의사결정트리를 이용한 효과적인 호스트 기반의 서비스 거부 공격 탐지에 관한 연구

두선정*, 황현진*, 조재익**, 김낙훈*
*동덕여자대학교 컴퓨터학과
**고려대학교 정보경영공학전문대학원
e-mail : 20051278@dongduk.ac.kr

Study on the Host-based Detection for DoS Attack using the Decision Tree Method

Sun-Jeong Doo*, Hyun-Jin Hwang*, Jae-Ik Cho**, Nak-Hoon Kim*
*Dept of Computer Science, Dong-Duk Women's University
**Graduate School of Information Management and Security, Korea University

요 약

서비스 거부 공격은 현재의 서비스를 불법적으로 중단시켜 여러 사용자의 접근을 제한하는 공격 방법이다. 이러한 서비스 거부 공격 탐지 기법에 관한 연구가 활발히 진행되어 왔지만 기존의 네트워크 기반의 공격 탐지 기법은 많은 문제점을 낳고 있다. 따라서 본 논문에서는 기존의 탐지기법의 취약점을 보완하기 위해 호스트기반의 데이터를 이용해 더 효과적으로 서비스 거부 공격을 탐지할 수 있는 방법을 제안한다.

1. 서 론

네트워크 환경의 발달에 따라 많은 온라인 컨텐츠의 개발과 더불어 다양한 시스템으로 네트워크가 확장되고 있다. 그러나 네트워크의 확장과 함께 다양한 공격이 발생되고 있는 것 또한 사실이다. 근래에 가장 문제가 되는 공격으로 서비스 거부 공격이 있다.

서비스 거부 공격은 현재의 서비스를 불법적으로 중단시켜 여러 사용자의 접근을 제한하는 공격 방법으로, 공격자의 추적이 힘들고 공격 차단 방법이 정확하지 않아 이로 인한 많은 문제점을 일으키는 공격이다. 또한 일시적으로 대규모의 네트워크를 마비 시킬 수 있기 때문에 국가 보안의 차원에서도 반드시 차단 되어야 하는 공격이다[1].

본 연구는 이러한 많은 문제점을 발생시키는 서비스 거부 공격을 호스트 기반에서 정확히 탐지하기 위하여 커널 기반 데이터를 이용해 탐지하였다. 기존의 패킷 기반 보안 시스템의 경우 일부 네트워크에 대한 서비스 거부 공격을 탐지 하기 때문에 내부의 공격이나 분산된 서비스 거부 공격을 탐지하는데 한계가 있다. 그러나 본 논문에서 제안하는 방법의 경우 호스트 기반 탐지를 위해 커널 기반 데이터를 이용하기 때문에 보다 실질적인 시스템 사용에 대한 분석을 통한 정확한 탐지가 가능하다.

본 논문의 구성은 다음과 같다. 2 장에서 서비스 거부 공격의 일반적인 내용과 기존의 탐지 방법을 설명하고 3 장에서 본 논문에서 연구된 서비스 거부 공격의 탐지 기법을 설명하며, 4 장에서는 3 장에서 제안한 방법을 이용해 실험하여 제안 방법의 효과 및 결과를 확인 한다. 마지막으로 5 장에서 결론을 맺는다.

2. 관련 연구

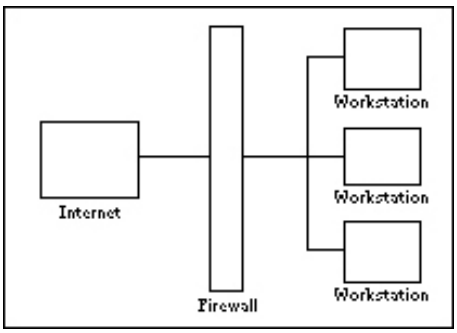
현재 네트워크 공격 기술은 기존의 여러 가지 악성 코드를 이용한 방법을 확장하여 봇넷을 이용한 원격지 공격, 다중 시스템을 이용한 대량의 서비스 거부 공격, 이를 확장한 분산된 원격지를 이용한 다중 서비스 거부 공격으로 확장되고 있다. 또한 이러한 서비스 거부 공격은 가장 위험한 공격으로 선정되고 있다[2]. 서비스 거부 공격은 스크립트 키즈로부터 단순한 호기심 위주의 공격에서 상업적 목적을 위한 고급 해커로 구성된 공격으로 발전되고 있으며, 다른 국가 혹은 네트워크 내부로부터 공격 되는 다양한 공격지로 확장되고 있다.

그러나 서비스 공격을 탐지 하기 위한 방법으로서 기존의 침입 탐지 시스템이나 침입 차단 시스템에 간단한 확률 탐지 방법이 이용되기 때문에 오탐지율이 높고 내부 개개의 호스트에서는 내부 공격에 대해 탐지가 불가능한 단점이 있다.[3]

본 관련 연구에서는 서비스 거부 공격에 대한 기존의 침입 탐지 및 침입 차단 기술을 살펴 본다.

2.1 기존의 침입 탐지 및 침입 차단 방법

기존의 침입 탐지 및 침입 차단 시스템의 경우 일반적으로 네트워크 패킷 데이터에서 데이터그램 부분을 분리하여 보안 시스템에 구성되어 있는 공격 종류별 시그너쳐와 비교하여 탐지하는 방법을 이용한다[4].



(그림 1) 침입 탐지 시스템의 구조

또한 위의 그림 1 과 같이 일반적으로 침입 탐지 및 침입 차단 시스템은 외부 네트워크로부터 유입되는 데이터를 최초의 단말에서 확인하여 탐지하는 방법을 사용한다.[5] 이러한 구조 및 탐지 방법은 여러 가지 문제점이 있다. 첫 번째로 보안 시스템의 위치가 공격자에 의해서 확보되면 일반적으로 시스템 거부 공격을 통해 보안 시스템을 공격자가 장악하며, 대량의 패킷 데이터 유입으로 보안 시스템은 기본적으로 유입 패킷을 검사하지 않고 내부 네트워크로 전송한다. 즉, 내부의 네트워크가 외부에 아무런 보안 시스템 없이 노출된다. 두 번째로 대규모의 봇넷을 이용한 내부 네트워크를 대상으로 하는 서비스 거부 공격이 가능하다. 현재의 침입 탐지 시스템에서는 단순히 동일 발신지에서 수신되는 패킷 데이터의 수 혹은 양을 확인하여 차단하는 방법을 이용하는데 공격자는 여러 개의 악성 코드에 감염된 호스트를 이용하여 여러 개의 내부 네트워크 호스트를 공격하는 방법을 이용한다. 몇 개의 호스트 공격은 차단 될 수 있지만 앞서 설명된 탐지 방법을 이용한다면 지속적인 다중 공격을 차단할 방법이 없다. 마지막으로 내부 네트워크의 일정 호스트가 악성 코드에 감염이 되거나 혹은 실제 공격자가 내부에서 공격하는 서비스 거부 공격은 탐지 할 수 없는 문제점이 있다.

이러한 내용을 볼 때에 내부 네트워크에서 호스트 기반 서비스 거부 공격을 탐지하는 것은 매우 중요하며 현재의 보안 시스템에서 사용하고 있는 방법과 더불어 추가적인 호스트 기반 서비스 거부 공격 탐지 방법이 필요하다.

본 논문에서는 커널 기반 데이터를 이용하여 호스트 기반 서비스 거부 공격을 탐지하는 방법을 제안한다.

3. 제안하는 방법

3.1 데이터의 수집

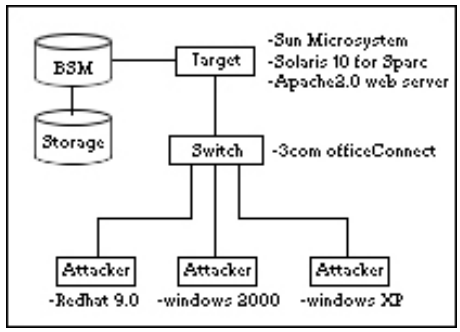
네트워크에서 발생한 서비스 거부 공격을 탐지하는 방법을 실험하기 위하여 고려대학교 내부 네트워크에 구성된 실제 유닉스 기반의 웹 서버 시스템에서 커널 기반 데이터를 수집하였다. 커널 기반 데이터 수집은 선 마이크로시스템의 솔라리스 환경에서 Basic Security Module (BSM) 을 이용하여 수집하였다. 수집은 2 주 동안 해당 시스템에서 수집 되었으며, 일반적인 공개 침입 탐지 시스템인 SNORT 를 이용하여 공격이 없었음을 확인하였다.

수집된 데이터의 상세 내용은 아래 표와 같다.

수집 기간	2008 년 5 월 2 일 ~5 월 16 일
데이터 크기	약 15Gbyte
수집 대상 시스템	Sun Solaris 10 for Sparc, Apache 2.0 환경

<표 1> 수집된 정상 커널 기반 데이터

공격 데이터의 경우 아래의 그림과 같이 가상의 폐쇄 네트워크를 구성하여 수집하였다.



(그림 2) 공격 데이터 수집용 가상 폐쇄 네트워크 구조

가상의 폐쇄 네트워크는 공격이 발생 하였을 때를 가상화 하여 공격의 목적이 되는 시스템의 커널 데이터를 수집하기 위해 구성하였다. 가상의 폐쇄 네트워크에서 두 개의 공격자 시스템은 윈도우시스템에서 Syn 패킷을 대량 발생시켜 공격 목적이 되는 시스템으로 전송하였으며 동시에 공격 목적이 되는 시스템의 BSM 에서 커널 데이터(이하 BSM 데이터)를 기록하였다. 기록된 BSM 데이터는 선 마이크로시스템에서 제공되는 전용 파서 소프트웨어를 이용하여 ASCII 로 변환하였으며 변환된 ASCII 데이터를 이용하여 실험에 이용하였다.

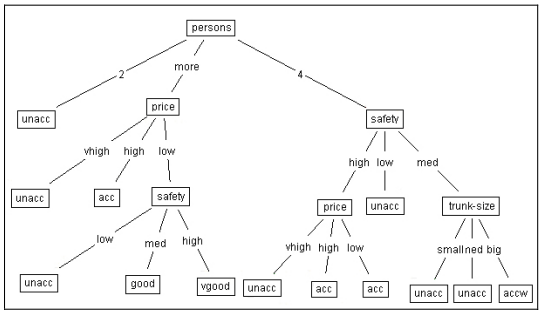
3.2 공격 탐지 방법

본 논문에서는 의사결정트리를 이용하여 공격을 탐지하였다. 의사결정트리(Decision Tree) 기법은 주로 분류 목적으로 사용되는 지식발견(knowledge discovery) 기법이다. 나무 모형은 뿌리에서 가지를 거쳐 잎으로 분화되는 나무의 형태를 취한다. 가지의 나뉨, 즉 분

지는 마디에서 이루어지는데 단순한 결정나무의 각 마디는 “If A then B1. Else B2."의 논리적 구조로 구성되어 있다. 즉 A 의 경우이면 B1 으로 가고 B2 로 가라는 것이다. 즉 Decision Tree 는 과거에 수집된 데이터의 레코드들을 분석하여 이들 사이에 존재하는 패턴을 적절한 속성의 조합으로 분류하여 생성한 플로우차트 형태의 트리 구조를 의미한다.

price	maint cost	doors	persons	trunk size	safety	accep table
vhigh	med	2	2	big	low	unacc
vhigh	vhigh	5more	more	big	high	unacc
vhigh	high	2	2	small	low	unacc
…	…	…	…	…	…	…
low	low	5more	more	med	med	good
low	low	5more	more	med	high	vgood
low	low	5more	more	big	med	good

<표 2> 자동차 구매 결정 데이터 셋



(그림 3) 의사결정트리(Decision Tree) 예제

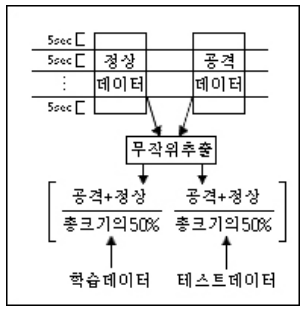
그림 3 은 자동차 구매에 관련된 데이터 셋(표 2)을 이용한 의사결정트리(Decision Tree)를 나타낸다. 이 트리에 나타난 사각형의 글상자들을 ‘노드(Node)’라 하며, 종단 부를 제외한 노드에는 데이터의 특성을 분류할 수 있는 속성이 포함된다. 노드가 나타내는 속성에 따라 분류된 데이터는 실선 형태의 ‘가지(Branch)’를 따라서 분류되어 하위 노드와 연결되며, 하위노드 중 더 이상의 하위 노드를 갖지 않는 종단부의 노드를 ‘잎(Leaf)’이라 하며, 잎에는 최종적인 분류 결과가 표시된다. 트리를 구성하기 위해서는 먼저 정보이론에 따라 어떤 속성을 상위 속성으로 할 것인지 결정해야 하며, 이렇게 결정된 속성에 의하여 데이터들을 분류한 후, 데이터 상의 오류로 인한 불필요한 가지들을 제거하는 ‘가지치기(pruning)’ 작업을 수행한다.

의사결정트리(Decision Tree)는 분류나 예측의 근거를 알려주기 때문에 이해하기가 쉽고, 모델 구축이 분류에 영향을 미치지 않는 속성들을 자동으로 제외시키기 때문에 데이터 선정이 용이하다. 또한 연속형이나 이산형 데이터 값들을 기록된 그대로 처리할 수 있기 때문에 지식발견 프로세스 중 데이터의 변환단계에서 소요되는 기간과 노력을 단축시키며, 어떠한 속성들이 각각의 분류 값에 결정적인 영향을 주는가를 쉽게 파악할 수 있고, 모델구축에 소요되는 시간이 짧다. 그러나 다음과 같은 단점으로 인하여 적용에 있어서 제약이 있다. 연속형 데이터를 처리하는 능력이 신경망이나 다른 통계기법에 비해 떨어지며, 결과적으로 예측력도 감소한다. 따라서 주가나 주택의 가격등과 같은 연속형 변수를 이용하는 모델을 구축하는 것이 목적일 경우에는 적합하지 않다. 또한 모델을 구축하는데 사용되는 표본의 크기에 지나치게 민감하다. 그러므로 정확한 모델을 만들기 위해서는 서로 상이한 값을 갖는 레코드들을 가능한 한 많이 포함하는 데이터가 필요하다.

4. 실험 및 결과

4.1 공격 및 정상 데이터의 구성

공격 데이터와 정상 데이터는 3. 에서 설명된 수집 방법을 이용해 수집한 후 ASCII 데이터로 변경하여 전처리 하였다. 데이터의 전처리는 5 초 간격으로 데이터를 구분하여 각 단위 시퀀스에서 발생된 이벤트 종류의 빈도를 기록하였다. 기록된 빈도에서 각각의 시퀀스의 이벤트 종류를 합하여 발생되지 않은 빈도는 0 으로 구성하였으며 총 빈도의 합은 시퀀스 수와 동일하게 구성하였다. 학습 데이터와 테스트 데이터는 무작위로 선정되어 아래 그림과 같이 비율을 동일하게 구성 하였다.



(그림 4) 데이터의 구성

4.2 공격 탐지 결과

본 실험의 결과는 1-level, 즉 최상위 노드가 한 단계만의 하위노드를 가지는 의사결정트리(Decision Tree)를 통해 분석하였다. <표 3>은 탐지율 및 confusion matrix 를 나타낸 것이다.

탐지율 (%)	정상	공격	
99.91	8632	0	정상
	7	37	공격

<표 3> 탐지율 및 Confusion Matrix

의사결정트리를 이용해 64 개의 Attribute 를 갖는 8676 개의 관측치를 분석한 결과, 정분류율은 99.91%로 굉장히 높았고 오분류율은 0.08%에 그쳤다. 또한 테스트데이터 8676 개의 관측치 중 공격에 해당하는 관측치는 44 개로 공격 event 개수 자체가 굉장히 적

은데, 이와 같은 데이터에서는 데이터에 대한 정분류율보다는 공격에 대한 분류 능력인 민감도가 더 중요하게 작용한다. 본 논문에서의 실험 결과에 나타난 민감도(TP Rate)는 84%로, 이를 통해 의사결정나무를 이용한 공격 탐지율이 높게 나타났다고 볼 수 있다.

5. 결론

본 논문에서는 네트워크 기반의 침입 탐지 시스템과는 다르게 커널 기반 데이터를 이용한 호스트 기반의 침입 탐지 기법을 제안하였다.

제안하는 방법에 대한 실험을 위해 커널 기반 데이터를 수집하여 전처리 한 후, 데이터마이닝 기법 중 하나인 의사결정트리에 적용하였다. 그 결과 서비스 거부 공격에 대한 오탐지율이 낮고 적중률과 민감도가 모두 높은 결과를 도출하였다.

본 논문에서는 의사결정트리를 사용하여 단순 Syn flooding 방법을 이용한 서비스 거부 공격만을 대상으로 탐지 실험을 진행하였지만, 차후에는 베이지안 결정 이론 등의 기계학습 알고리즘을 적용하여 분산 서비스 거부 공격 등 다양한 서비스 거부 공격 방법에 대한 연구를 진행 해야 한다.

참고문헌

[1] 문경원, 황병연, “ 서비스 거부 공격 대응을 위한 위험 탐지 모델링”, 한국정보과학회 가을 학술발표논문집 Vol.31, No.2

[2] Arbor Networks, http://www.itdaily.kr/news/quickViewArticleView.html?idxno=5979

[3] Stephen Northcutt, Judy Novak, “ Network Intrusion Detection : An Analyst’s Handbook(2nd Edition)”, New Riders Publishing, September 2000

[4] 이종엽, 윤미선, 이 훈, “ **DoS** 공격의 유형 분석 및 탐지 방법”

[5] 조현정, “ 차세대 네트워크 보안기술 기반의 침입방지시스템(IPS)”, 정보과학회지 제 23 권 제 1 호

[6] Ian H. Witten, Eibe Frank, “ DATA MINING Practical Machine Learning Tools and Techniques”