 그리고 공격이 끝났을 경우에는 트래픽 양의 변화량이 감소한다. 이러한 특성을 활용하여, 트래픽 양을 1차 차분(differencing)한 값을 지수평활법으로 관찰하여, 이상상태가 종료했음을 알 수 있다.

3. 제안하는 기법

본 논문은 광대역 기간망에서의 포트별 트래픽 양을 바탕으로 네트워크의 이상을 탐지하려고 한다. 기간망의 경우, 방대한 양의 트래픽이 통과하기 때문에 패킷이 발생하는 매 시점마다 이상을 판단하는 것은 거의 불가능하다. 따라서 본 논문에서는 각 포트 별 트래픽의 양을 일정한 주기(1분 또는 더짧은 기간)마다 입력 받는 상황을 고려하였다.

시스템의 작동 과정은 [그림 3]과 같다.

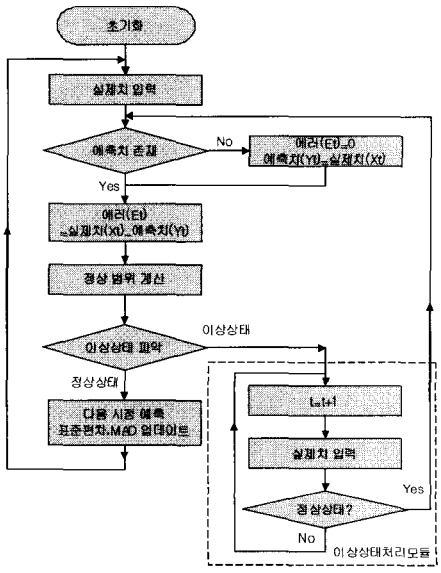


그림 3. 시스템 작동 과정

□ 초기화 : 처음 예측, 분석을 시작하기 앞서, 초기의 MAD 또는 표준편차를 결정한다.

□ 실제치 입력 : 트래픽 측정 장치로부터, t시점에서 실제로 측정된 트래픽 X_t를 입력 받는다.

□ 예측치 존재 판단 : 예측을 처음 시작하였거나 이상상태가 끝난 후, 지수평활법이 재실행되었을 경우에는 그 시점에 대한 예측치(Y_t)가 존재하지 않는다. 예측치가 존재하지 않을 경우에는 예측치가 실제치와 같다고 설정한다 ($Y_t = X_t$). 이 때, 예측오차(E_t)는 0이 된다. 예측치가 존재하는 일반적인 경우에는 예측오차는 실제치와 예측치의 차이가 된다 ($E_t = X_t - Y_t$).

□ 정상 범위 계산 : (6)을 바탕으로 예측치 (Y_t)와 MAD 또는 표준편차를 이용하여, 관리한계의 상한을 결정한다.

□ 이상 상태 파악 : 실제치(X_t)가 관리한계를 벗어날 경우($X_t > Y_t + z_a\sigma_t$)에는, 네트워크에 이상이 발생하였다고 판단하여, 지수평활법을 정지하고, 이상상태처리모듈을 수행한다. 실제치가 정상 범위에 있을 경우($X_t \leq Y_t + z_a\sigma_t$)에는 다음 단계로 진행하여 지수평활법을 계속 수행한다.

□ 다음 시점 예측, 표준편차, MAD 업데이트 : 관리한계 내의 실제치(X_t)를 사용하여 식 (1)에 따라, 다음 시점의 예측치(Y_{t+1})를 계산한다. 또한 현재의 예측 오차를 바탕으로 식 (3), (5)를 사용하여, 다음 시점의 정상 범위 계산에 필요한 표준편차 또는 MAD를 업데이트한다. 그 후, 다음 시점의 실제치가 입력될 때까지 대기한다.

□ 이상 상태 처리 모듈 : 만약, 이상 상태 파악 단계에서 트래픽의 이상상태가 발견되면, 지수평활법에 의한 예측을 중지하고, 트래픽이 다시 정상 상태가 되기를 기다린다. 현재의 이상 트래픽은 무시하고, 다음 시점의 트래픽을 입력 받는다. 현 시점의 트래픽이 정상적이라고 판단되면, 지수평활법을 다시 시작한다. 다시 시작되는 지수평활법에서, 표준편차 또는 MAD는 이상 발생 이전의 값이 그대로 사용되고, 그 시점에 대한 예측치가 없기 때문에 예측치의 초기값은 실제치와 같다고 설정된다. 입력된 트래픽이 여전히 이상상태로 판단되면, 정상상태의 트래픽이 도착할 때까지 대기한다.

V. 실험 및 결과

본 논문에서는 한국과 미국을 연결하는 기간망의 2003년 7월 20일에서 27일 사이의 트래픽 데이터를 바탕으로 실험을 수행하였다. 각 포트의 1분간의 트래픽의 양을 추출하여, 제안한 기법에 적용하였다.

본 논문에서 활용한 데이터는 제어 가능한 환경하에서 공격과 정상 상태의 트래픽을 생성한 것이 아니라 실제 네트워크 상의 데이터이기 때문에 이상상태와 정상상태를 정확히 규명할 수 없다. 그러므로 제안한 기법을 통해 이상유무를 탐지함에 있어서, 탐지율이나 오탐률을 활용한 객관적인 성능 평가가 불가능한 문제점이 존재한다.

1. 탐지 결과

본 논문에서는 트래픽이 많이 발생한 4개의 포트를 대상으로 실험을 수행하였다. 대상 포트는 1434, 80, 445, 137 포트이다. 1434 포트는 MS-SQL 슬래머 웜의 공격에 사용되는 포트로 2003년 7월에도 지속적인 공격이 발생하였다. 80 포트는 http를 이용한 웹 서비스에 사용되는 포트이며 CodeRed, Nimda 등의 공격에 사용하는 포트이다. 445 포트와 137포트는 각각 SMB, NETBIOS 서비스를 위한 포트이며, 각종 인터넷 웜과 바이러스의 공격에 사용된다[13].

1434 포트를 사용하는 MS-SQL 슬래머 웜의 경우 감염된 시스템은 매우 많은 패킷을 무작위로 발생시킨다고 알려져 있다[3]. [그림 4]에서, 제안한 기법은 이러한 트래픽 양의 큰 변화를 잘 감지한다는 사실을 알 수 있다.

80 포트는 웹 서비스를 위해 널리 사용되며 또한 다양한 공격이 발생하는 포트이다. [그림5]를 살펴보면, 정상적이라고 생각되는 트래픽의 변동을 비교적 정확하게 예측하고 있으며, CodeRed 혹은 Nimda로 의심되는 트래픽의 급격한 증가는 민감하게 탐지해 낸다는 것을 알 수 있다.

[그림 6]과 [그림 7]에서 제안한 기법은 445 포트와 137 포트에 대해서도 공격으로 추정되는 트래픽 양의 급격한 증가를 적절히 탐지한다는 것을 알 수 있다.

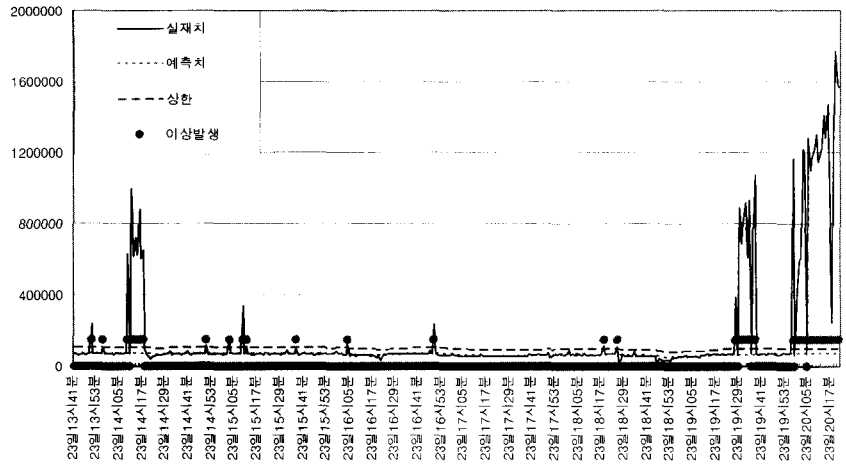


그림 4. 1434 포트의 이상 탐지 결과

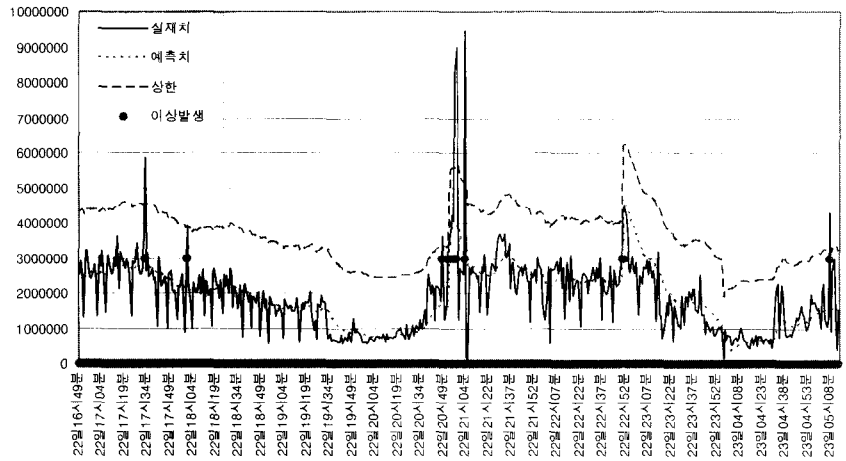


그림 5. 80 포트의 이상 탐지 결과 ($\alpha = 0.1, \gamma = 0,001, z_{\alpha} = 3$)

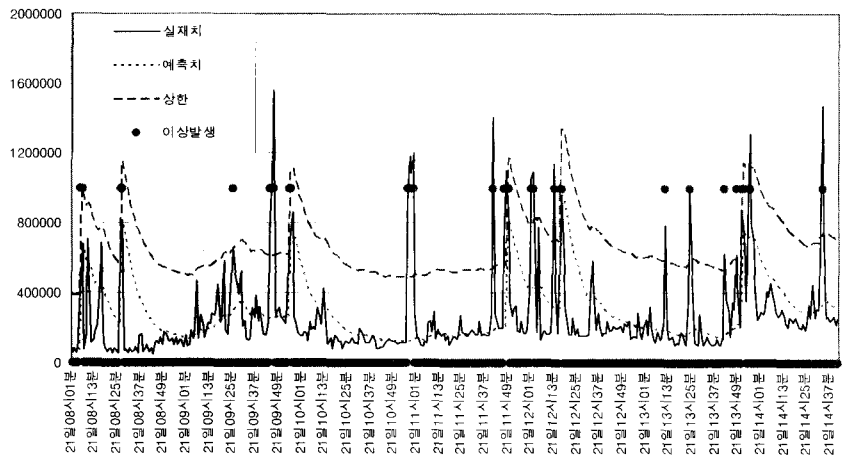


그림 6. 445 포트의 이상 탐지 결과 ($\alpha = 0.1, \gamma = 0,001, z_{\alpha} = 3$)

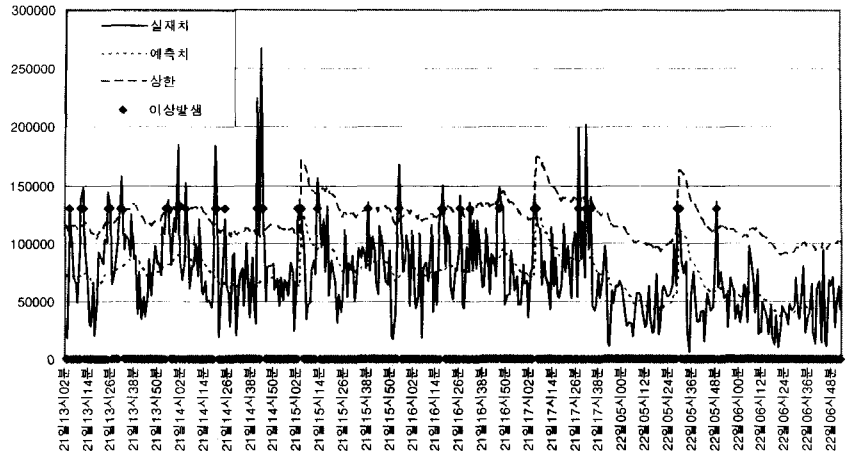


그림 7. 137 포트의 이상 탐지 결과 ($\alpha = 0.1, \gamma = 0,001, z_{\alpha} = 3$)

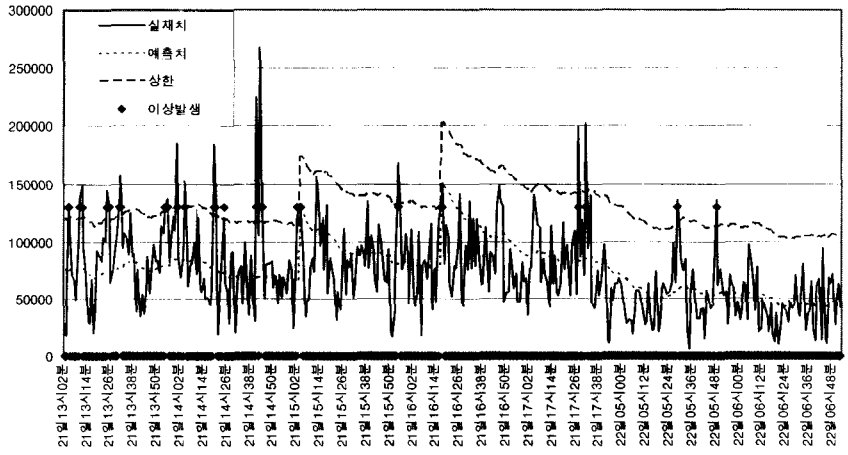


그림 8. 137 포트의 이상 탐지 결과 ($\alpha = 0.05, \gamma = 0,001, z_{\alpha} = 3$)

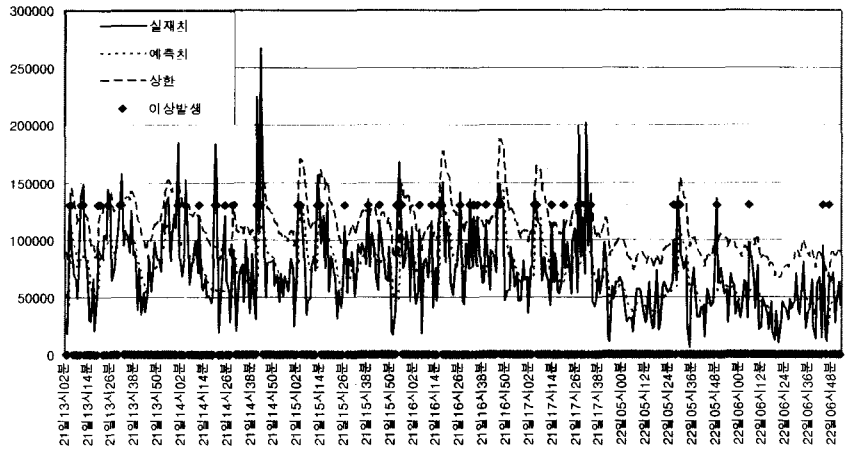


그림 9. 137 포트의 이상 탐지 결과 ($\alpha = 0.3, \gamma = 0,001, z_{\alpha} = 3$)

2. 평활 상수의 설정에 따른 이상 탐지 결과

지수평활법에 있어서 평활 상수의 결정은 모형의 성능에 큰 영향을 준다[12]. 훈련 데이터를 사용할 수 있을 경우, 예측치와 실제치의 오차를 최소화하는 평활 상수를 선택할 수 있다. 그러나 본 논문에서는 훈련 데이터를 사용하지 않는 시스템을 제안하였다. 따라서 평활 상수의 선택에 따른 이상 탐지 특성을 관찰하였다.

〔그림 8〕을 〔그림 7〕과 비교해 보면, 〔그림 8〕의 경우, 평활 상수가 작기 때문에 실제 트래픽 양의 변화에 둔감하게 반응한다는 것을 알 수 있다. 또한 트래픽 양의 변화에 민감하게 반응하지 못하기 때문에 예측 오차가 커져서 관리 한계는 넓어진다. 그러므로 평활 상수를 너무 작게 설정할 경우, 시스템의 탐지율이 저하될 가능성이 있다.

〔그림 9〕와 같이 평활 상수가 클 경우, 트래픽의 변화에 민감하게 반응하지만, 관리 한계는 좁아진다. 그러므로 평활 상수를 너무 크게 설정할 경우, 시스템의 오탐률(false positive error)이 증가할 위험이 있다.

〔표 2〕에서 일반적으로 평활 상수가 작을수록 이상 탐지 횟수가 작아지고, 평활 상수가 클수록 이상 탐지 횟수가 많아지는 경향을 확인할 수 있다.

표 2. 평활 상수에 따른 이상 탐지 횟수

	1434 포트	80 포트	445 포트	137 포트
0.01	1485	73	435	240
0.05	1723	143	480	194
0.1	1814	153	350	242
0.15	1944	252	534	285
0.2	1979	278	571	336
0.25	2013	329	588	390
0.3	2047	385	623	487
0.5	2233	695	823	1210

Ⅵ. 결 론

본 논문에서는 최근 증가하는 네트워크 전반에 대한 피해를 주는 인터넷 공격을 조기에 탐지하기 위하여 지수평활법을 활용한 광대역 네트워크 침입 탐지 기법을 제안하였다.

광대역 네트워크 상의 대량의 패킷을 대상으로 실시간 침입 탐지를 수행하기 위해서는 적절한 트래픽 특성 파라미터와 이상 탐지 기법의 선택이 중요하다. 네트워크를 통한 공격은 특정 포트를 사용하여 이루어진다. 따라서 트래픽을 각각의 포트에 따라 나누어 살펴보면 공격의 발생을 보다 정확히 탐지할 수 있다. 그러나 인터넷에서 사용되는 포트의 수가 많기 때문에, 시스템이 활용하는 정보의 양은 적어야 하고, 이상 탐지 기법은 간단해야 한다. 본 논문에서는 각 포트별 트래픽의 양을 지수평활법으로 분석하여 공격을 조기에 탐지하는 기법을 제안하였다.

8일간의 기간망의 트래픽을 대상으로 한 실험에서, 제안한 기법은 공격의 발생으로 추정되는 1434, 80, 445, 137 포트의 급격한 트래픽 양의 증가를 효과적으로 탐지함을 보여주었다. 또한 평활 상수의 변화에 따른 이상 탐지 특성의 변화도 살펴 보았다.

제안한 시스템은 현재 증가하는 네트워크의 전반적인 성능을 저하시키는 공격을 탐지하는데 효과적일 것으로 기대된다. 또한 기존의 호스트, 네트워크 침입 탐지 시스템과 함께 상호보완적으로 사용될 경우 더욱 뛰어난 성능을 보일 것으로 예상된다.

참 고 문 헌

〔1〕 Dorothy E. Denning, "An intrusion detection model", IEEE Transactions on Software Engineering, v.13 n.2, pp. 222-232, Feb. 1987

〔2〕 D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "The Spread of the Sapphire/Slammer Worm",http://www.caida.org/outreach/papers/2003/sapphire/sapphire.html

〔3〕 D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, N. Weaver, "Inside the Slammer worm", IEEE Security & Privacy Magazine, v.1 , i.4, pp. 33-39, July-Aug. 2003

〔4〕 J. L. Hellerstein, F. Zhang, P. Shahabuddin, "A statistical approach to predictive detection", Computer Networks, vol 35, pp.77-95, 2001

〔5〕 F. Zhang, J. L. Hellerstein, "An Approach to On-line Predictive Detection", In Proceedings of 8th International Symposium on Modeling,

Analysis and Simulation of Computer and Telecommunication Systems, Aug. 29-Sep. 2000

[6] N. K. Groschwitz and G. C. Polyzos, "A Time Series Model of Long-Term NSFNET Backbone Traffic", In Proceedings of IEEE International Conference on Communications, May 1994

[7] Y. Shu, M. Yu, J. Liu; Yang and O.W.W, "Wireless traffic modeling and prediction using seasonal ARIMA models", In Proceedings of IEEE International Conference on Communications, v.3, May 11-15, 2003

[8] P. Barford, J. Kline, D. Plonka and A.Ron, "A Signal Analysis of Network Traffic Anomalies", IMW'02, Nov.6-8, 2000

[9] B. Chen, S. Peng and K. Wang, "Traffic Modeling, Prediction, and Congestion Control for High-Speed Networks : A Fuzzy AR Approach", IEEE Transactions on Fuzzy Systems, v.8, n.5, Oct. 2000

[10] N. Ye, S. Vilbert and Q. Chen, "Computer Intrusion Detection Through EWMA for Autocorrelated and Uncorrelated Data", IEEE Transactions on Reliability, v.52, n.1, March 2003

[11] X. Gang, Z. Hui, "Advanced methods for detecting unusual behaviors on networks in real-time", In Proceedings of International Conference on Communication Technology Proceedings, v.1, pp.291-295, Aug. 2000

[12] G. Box , G. Jenkins, and G. Reinsel, Time Series Analysis, 3rd ed., Prentice Hall, 1994

[13] 한국정보보호진흥원, 2003년 7월 해킹바이러스 통계 및 분석 월보, 2003년 7월

[14] C. Zou, L. Gao, W. Gong, D. Towsley, "Monitoring and Early Warning for Internet Worms", In Proceedings of the 10th ACM Conference on Computer and Communication Security, October 2003

[15] J.B.D.Cabrera, L. Lewis, X. Qin, C. Gutierrez,W. Lee, R.K. Mehra, "Proactive intrusion detection and SNMP-based security management : new experiments and validation", In Proceedings of IFIP/IEEE Eighth International Symposium on Integrated Network Management, 24-28 March 2003

[16] J.B.D.Cabrera, L. Lewis, X. Qin, C. Gutierrez,W. Lee, R.K. Mehra, "Proactive Intrusion Detection and Distributed Denial of Service Attacks-A Case Study in Security Management", Journal of Network and Systems Management, vol. 10, num. 2, pp. 225-254, June 2002.

[17] J. Zhai, J. Tian, R. Du, J. Huang, "Network Intrusion Early Warning Model Based on D-S Evidence Theory", In Proceedings of 2003 International Conference on Machine Learning and Cybernetics, vol. 4, pp. 1972-1977, Nov. 2003

[18] J.Li, C. Manikopoulos, "Early statistical anomaly intrusion detection of DOS attacks using MIB traffic parameters", Information Assurance Workshop, 2003. IEEE Systems, Man and Cybernetics Society, pp. 53 - 59. June 2003

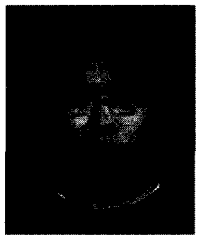
[19] J. Zhang, C Xiao, "A network early-warning architecture using mobile agent", In Proceedings of 2003 International Conference on Computer Networks and Mobile Computing, pp.349-352, Oct. 2003

[20] 조상현, 김한성, 이병희, 차성덕, "베이지언 추정을 이용한 웹 서비스 공격 탐지", 한국정보보호학회논문지, vol. 13, no. 2, pp. 115-126, April 2003

[21] D. C. Montgomery, Introduction to Statistical Quality Control, John Wiley and Sons, 1997

〈著者紹介〉

권 기 훈 (Ki Hoon Kwon) 학생회원
2001년 2월 : 한국과학기술원 산업공학과 졸업
2003년 2월 : 한국과학기술원 산업공학과 석사
2003년 3월~현재 : 한국과학기술원 산업공학과 박사과정
〈관심분야〉 정보보호, 네트워크 보안, 침입 탐지



한 영 구 (Young Goo Han) 학생회원
2002년 2월 : 한국과학기술원 산업공학과 졸업
2004년 2월 : 한국과학기술원 산업공학과 석사
2004년 3월~현재 : 한국과학기술원 산업공학과 박사과정
〈관심분야〉 정보보호, 네트워크 보안, 침입 탐지

정 석 봉 (Seok-Bong Jeong)
1999년 2월 : 한국과학기술원 산업경영학과 졸업
2001년 2월 : 한국과학기술원 산업공학과 석사
2001년 3월~현재 : 한국과학기술원 산업공학과 박사과정
〈관심분야〉 정보보호, 네트워크 보안, 침입 탐지

김 세 헌 (Sehun Kim) 정회원
1972년 : 서울대학교 물리학과 학사
1977년 : 스탠포드대학교 물리학과 석사
1981년 : 스탠포드대학교 OR 박사
1982년~현재 : 한국과학기술원 산업공학과 교수
2003년 : 정보통신부 정보보호실태 조사단 단장
2003년 : 한국정보보호학회 회장
2003년~현재 : 한국PKI포럼 이사
2004년~현재 : 국가정보원 국가정보보안협의회 산학연 회장
〈관심분야〉 OR, 이동통신, 정보보호, 네트워크 보안

이 수 형 (Soo Hyung Lee)
1993년 2월 : 한양대학교 전자공학과 석사
1993~1999년 : (주)데이콤 종합연구소 전임연구원
1999~2000년 : (주) ifeelnet 개발 팀장
2000년~현재 : 한국전자통신연구원 선임연구원
〈관심분야〉 네트워크 안정성 · 보안, 유비쿼터스 컴퓨팅 · 보안, 정보 보안

나 중 찬 (Jung Chan Na)
1986년 2월 : 충남대학교 계산통계학과 학사
1989년 2월 : 숭실대학교 전자계산학과 석사
2004년 2월 : 충남대학교 컴퓨터과학과 박사
1989.2~현재 : 한국전자통신연구원 책임연구원
〈관심분야〉 네트워크 보안, 실시간 시스템, 웹 서비스 보안