침입탐지를 위한 네트웍 트래픽 데이터 도시

곽미라, 조동섭
이화여자대학교 과학기술대학원 컴퓨터학과

Graphical Representation of Network Traffic Data for Intrusion Detection

Mira Kwak, Dong-sub Cho
Dept. of Computer Science and Engineering Ewha Womans University

요 약

침입 탐지를 위하여 수집되는 네트웍 트래픽은 보통 분석 처리 프로그램으로 입력되기 위해 수치적으로 표현된다. 이러한 데이터로부터 그 가운데 드러나는 경향을 한 눈에 발견하는 데에는 어려움이 있어, 이에 대해 프로토콜, 서비스 및 세션 등을 기준으로 분류하는 처리를 수행한 결과를 바탕으로 세세한 분석과정을 거치는 것이 일반적이다. 네트웍 트래픽 데이터를 도시하여 그 추이를 직관적으로 살필 수 있게 한다면 여러 기준에 따라 분류된 각 트래픽이 가지는 특징을 쉽게 발견할 수 있다. 이러한 트래픽 추이와 특징 파악의 용이함은 트래픽에서 비정상적인 부분을 식별해내는 것을 쉽게 한다. 이것은 시스템 관리자가 현재 해당 시스템에 설치되어 작동되고 있는 침입탐지 시스템이나 방화벽 시스템에 대해 독립적으로 편리하게 네트웍 트래픽의 특징을 살피고 이상을 발견할 수 있도록 하며, 경고되거나 차단되지 않은 이상에 대해 신속히 대응할 기회를 준다. 이에 본 연구에서는 네트웍 트래픽들의 특징을 설명할 수 있는 요소들을 조합하여 표현함으로써 네트웍 트래픽의 특징과 이상 파악에 편리한 데이터 도시 방법을 제안한다.

1. 서론

인터넷 사용이 증가함에 따라 침입 네트웍 침입 탐지에 관한 연구는 그 중요성이 인식되어 활발히 진행되고 있다. 침입을 탐지하는 첫 단계는 네트웍 트래픽으로부터 이상 유무를 발견하는 것이고 이를 위해서는 이상을 설명할 수 있는 속성의 관찰이 선행되어야 한다. 본 연구에서는 이러한 속성을 시각적으로 관찰하는 방법을 제안한다.

침입 탐지를 위하여 수집되는 네트웍 트래픽은 보통 분석 처리 프로그램으로 입력되기 위해 수치적으로 표현된다. 이러한 데이터로부터 그 가운데 드러나는 경향을 한 눈에 발견하는 데에는 어려움이 있어, 이에 대해 프로토콜, 서비스 및 세션 등을 기준으로 분류하는 처리를 수행한 결과를 바탕으로 세세한 분석과정을 거치는 것이 일반적이다. 네트웍 트래픽 데이터를 도시하여 그 추이를 직관적으로 살필 수 있게 한다면, 서비스 및 프로토콜별 트래픽의 특징, 시간대별 트래픽의 특징과 같이 여러 기준에 따라 분류된 각 트래픽이 가지는 특징을 쉽게 발견할 수 있다. 이러한 트래픽 추이와 특징 파악의 용이함은 트래픽에서 비정상적인 부분을 식별해내는 것을 쉽게 한다. 이것은 시스템 관리자가 현재 해당 시스템에 설치되어 작동되고 있는 침입탐지 시스템이나 방화벽 시스템에 대해 독립적으로 편리하게 네트웍 트래픽의 특징을 살피고 이상을 발견할 수 있도록 하며, 경고되거나 차단되지 않은 이상에 대해서도 보다 신속히 대응할 수 있는 기회를 준다.

이에 본 연구에서는 네트웍 트래픽의 특징과 이상 파악에 편리한 데이터 도시 방법을 제안한다. 본 논문은 다음과 같이 구성된다. 2장에서는 네트웍 트래픽 분석을 위한 기존의 데이터 가시화 연구를 살펴본다. 3장에서는 다양한 기준으로 분류된 네트웍 트래픽에서 특징을 설명할 수 있는 속성 요소들을 찾고 그것을 바탕으로 직관적인 데이터 특징 파악을 가능하게 하는 데이터 도시 방법을 제안한다. 4장에서는 3장에

※ 이 논문은 2003년도 두뇌한국21사업에 의하여 지원되었음.

서 제안한 방법을 샘플 데이터에 적용하여, 그 방법이 네트웍 트래픽의 추이와 이상 유무의 파악하는데 얼마나 효과적인지 실험한다. 5장에서는 본 논문에서 논한 주제와 제안한 방법, 실험 결과에 대해 정리하고 향후 연구과제를 생각해봄으로써 논문을 맺는다.

2. 네트웍 트래픽 데이터의 도시에 관한 기존 연구

네트웍 모니터링 도구들 중 관찰되는 정보를 그래프 형태로 사용자에게 보이는 것들이 있다. 이들은 대개 관찰 대상 각각에 대해 시간 축에 따라 변화하는 관찰 값의 1차원 그래프를 보이는 것에 그치고 있다. 그림 1은 이러한 그래프의 예이다. 이와 같은 그래프는 기본적인 형태로, 시간에 따른 TCP 트래픽의 유량과 같이 한 정보의 시간에 따른 변화를 관찰하는 데에는 도움이 된다. 그러나 연관관계를 명확히 알지 못하는 여러 정보들로부터 연관성 유무를 찾아내기 위한 정보의 조합에 대한 관찰을 가능하게 하기에 무리가 따른다.

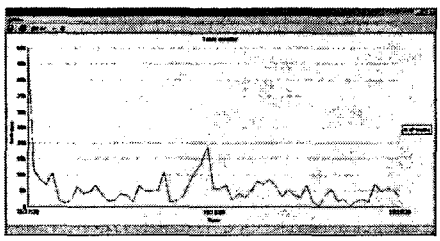

그림 1. 네트웍 트래픽에 관한 1차원 그래프 예

이보다 발전된 그래픽 표현의 예도 몇 가지 찾아볼 수 있다. 그림 2는 T. Dunigan, G. Ostrouchov가 제안한 침입 탐지를 위한 네트웍 트래픽의 성격에 따른 분류 연구에서 나타난 그래프이다[4]. 이 그래프는 그림 1의 예를 들어 위에서 이야기한 기본 형태의 그래프보다 많은 정보를 한 번에 보이며, 정보를 조합하여 관찰함으로써 기본 데이터에서는 드러나지 않던 정보를 발견하도록 하는데 보다 유리하다. 그러나 이 그래프에 익숙하지 않은 사람이 직관적으로 그 내용을 이해하는 데에는 불편이 있을 수 있다.

3. 제안하는 방법

3.1. 네트웍 트래픽의 속성 요소

우리는 네트웍 트래픽으로부터 침입 탐지에 유용한 정보를 얻고자 한다. 이를 위해서는 침입이 있는 경우

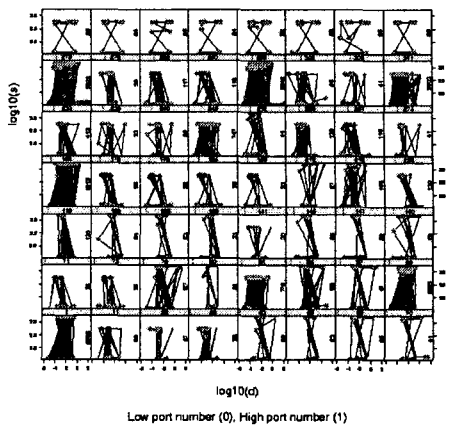

그림 2. FTP 데이터 전송 트래픽에 관한 그래프

정상 상태의 트래픽과 비교하여 그 값이나 값의 변화가 다르게 나타나는 요소들을 찾아내는 작업이 선행되어야 한다. 본 연구에서는 tcpdump라는 소프트웨어 도구를 사용하여 실험 대상이 되는 호스트 시스템에 들어오고 나가는 모든 네트웍 트래픽의 패킷의 정보를 거르지 않고 수집하였다. 그 결과 네트웍 트래픽을 구성하는 패킷에 대한 정보는 아래와 같았다.

- 해당 패킷이 tcpdump에 의해 관찰된 시간
- 패킷의 출발 호스트의 IP 주소
- 패킷의 출발 호스트가 해당 연결에서 사용한 포트
- 패킷의 목적 호스트의 IP 주소
- 패킷의 목적 호스트가 해당 연결에서 사용한 포트
- 전달된 패킷의 크기
- 패킷의 분할 정보
- 패킷의 순서 번호
- 그 밖의 옵션

이러한 정보를 바탕으로 이미 그 안에 발생한 침입에 대한 정보를 알고 있는 네트웍 트래픽 데이터와 정상 네트웍 트래픽 데이터를 비교, 분석하였다. 그 결과 다음과 같은 정보를 통해 침입 여부를 판단하고, 서비스나 프로토콜에 따른 특징을 파악할 수 있음을 알았다.

양 호스트 사이의 모든 연결에 대해
- 호스트 a가 사용한 모든 포트 수
- 호스트 b가 사용한 모든 포트 수

양 호스트 사이에 맺어진 각 연결에 대해
- 총 전달된 패킷 수
- 총 전달된 패킷들 중 방향 A 패킷들의 비율
- 방향 A 패킷들의 평균 크기
- 방향 A 패킷들의 평균 출발시각 간격
- 총 전달된 패킷들 중 방향 B 패킷들의 비율
- 방향 B 패킷들의 평균 크기

- 방향 B 패킷들의 평균 출발시각 간격
- 연속으로 방향 A 패킷이 전달된 비율
- 방향 A 패킷 후 방향 B로 패킷이 전달된 비율
- 방향 B 패킷 후 방향 A로 패킷이 전달된 비율
- 연속으로 방향 B 패킷이 전달된 비율

3.2. 네트웍 트래픽의 효율적 도시

본 절에서는 그 패킷들에 관한 모든 정보를 수집하고 구성하여 3.1 절에서 설명한, 특징 및 이상 여부 발견에 효율적인 속성들을 파악하기 쉽도록 네트웍 트래픽 데이터를 도시하는 방법을 보인다. 이는 전체 네트웍 트래픽을 구성하는 각 패킷을 3차원 좌표의 한 선분으로 표현함으로써 이루어진다. 이 때 좌표의 각 축과 그 값의 의미는 다음과 같다.

- x축 : 시간 축. 이 좌표의 값은 tcpdump에 의한 패킷의 타임스탬프 값으로, tcpdump에 의해 해당 패킷이 관찰된 시간을 의미한다. 이는 1970년 1월 1일 0시 정각으로부터 경과된 값으로 표현되며, 단위는 $1/10^6$초이다.
- y축 : 호스트 축. 호스트의 IP주소인, 점('.')으로 구분된 네 개의 십진수 중 첫 번째 수에는 10^9을, 두 번째 수에는 10^6을, 세 번째 수에는 10^3을 곱한 후 네 개의 숫자를 모두 더하고, 내부 호스트인 경우 같은 크기의 음수로 변환하고 외부 호스트인 경우 그대로 취하여 이 좌표의 값으로 삼는다.
- z축 : 포트 축. 연결에 사용된 포트 값을 곧 이 좌표의 값으로 삼는다.

이러한 좌표 위에서 한 패킷을 나타내는 선분은 다음과 같은 속성을 가지고 그려진다.

- 시작점 : 패킷을 송신한 호스트와 사용된 포트, 그 패킷이 관찰된 시간으로 이루어진 점
- 끝점 : 패킷을 수신한 호스트와 사용된 포트, 그 패킷이 관찰된 시간으로 이루어진 점
- 선분의 굵기 : 전송된 패킷의 크기. 여기에서는 바이트 단위의 크기 값을 100으로 나누어 사용한다.
- 선분의 색 : 여러 연결들을 함께 나타낼 때는 각 연결마다 고유한 색을 가지고, 연결 내의 패킷들은 모두 같은 색을 가진다. 한 연결만 나타낼 때는 각 패킷 선분에 각기 다른 색을 준다. 이 때 색은 임의로 결정한다.
- 선분의 모양 : 연결을 시작한 호스트로부터 출발한 패킷의 선분은 '—' 모양, 그 반대 방향으로 전달되는 패킷의 선분은 '-·'모양을 가진다.
- 선분의 설명(label) : 시퀀스 번호, 프로토콜 이름, 윈도 크기, 플래그 내용 등을 요청에 따라 보인다.

호스트 주소와 포트 번호로 이루어진 호스트 정보의 쌍으로 연결들을 구분하여, 전체 트래픽 및 연결별 트래픽, 연관된 연결의 모든 트래픽, 각 패킷 등 다양한 수준에서 데이터를 관찰할 수 있도록 하였다.

4. 실험

본 장에서는 3장에서 설명한 방법을 샘플 데이터에 적용한 예를 보인다. 사용한 데이터는 MIT Lincoln 연구소에서 DARPA의 침입 탐지 시스템 평가 프로젝트를 위해 준비한 분석 대상 데이터 중 일부이다. 제안한 방법을 통해, 사용된 네트웍 트래픽 데이터로부터 모든 연결들을 구분하고 사용자로 하여금 각 연결마다, 연관된 연결들을 묶어서, 임의의 연결들을 묶어서, 또한 모든 연결들을 함께 3차원 좌표에서 볼 수 있었다.

그림 3부터 그림 8까지의 그림들은 기본적인 그래프의 예이다. 그림 3, 4, 5는 비슷한 시간대에 일어난 100개의 연결들에 관한 일반적인 그래프의 예이고, 그림 6, 7, 8은 한 연결에 관한 일반적인 그래프의 예이다.

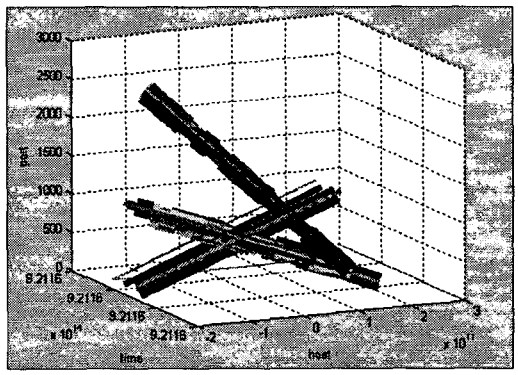

그림 3. 입체 관찰

그림 3은 100개 연결을 구성하는 패킷 선분들에 대한 3차원적인 입체 조망이다. 이러한 관찰을 통해 해당 연결들이 네트웍의 어떠한 내/외부 호스트들 사이에 어떠한 포트를 통해 이루어진 것이며 어느 정도 오랫동안 지속되었는지 등에 관한 대략의 파악이 가능하다. 관심을 끄는 연결에 대해서는 선택 기능을 통해 자세히 볼 수 있다. 그림 4와 5는 그림 3과 같은 내용을 다른 위치에서 본 그림이다. 다양한 각도에서 다른 관심의 기준으로 데이터를 관찰할 수 있어, 시점을 바꾸었을 때 눈에 띄는 속성에서 특이한 점을 발견하는 것이 용이하다. 이와 같이 차원을 하나 줄인 관찰에서는 개별 패킷에 대한 특징이 보다 눈에 띈다. 그림 6은 하나의 연결을 구성하는 패킷들 사이의 시간 간격을 나타낸 그래프이다. 이를 통해 여러 공격의 징후를 발견하는 데 필요한 속성인 패킷 간 시간 차

이를 관찰할 수 있다. 그림 7은 패킷의 크기를 높이로 그 방향을 음/양으로 나타낸 그래프로, 전달된 패킷의 크기와 그 방향의 변화를 쉽게 파악할 수 있도록 한다.

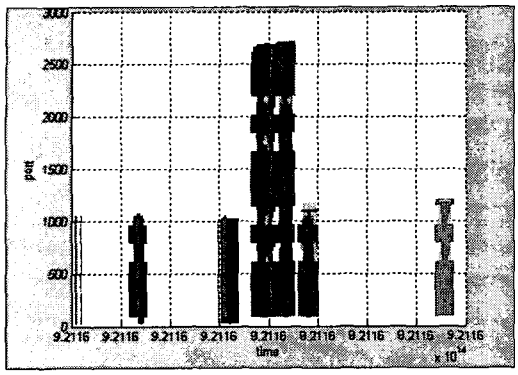

그림 4. 시간과 포트 관찰

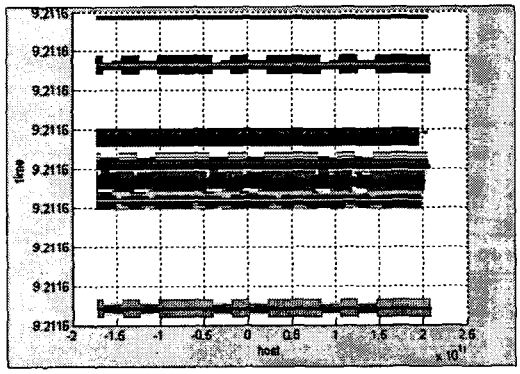

그림 5. 호스트와 시간 관찰

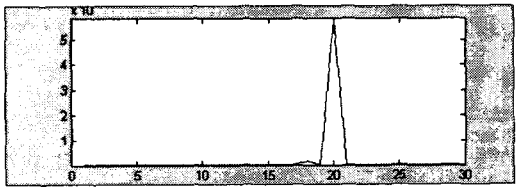

그림 6. 연속된 패킷 사이의 시간 간격

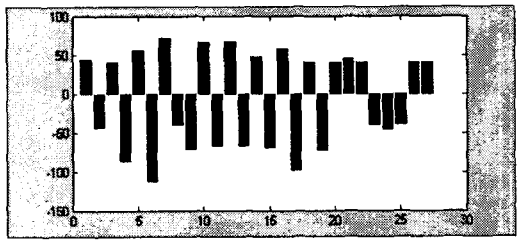

그림 7. 패킷 크기와 방향 그래프

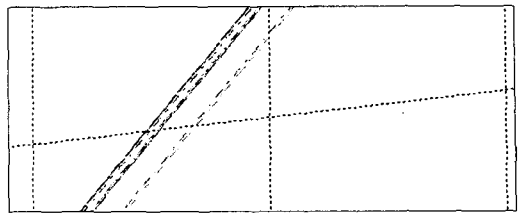

그림 8. 한 패킷의 확대 관찰

그림 8은 한 패킷에 대한 그림 1과 같은 입체 조망을 확대한 것이다. 이러한 확대를 통해 개별 패킷을 자세히 볼 수 있다.

5. 결론

본 연구에서는 네트웍 트래픽 데이터를 도시하여 시각적으로 그 특징과 특이점을 쉽게 발견할 수 있도록 하는 방법을 제안하였다. 4장에서 사용한 데이터를 실험하여 패킷의 특징 및 연결의 특징을 함께 관찰하여 쉽게 특징과 침입 여부를 직관적으로 찾아낼 수 있음을 확인하였다. 현재는 이미 수집된 데이터를 입력으로 하였으나, 진행되고 있는 네트웍 트래픽을 입력으로 하여 실시간으로 이와 같은 도시를 가능하게 하며, 침입 탐지 기능과 연동하여 침입 트래픽에 대한 상세한 그래프 정보를 함께 남기도록 한다면 더욱 유용할 것이다.

[참고문헌]

[1] K. Claffy, G. Polyzos and H. W. Braun, "A Parameterizable methodology for Internet traffic flow profiling," Mar 1995, IEEE JSAC Special Issue on the Global Internet

[2] K. Claffy, "Internet traffic characterization," Ph.D. Dissertation, UC, San Diego, June 1994

[3] M. Luoma, M. Ilvesmäki and M. Peuhkuri, "Source characteristics for traffic classification in Differentiated Services type of networks," Voice, Video & Data Communications '99 in Boston, MA, USA.

[4] T. Dunigan, G. Ostrouchov, "Flow characterization for intrusion detection," Oak ridge national laboratory, Technical report TM-2000, Nov. 27, 2000