主題

보안측면에서의 네트워크 이상징후 분석기술

한국전자통신연구원 정보보호연구본부 나 중 찬
충남대학교 전기정보통신공학부 김 영 국

차 례

Ⅰ. 서 론

인터넷상에서 이루어지는 사이버공격의 발생 빈도는 점점 증가하고 있다. 실제로 코드레드(CodeRed), 님다(Nimda), SQL-worm 등 매년 우리는 일상적으로 보안사고를 거치고 있지만, 마음처럼 그렇게 쉽게 보안사고에 대비하지 못하고 있는 것이 현실이다.

이로 인해 인터넷 이용자들은 인터넷 환경에 대한 신뢰성에 의구심을 갖기도 하며, 시간적 경제적인 피해를 입게 되기도 한다. 또한 90년대 중반이후부터 시작된 초고속 인터넷 구축과 함께 인터넷의 급격한 확산에 많은 투자를 하였거나 최근 들어 인터넷 장비를 구매한 경험자들도 상당수 느끼는 일들이겠지만, 장비의 가격이 만만치 않고 현재 장비의 기능과 성능의 효용가치에 대한 회의론도 다시 짚어보게 된다.

아주 비관적인 경우, 향후에는 인터넷이 완전히 마비되어 쉽게 복구되지 않는 최악의 상황을 상상하는 것도 어렵지 않다. 이를 뒷받침할 만한 충분한 근거로는 자기 복제 능력을 가지고 빠른 속도로 인터넷을 통해 확산되는 웜(worm) 배포와 이로 인해 결과적으로 네트워크 서비스 제공 자체가 위협받는 단계에 이르렀던 점을 들 수 있다.

반면 현재의 네트워크 보안 기술은 공격이 발생하면 사후에 해당 공격에 대한 특징을 분석하여 공격패턴을 생성하고, 이후에 발생하는 공격을 탐지한다. 즉, 이는 각 로컬 도메인의 입구에 존재하는 방화벽 시스템과 결합하여 각 도메인으로 유입되는 공격 트래픽을 차단하는 대응방식으로 이용된다. 그러나 과다 트래픽을 유발하는 새로운 공격 방식이 출현하거나, 기존의 공격 방식에 의한 부수적인 효과에 의해 네트워크의 성능을 침해할 수준의 트래픽이 발생하는 경우, 이러한 방법을 통한 적절한 대응은 어렵다. 또한 더욱 다양한 보안 취약성을 이용해서 다양한 전파 경로를 통해 시스템을 공격하는 경우, IDS

(Intrusion Detectior System)와 같은 단일 보안 솔루션으로 이를 방어하는 것은 어렵다.

이를 극복하기 위해서는 네트워크 성능 저하(network performance failure)를 가져올 수 있는 공격으로 인한 네트워크 이상 트래픽에 대한 감지 능력과 감지된 트래픽에 대해 네트워크 차원에서 대응할 수 있는 기능이 요구된다고 할 수 있다. 이와 관련하여 보안 측면에서 트래픽 분석과 관련된연구는 최근에 활발히 이루어지고 있는 상태이나 발표되는 내용은 주로 학계에서의 이론적인 측면이 강하다. 실제 적용하기 위한 연구는 DARPA(Defence Advanced Research Projects Agency) 주도로 이루어진 LightHouse 프로젝트에서 일부 다루어졌다.

본 논문에서는 네트워크의 안정적인 운영을 위협하는 트래픽 기반의 공격을 분석함에 있어 정상 상태의 트래픽 패턴을 모델링하고, 이를 바탕으로 이상 트래픽을 판별하는 방법에 관하여 기술한다.

II. 네트워크 이상징후 분석

네트워크 공격의 인지는 직관적으로 네트워크가 느려지거나, 접속 불능 등의 네트워크 이상징후에 대한 신고가 접수되거나, 각 사업자와 이상징후를 분석하는 중앙센터를 연계하는 인터페이스를 통해 경보가 전달될 경우, 이를 확인하기 위해 트래픽 정보를 분석함으로써 이루어질 수 있다. 또는 주기적인 네트워크 원시 트래픽 정보 수집을 통한 지속적인 감시 활동에 의한 인지가 가능하다.

여기서 트래픽 분석이라고 하는 것은 넓은 의미로서, 트래픽 분석의 목적에 따른 판단자료를 산출하는 것을 말한다.

2.1 사이버 공격 패턴의 변화

2002년 발생된 코드레드 웜은 인터넷에 배포되어 9시간만에 전세계 25만대의 윈도우 시스템을 감염시켰고, 채 하루도 지나기 전에 윈도우 서버 40만대를 감염시키면서 네트워크 마비로 일손을 놓을 수밖에 없었다. 또한 워홀 웜에서 알 수 있듯이 웜의 전파속도를 향상시키면서 15분만에 인터넷에 존재하는 모든 취약한 시스템을 감염시켰다. 또한 2003년 1.25 인터넷 침해사고를 일으킨 SQL 슬래머 웜의 경우, 웜 감염 직후에 1434 포트의 트래픽이 급격히 증가하였으며, 발생 10분 이내에 전세계의 취약 서버의 90% 이상을 감염시키는 매우 빠른 전파속도를 보였다.

이상에서와 같이 최근의 인터넷 웜은 기존의 사이버 공격 패턴과는 달리 사용자의 도움이 없어도 스스로 동작해 복제해 나가는 자가복제 능력을 갖고 있으며, 사람이 발견하여 대응하기에는 역부족일 만큼 전파속도가 매우 빨라 네트워크 전체에 큰 피해를 주고 있다.

향후의 사이버 공격은 전파방식에 있어 이전의 공격보다 더욱 다양한 보안 취약성(multi-holes)을 이용해 다양한 전파경로(multi-modes)로 다양한 시스템을 공격시킬 것이다. 또한 공격의 형태를 스스로 변화시키면서 등장할 수 있거나, 일단 넓게 전파된 후 논리폭탄과 같이 특정 시간이나 특정 순간에 동작을 하는 방식으로 피해를 줄 수 있다.

2.2 트래픽 데이터

보안 측면에서의 트래픽 분석을 위해 기초가 되는 네트워크 원시 트래픽 데이터는 네트워크 상에서 현재 처리되고 있는 트래픽 현황을 나타내는 다음과 같은 항목들이 그 대상이 될 수 있다.

- SNMP MIB-II

  라인 인터페이스 단위의 트래픽 정보를 제공

o RMON MIB
네트워크 세그먼트 단위의 트래픽 정보를 제공
o Netflow
플로우 단위의 트래픽 정보를 제공

SNMP(Simple Network Management Protocol) MIB(Management Information Base)는 범용성으로 인해 네트워크 전체를 대상으로 하기가 용이한 반면에, 라인 인터페이스 단위에 대한 트래픽 양에 대한 정보만을 제공하므로 보안 측면에서의 트래픽 분석을 하는데 있어 이것만으로는 충분치 않다. RMON(Remote MONitoring) MIB은 네트워크 세그먼트 단위의 트래픽 정보를 이용하며 네트워크 이상 징후를 판단할 수 있는 유용한 통계 자료를 제공하는 반면에, 송신자와 수신자의 연속적인 데이터 흐름을 나타내는 플로우 개념이 없다. 또한 RMON MIB은 데이터 생성을 위해 지나친 부담을 초래할 수 있기 때문에 네트워크 전체를 대상으로 적용하기는 어렵다. 마지막으로 Netflow는 누적된 통계 기능은 제공하지 않으나 제한된 시간 내에 도착하는 IP 패킷들의 흐름(응용 주소 쌍, 호스트 쌍 등) 정보를 제공함으로서 보안 측면에서의 트래픽 분석을 수행함에 있어 유용하게 이용할 수 있다.

2.3 네트워크 이상 징후 분석 시스템 고려사항

본 절에서는 보안 측면에서 효율적인 네트워크 이상징후 분석 시스템을 개발함에 있어 고려해야 할 사항을 기술한다.

2.3.1 네트워크 이상징후 특성 매개변수

보안 측면에서 네트워크 이상징후 현상을 정확히 분석하기 위해서는 네트워크 원시 트래픽 중에서 네트워크 이상 징후를 결정짓는 트래픽 특성 매개변수를 정의하고, 정의된 매개변수의 정상상태를 규명 짓는 일이 무엇보다도 중요하다. 이러한 트래픽 특성 매개변수는 사이버 공격을 분석함으로써 얻을 수 있다. 일반적으로 알려진 매개변수는 다음과 같다.

o CPU 사용량 및 Load 양 :
갑작스런 CPU 사용율이 높거나 갑작스런 패킷량(PPS: Packet Per Second)가 폭주할 경우 비정상적인 트래픽 유발 가능성에 대해 의심하기를 권고하고 있다. 즉, SYN 플로딩과 같은 경우의 공격은 실제 대규모 패킷을 대역폭을 점유하기보다는 작은 패킷들로 다량으로 보내어 PPS를 높이고, 결과적으로 라우팅 장비나 대상 시스템에서 많은 CPU를 소모하게 되는 경우가 많다.
o 패킷 분포 :
SYN 플로딩과 같은 경우의 공격은 일반적인 분포와는 달리 특정 크기의 패킷 특히, 아주 작거나 큰 패킷의 분포가 비정상적으로 늘어나는 것을 확인할 수 있다.
o 응용 프로토콜 분포 :
네트워크 서비스의 종류별 분포를 나타내는 것으로, 공격 발생 시 특정 서비스나 기타 알려지지 않는 서비스의 흐름이 급격하게 증가되는 현상을 보인다.
o 네트워크 플로우 정보 :
네트워크 플로우 정보는 특정 호스트에 대한 플로우 집중 현상, 위장 주소를 이용하는 플로우에 대한 모니터링, 네트워크 상에서의 플로우 급증 현상 등을 파악하는데 이용될 수 있다.

하지만 보다 궁극적인 원인을 분석하기 위해서는 단일 매개변수보다는 여러 개의 조합으로 구성된 매개변수들을 이용하여 정확도를 높여야

할 것이며, 현시점에서 이상징후를 확인하기 위한 알려지지 않은 매개변수들을 지속적으로 정의해야 할 것이다.

2.3.2. 트래픽 데이터 축약(Reduction)

최근의 공격에서 알려져 있듯이 공격으로 인한 네트워크 성능의 이상을 가져오기까지 걸리는 시간은 10분 이내이다. 따라서 네트워크 이상 징후를 분석함에 있어 상세 분석보다는 빠른 분석을 통해 우선적으로 유해 트래픽 가능성 정도에 대해 판단하고, 유해 가능성이 높은 트래픽 부분에 대해서 상세 한 분석을 수행하는 것이 효율적이다. 수집 트래픽 데이터의 축약을 수행함으로써 상세 분석 대상 트래픽의 범위를 좁힐 필요가 있고, 이의 성능 및 가능성이 중요시된다.

2.3.3. 트래픽 유발의 공격 징후 탐지

실제로 과다 트래픽을 유발하는 새로운 공격 방식이 출현하거나 기존의 공격 방식에 의한 부수적인 효과에 의해 네트워크의 성능을 침해할 수준의 트래픽이 발생하는 경우에 탐지를 통한 적절한 대응이 필요하다. 이를 위해 보다 정확한 네트워크의 정상적인 행위에 대한 프로파일의 구축이 요구된다.

2.3.4 네트워크 차원의 이상징후 분석

향후, 더욱 다양한 보안 취약성을 이용해서 다양한 전파경로를 통한 시스템 공격이 전개될 것으로 예상된다. 이와 같이 복잡한 공격방식을 사용하는 경우 단일 보안솔루션에서 분석한 이상징후는 전체 네트워크 이상징후에는 영향이 미치지 않을 수 있으며, 역으로 매우 크게 영향을 미칠 수 있다. 따라서 전체 네트워크 차원에서의 비정상적인 행위에 대한 분석은 단일 보안솔루션에서의 분석과 함께 연계하여 수행되어야 할 것이다.

Ⅲ. 트래픽 분석 모형

본 장에서는 네트워크의 안정적인 운영을 위해 기존에 연구되었던 트래픽 분석 동향에 대해 조사, 정리한다.

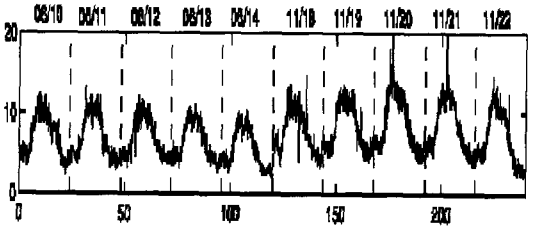
3.1 자기회기(AR: Auto-Regressive) 모형

자기회귀모형은 시계열 자체에 대한 회귀형태를 취하는 모형을 가리킨다.

3.1.1 ANOVA 분석을 통한 AR(2) 모형의 적용

네트워크 기반의 서비스들이 급속히 증가함에 따라, 관리자들이 서비스 문제를 빨리 파악하고 해결하는 것에 대한 중요성이 커지고 있다. 해당 연구에서는 이 모형을 서비스 관리 측면에서 웹 서버 상의 초당 http 동작 횟수를 대상으로 한계치 위반에 대한 예측에 적용함으로써 이상 상태에 대한 판별 가능 여부를 수행하였다.

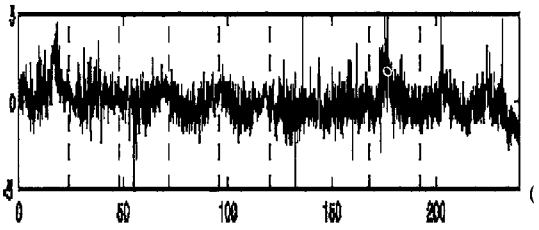
네트워크 상의 트래픽은 아래의 그림에서 알 수 있듯이 시간에 따라 변화하는 비정류(non-stationary)적인 특성을 가지고 있다.



(그림 1) 네트워크 트래픽 데이터

그러나 분산분석(ANalysis Of VAriance : ANOVA)을 사용하면, 트래픽의 월별 요인, 요일별 요인, 시간대별 요인을 파악할 수 있다. 원

래의 데이터에서 이들 요인을 제거하면 아래와 같은 정류적인 특성을 가진 데이터를 얻을 수 있다.



(그림 2) 비정류성을 제거한 트래픽 데이터

이 자료는 시간에 따른 평균의 변동이 없는 stationary한 시계열 자료이기 때문에 AR 모형을 적용할 수 있다. 위의 자료의 자기상관함수(Auto Correlation Function : ACF)를 계산한 결과를 정형화하기 위해서는 2차 자기회귀(AR(2)) 모형이 적합하다는 것을 알 수 있다. 2차 자기회귀모형의 계수는 쉽게 추정될 수 있다.

위의 AR(2) 모형을 사용하여, 향후의 자료를 예측할 수 있다. 또한 미래의 측정치가 한계치를 벗어날 확률을 계산할 수 있다. 그러나 분산분석을 수행하기 위해서는 많은 양의 훈련 데이터를 필요로 한다.

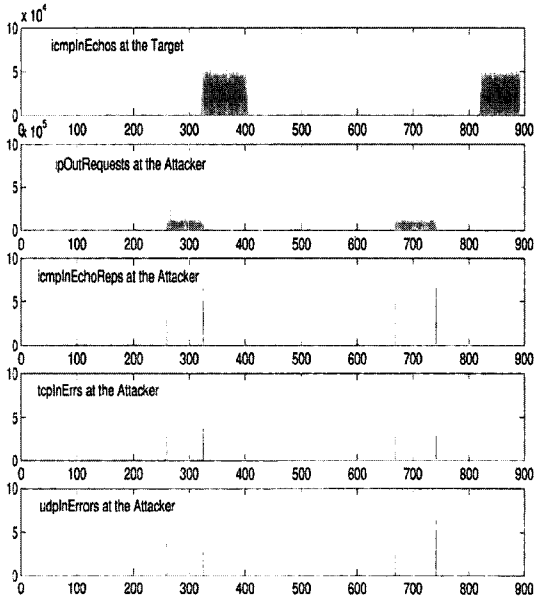
3.1.2 Kalman Filter를 적용한 AR(2) 모형의 적용

앞서 살펴본 바와 같이 비정류성을 가진 트래픽에서 월별 요인, 요일별 요인, 시간대별 요인을 파악하여 제거하면 정류적인 특성을 가지는 데이터를 얻을 수 있다. 그러나 분산분석을 통하여 트래픽의 비정류성을 제거하기 위해서는 많은 양의 사전 데이터가 필요하다. 필요한 데이터의 양을 최소화하기 위하여 Kalman 필터를 사용할 수 있다. Kalman 필터를 사용하면 분산분석 기법을 적용한 것과 마찬가지로 요일별 요인, 시간대별 요인들을 제거할 수 있고, 이렇게 구한 정류적인 데이터를 AR(2) 모형에 적용시켜 미래에 대해서 예측을 할 수 있다.

이 모형은 요일별 요인과 시간대별 요인을 제거하였으나, 긴 시간에 걸쳐서 나타나는 평균 변화를 고려하기는 어렵다. 이러한 장기적인 변동을 고려하기 위하여, 일반화된 우도비(Generalized Likelihood Ratio)를 사용하여 자료에 전체적인 변동이 발생했는지 탐지할 수 있다. 만약에 자료에 변동이 있을 경우에는 AR(2) 모형의 파라미터를 다시 추정하여 예측을 수행한다.

3.2 MIB 변수를 이용한 DDoS 사전감지

SNMP의 MIB의 인터페이스 그룹 변수를 이용하여 DDoS(Distributed Denial of Service) 공격을 조기에 감지하는 것이 가능한가에 대한 연구를 수행한 부분이다. TFN2K Ping 플로딩 공격의 경우, 아래의 그림을 보면 공격자가 실재로 공격을 수행하기 이전에도 MIB 변수가 크게 증가한다.

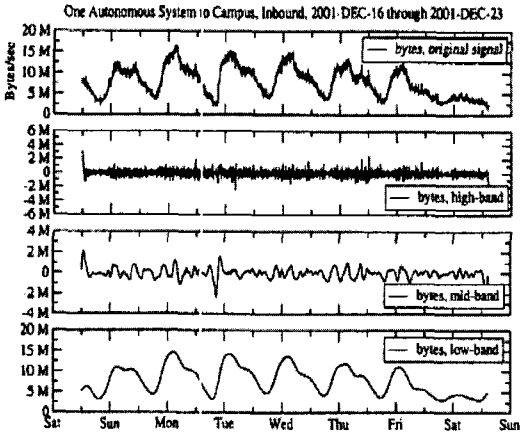


(그림 3) TFN2K 공격에 따른 MIB변수

즉 실제 공격 이전에 공격의 사전징후를 발견할 수 있고, MIB 변수의 시계열 자료에 Granger Causality Test를 적용하면 이러한 공격의 징후가 어떤 MIB 변수에 나타나는지 파악할 수 있다. 공격의 징후가 나타나는 MIB 변수의 갑작스러운 증가를 감지하여 DDoS 공격의 발생을 사전에 예측하여 피해를 최소화할 수 있다.

3.3 웨이블릿(wavelet) 분석

주어진 어플리케이션에 적합한 wavelet 변형을 선택하는 것에 크게 좌우된다.



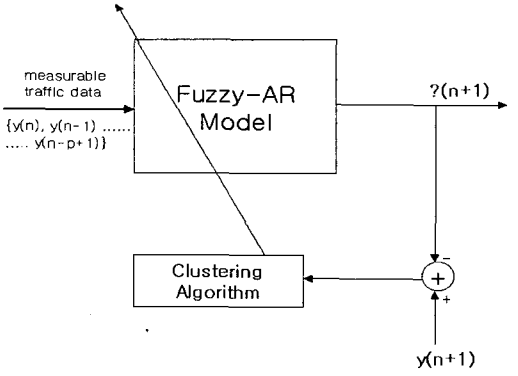
(그림 4) Aggregate byte traffic for a typical week & high/mid/low decomposition

위의 그림의 예는 Wavelet에 의해 원래 신호를 시간 주기로 필터링한 새로운 신호들을 나타낸다. 이 예에서 볼 수 있듯이 각각의 필터링된 신호들을 살펴봄으로써 단기적인 비정상 트래픽과 장기적인 비정상 트래픽이 원래 신호에 비해 더욱 명확하게 구분될 수 있다. 이러한 wavelet 분석은 비정상적인 트래픽을 구분하는데 좋은 성능을 발휘할 수 있으며 단시간의 비정상과 장시간의 비정상을 구분해내는데 효과적으로 쓰일 수 있다.

3.4 퍼지-자기회기 (Fuzzy-AR) 모형

멀티미디어 통신 서비스가 다양해지고 네트워크의 스위치, 다중 송신 기술이 발달하면서 고속 네트워크 트래픽은 점점 비선형적이고 비정류적인 특성을 지닌 프로세스가 되어 가고 있다. Fuzzy-AR 모형은 입력 데이터의 비선형적 대응을 통해서 이러한 트래픽 특성을 정확히 묘사할 수 있는 기법이다.

Fuzzy-AR 모형은 비선형적이고 시간에 따라 변동하는 프로세스를 퍼지 클러스터링 기법을 이용하여 선형의 국소적 AR 프로세스의 결합으로 근사화한다. 이러한 클러스터링 알고리즘은 두 단계의 조율(거친 조율, 미세한 조율)을 통해서 Fuzzy-AR 모형이 실제 트래픽 데이터를 정확히 반영하도록 구성된다.



(그림 5) Fuzzy-AR 트래픽 모델링

Fuzzy-AR 트래픽 모델은 다음의 네 가지 구성요소로 이루어진다.

- Rules :
  트래픽 프로세스를 선형 AR 모형들로 세분하는 퍼지 규칙
- Fuzzifier :
  입력 데이터의 가능성 정도와 그들의 membership function을 고려하여 입력 값들

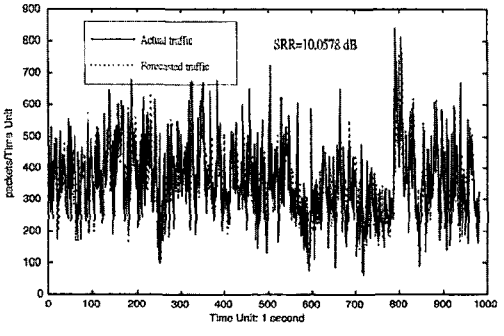
을 퍼지 집합에 대응시킨다.

o Defuzzifier :
출력 퍼지 집합을 출력 데이터에 대응시킨다.

o 퍼지 추론 엔진 :
퍼지 규칙을 통하여 출력 퍼지 집합을 결정하기 위한 추론을 행한다.

퍼지 규칙이나 AR 모형의 계수의 증가는 계산의 복잡성을 증가시키지만 보다 정확한 모델을 제시한다. 따라서 이러한 파라미터를 적절하게 결정하는 것은 Fuzzy-AR 모형의 성능에 중요한 영향을 미친다.

Fuzzy-AR 모형은 예측 기반 모델링으로써 현재의 트래픽 정보, 과거의 트래픽 정보, 트래픽에 대한 예측 정보를 결합하여 밀집된 트래픽 환경에서 신뢰성 있고 정확한 예측을 수행한다. Fuzzy-AR 모형은 다른 모형에 비해 실제 트래픽에 대해 뛰어난 예측 능력을 보여 주는데 아래 그림에 그 예를 보였다.



(그림 6) Fuzzy-AR model을 적용한 경우 실제 트래픽과 예측 트래픽

네트워크 기반의 서비스들이 급속히 증가함에 따라, 관리자들이 서비스 문제를 빨리 파악하고 해결하는 것에 대한 중요성이 커지고 있다. 해당 연구에서는 이 모형을 서비스 관리 측면에서 웹 서버 상의 초당 http 동작 횟수를 대상으로 한계치 위반에 대한 예측에 적용함으로써 이상 상태에 대한 판별 가능 여부를 수행하였다.

네트워크 상의 트래픽은 아래의 그림에서 알 수 있듯이 시간에 따라 변화하는 비정류적인 특성을 가지고 있다.

V. 결 론

최근 네트워크 상의 사이버 공격의 특징은 특정 시스템을 대상으로 하는 것이 아니라, 의도하였든 그렇지 않든 네트워크 서비스 전체의 마비를 가져올 수 있는 공격이 출현하고 있다는 점이다. 이들 공격은 네트워크 상에서 처리할 수 있는 적정 수준 이상의 트래픽을 과도하게 발생시킴으로써 네트워크의 부하를 급격히 증가시켜 합법적인 사용자의 네트워크 서비스 이용을 방해하게 된다.

이에 대응하기 위해서는 네트워크 상에 처리되는 정상 트래픽을 모델링하고, 이를 현재의 트래픽 데이터와 비교 분석함으로써 이상 트래픽을 감지하고, 감지된 이상 트래픽에 대해서 네트워크 차원에서 대응을 수행하는 방법이 효과적이다. 이로 인해 보안 측면에서 트래픽 분석은 이제 시작되고 있는 단계이다. 이론적으로는 트래픽을 모델링하기 위한 모형들에 대한 연구가 많이 진행되었으나, 이들 중 많은 부분은 네트워크 엔지니어링 관점에서 이루어진 것이어서 네트워크 보안 부문에 바로 적용하기에는 어려움이 많다.

하지만 앞으로 이와 관련된 연구가 진척되고 해결책을 찾아서 이상 트래픽을 사전에 감지하고, 이에 대한 네트워크 수준의 강력한 대응을 수행할 수 있다면, 네트워크의 안정적인 운용을 보장할 수 있을 뿐만 아니라 사이버 공격으로 인한 많은 피해를 줄일 수 있을 것으로 본다.

참고문헌

[1]정현철, 변대용, "트래픽 분석을 통한 서비스 거부공격 추적", http://www.certcc.or.kr, 2003.1.

[2] Bor-Sen Chen, Sen-Chueh Peng, Ku-Chen Wang, Traffic Modeling, *Prediction, and Congestion Control* for High-Speed Networks: A Fuzzy AR Approach, IEEE Transactions on Fuzzy Systems, Vol. 8, No. 5, Oct. 2000

[3]Cisco Inc., Cisco's IOS Netflow Feature, http://www.cisco.com/warp/public/732/netflow

[4]David Moore, Gffrey M. Voelker and Stefan Savage, Inferring Internet Denial-of-Service Activity, http://www.dante.net/pubs/dip/42/42.html

[5]Denial of Service and Emerging Backbone Threats, Arbor Networks, June 12, 2002

[6]Internet Performance Measurement and Analysis Project Home Page, http://www.merit.edu/ipma

[7] Hong Fei, Wu Zhimei, A Novel Traffic Model Based on Wavelet Analysis, Proceedings of ICCT2003

[8]Joao B. D. Cabrera, Lundy Lewis, et. al., Proactive Intrusion Detection and Distributed of Denial of Service, Journal of Network and System Management, Vol. 10, No. 2, June 2002

[9]Paul Barford, Jeffery Kline, et. al., A signal analysis of network traffic anomalies - Proceedings of ACM IMW02

[10]Sheng Ma, Chuanayi Ji, Modeling Heterogeneous Network Traffic in Wavelet Domain, IEEE/ACM Transactions on Networking, Vol. 9, No. 5, Oct. 2001

[11]Yantai Shu, Yu Liu, et. al., Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models

나 중 찬

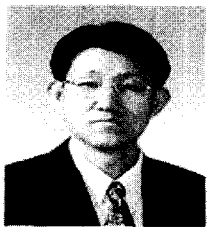
1986년 2월 : 충남대학교 계산통계학과 졸업

1989년 2월 : 숭실대학교 전자계산학과 석사

1998년 3월~현재 : 충남대학교 검퓨터과학과 박사과정

1989년 2월~현재 : 한국전자통신연구원 능동보안기술연구팀 팀장

<관심분야> 네트워크 보안, 실시간 시스템



김 영 국

1985년 2월: 서울대학교 계산통계학과 학사

1987년 2월: 서울대학교 계산통계학과 석사

1995년 5월: University of Virginia, Computer Science 박사

1996년 3월 ~ 현재: 충남대학교 전기정보통신공학부 교수

<관심분야> 실시간 정보시스템, 이동정보시스템, 전자상거래 시스템