제18회 한국정보처리학회 추계학술발표대회 논문집 제9권 제2호 (2002.11)

기계학습 기법에 의한 비정상행위 탐지기반 IDS의 성능 평가

노영주*, 조성배
연세대학교 컴퓨터과학과
e-mail: yjnoh, sbcho@candy.yonsei.ac.kr

Performance Evaluation of IDS based on Anomaly Detection Using Machine Learning Techniques

Young-Ju Noh*, Sung-Bae Cho
Dept of Computer Science, Yonsei University

요 약

침입탐지 시스템은 전산시스템을 보호하는 대표적인 수단으로, 오용탐지와 비정상행위탐지 방법으로 나눌 수 있는데, 다양화되는 침입에 대응하기 위해 비정상행위 탐지기법이 활발히 연구되고 있다. 비정상행위기반 침입탐지 시스템에서는 정상행위 구축 방법에 따라 다양한 침입탐지율과 오류율을 보인다. 본 논문에서는 비정상행위기반 침입탐지시스템을 구축하였는데, 사용되는 대표적인 기계학습 방법인 동등 매칭(Equality Matching), 다층 퍼셉트론(Multi-Layer Perceptron), 은닉마르코프 모델(Hidden Markov Model)을 구현하고 그 성능을 비교하여 보았다. 실험결과 다층 퍼셉트론과 은닉마르코프모델이 높은 침입 탐지율과 낮은 false-positive 오류율을 내어 정상행위로 사용되는 시스템감사 데이터에 대한 정보의 특성을 잘 반영하여 모델링한다는 것을 알 수 있었다.

1. 서론

한국정보보호진흥원에 따르면 국내·외 컴퓨터 시스템에 대한 공격은 인터넷이 보급되기 시작한 98년 이래로 폭발적으로 증가하고 있으며, 한국의 경우 연 평균 300% 이상의 증가율을 보인다고 한다[1]. 공격을 위한 도구 또한 예전에는 단순히 시스템의 버그를 이용하는 것들이 주류를 이루었으나, 최근에는 은닉화(stealth), 분산화(distributed), 그리고 자동화(automation)의 특성을 갖는 공격 방법들이 늘어나고 있다[2].

침입탐지시스템은 시스템의 불법적인 사용이나 오용, 남용 등에 의한 침입을 탐지해내는 것으로, 침입탐지시스템의 탐지방법은 크게 오용탐지와 비정상행위탐지로 나눌 수 있다. 오용탐지 기법은 알려진 공격에 대한 정보를 구축한 후 사용자나 시스템 또는 프로그램의 현재 행동이 공격패턴과 일치하는지를 검사한다. 공격패턴 정보를 가지고 있으므로 정상행위를 공격행위로 간주하는 오류(false-positive error)가 낮고, 공격패턴만 검색하면 되므로 경제적인 장점이 있는 반면 알려지지 않은 새로운 공격은 탐지할 수 없다는 단점을 가지고 있다. 이에 비해 비정상행위탐지 기법은 모델링된 정상행위에서 벗어나는 행동을 공격행위로 간주하기 때문에 공격행위를 정상행위로 간주하는 오류(false-negative)가 낮다. 그러나 정상행위 모델링을 위해서 다량의 데이터를 분석해야 하므로 구현 비용이 높고, 학습되지 않은 정상행위는 비정상행위로 간주되므로 정상행위가 공격행위로 간주되는 오류가 높다[3].

본 논문에서는 비정상행위 탐지방식을 사용한 최적의 침입탐지시스템 구현을 위해 정상행위 모델링 방법에 따른 성능을 비교 분석하였다.

2. 비정상행위 탐지기반 IDS

침입탐지시스템의 구조는 그림 1과 같이 전처리 과정에서 정상행위모델인 프로파일을 생성하고 이를 이용해 침입판정 과정에서 침입탐지를 한다. 전처리 과정은 모델링을 위한 정상행위 감사자료를 필터링하고 축약한 후 정상행위 모델을 생성하며, 침입판정 과정은 프로세스들의 행위궤적을 필터링하고 축약하여 전처리 과정에서 생성한 정상행위 모델과의 비교를 통해 진행된다[4].

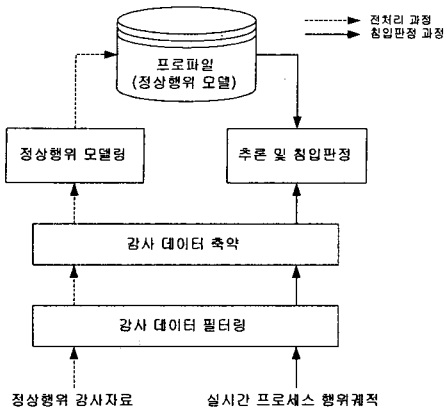

그림 1. 시스템 구조도

2.1 감사자료의 수집 및 전처리

권한 이동의 정보를 추출하기 위해 SunOS의 BSM(Basic Security Module)[4]을 사용하는데, BSM 데이터를 통하여 추출되는 막대한 양의 감사데이터를 이용하기 위해서는 설정파일 조정 등을 통해 정보의 손실을 최소화시키고 탐지에 필요한 효율적인 대표 값들을 추출하는 작업이 필요하다[5]. 본 논문에서 추출된 다량의 레코드 정보 중 UID, EUID, 소유권, 시스템호출 이벤트 등을 사용한다. 순서적으로 생성되는 이벤트는 일정 크기의 윈도우를 옆으로 이동시켜가면서 윈도우 크기만한 시퀀스로 추출하였다.

2.2 권한이동

침입탐지시스템은 사용자의 키입력, 시스템호출, 접속당 사용시간, 시스템의 평균부하 등 다양한 관찰심볼을 기반으로 침입여부를 판정한다[6]. 침입탐지 문제는 관찰된 사용패턴을 침입행위와 정상행위로 분류하는 문제로 생각할 수 있다. 효과적인 침입탐지를 위해 감사자료의 양을 줄이면서 침입탐지율을 높일 수 있는 사용자 행위 시퀀스가 요구된다.

일반적으로 호스트 기반 침입탐지시스템이 사용하는 감사자료는 시스템 호출 시퀀스이다[5]. 침입의 궁극적 목표는 루트 권한의 획득이라고 할 수 있다. 이러한 침입은 사용하지 않거나 암호가 없는 일반 유저의 권한으로 침입하여 고도화된 공격 기법을 이용, 루트의 권한을 획득한다. 이때 Real User ID(UID) 값과 Effective User ID(EUID) 값의 차이가 생긴다. 또한 유닉스 시스템에서 지원하는 일반적인 권한 이동과 비교, 분석하였을 경우 차이점을 발견할 수 있다. 대부분의 침입은 SETUID로 설정된 파일 버그를 이용하여 권한을 획득하게 되므로, 권한 이동은 정상행위와 비정상 행위를 구분하는 중요한 특징이 된다.

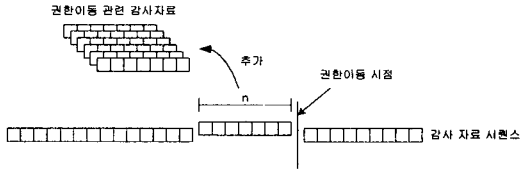

그림 2. 권한이동 관련 감사자료 추출방식

하나의 시스템 호출에 대해 기록되는 정보는 다양하지만 권한이동 모델에 적용하기 위해 UID, EUID, 소유권, 시스템호출 이벤트를 추출하여 사용한다. 순서적 이벤트를 처리하기 위해서는 고정된 단위로 이벤트를 분할할 필요가 있는데 슬라이딩 윈도우 방식을 사용하여 수집된 시퀀스 중 권한이 이동되는 시점과 관계된 시퀀스만을 추출해 정상행위 모델을 생성하는 자료로 이용한다[7].

2.3 정상행위 모델링과 침입판정

정상행위 모델링은 전처리 단계에서 생성된 정상행위 시퀀스를 기반으로 각 모델을 이용하여 정상행위 모델을 생성하는 과정이다[8]. 비정상행위 판정에서는 이미 구축되어 있는 모델링별 정상행위에 사용자 행위 시퀀스를 입력으로 넣고 각 정상행위에서 현재 행위가 생성되었을 유사도(MLP, EM)나 확률(HMM)을 구해 침입판정을 하는 과정이다.

3. 기계학습 모델링 기법

3.1 동등 매칭(Equality Matching)

동등 매칭은 두 패턴 요소간의 대응을 수행하여 유사도를 계산하는 방법으로 제시된 3가지 방법 중 가장 쉽게 구현되는 방법이다[9]. 정상행위 모델을 만들기 위해 27개씩 슬라이딩 윈도우 방식으로 잘려진 정상 시퀀스 모두를 테이블에 저장한다. 비정상행위 판정에서 새로운 사용자행위 시퀀스가 입력되면 저장된 정상행위 시퀀스와 유사도를 측정한다. 유사도 측정에서 일반적으로 사용하는 지수는 다음과 같은 유클리디안 거리(Euclidean distance)이다.

$$D(Q, I) = \sum_{i=1}^{n} \sum_{j=1}^{n} |E_{1_{ij}} - E_{2_{ij}}|$$

- Q : 사용자행위 시퀀스
- I : 정상행위 시퀀스
- $E_{1_{ij}}$: 사용자행위 특징값
- $E_{2_{ij}}$: 정상행위 특징값

3.2 다층퍼셉트론(MLP)

다층퍼셉트론은 입력층과 출력층 사이에 하나 이상의 은닉층(hidden layer)이 존재하는 신경망으로 그림4와 같은 계층구조를 갖는다.

MLP의 학습방법은 입력층의 각 노드에 데이터를 입력받고 이 신호들은 가중치 곱의 합으로 계산되어 은닉층으로 전달된 후 최종적으로 출력층으로 나오게 된다. 이 출력값과 원하는 출력값을 비교하여 그 차이를 감소시키는

방향으로 연결강도를 조정하는 것이 학습이다[10].

정상행위 모델링을 위하여 전처리 단계에서 생성된 정상행위 시퀀스와 랜덤생성기로 생성된 비정상행위 시퀀스를 기반으로 역전파(backpropagation) 알고리즘을 이용하여 MLP를 학습시킨다[11]. 역전파 신경망에 적용하기위해 슬라이딩 윈도우 방식으로 27개씩 나뉘어진 정상행위와 비정상행위 시퀀스를 입력한 후 다음 식에 의해 계산한다.

$$O = W_0 \sum_{i=1}^{n} W_i X_i$$

주어진 입력값으로 계산된 결과값과 실제값을 비교한후 오류를 산출한 뒤 그 오류를 최소화 하기위해 가중치를 조절하는데 시간 Δt 이후의 가중치 값은 아래와 같다.

$$w_i(t+\Delta t) = w_i(t) + \eta \delta x_i$$

비정상행위 판정에서는 이미 학습된 정상행위 MLP에 사용자행위 시퀀스를 입력으로 넣어 출력된 값이 정상인지 침입인지 판단을 내린다.

사용된 MLP는 입력층, 출력층 그리고 한개의 은닉층으로 구성되었으며, 입력층의 노드수는 27개, 은닉층의 노드수는 10~20개를 두었고, 출력층 노드수는 정상일때 [1.0, 0.0] 침입일때 [0.0, 1.0]을 나타내는 두개로 구성되었다. 학습은 200번부터 10000번까지 반복학습을 하였고, 각각 10회 반복하였다.

3.3 은닉마르코프모델(HMM)

은닉마르코프 모델은 상태라 불리는 N개의 노드와 노드간에 방향성을 갖는 전이를 나타내는 아크로 구성된 그래프 구조이다. 이 그래프의 각 노드에 공간적인 특성을 모델링하는 관측심볼 확률분포와 초기상태 확률분포가 저장되어 있고, 각 아크에는 관측열의 시간적인 특성을 모델링하는 상태전이 확률분포가 저장되어 있다. HMM은 주어진 관측열(순서적 이벤트)에 대해 비록 외부에서 그 상태전이과정을 직접적으로 관찰할 수는 없어도 마르코프 과정의 확률함수로 모델링할 수 있다. 그림 5는 3개의 노드가 연결된 HMM의 구조를 보여준다.

HMM은 관찰 시퀀스의 길이, 상태수, 심볼수와 학습에 의해 조정되는 전이확률, 관측확률, 초기 상태분포로 구성된다. 전이확률은 한 상태에서 다음상태로 전이할 확률을 나타내며, 관측확률은 한 상태에서 특정 심볼이 관측될 확률을 나타낸다. 초기 상태분포는 처음에 해당 상태에서 시작할 확률을 나타낸다. HMM은 다음과 같이 표현되며, A와 B를 만족하는 이중확률과정을 이산형 HMM이라고 하고 모델 λ는 간략히 (A, B, π)로 표현될 수 있다[12].

정상행위 모델링은 전처리 단계에서 생성된 정상행위 시퀀스를 기반으로 HMM의 매개변수를 결정하는 과정이다. HMM의 매개변수 결정은 주어진 시퀀스 O가 해당 모델 λ로부터 나왔을 확률인 $\Pr(O|\lambda)$ 값이 최대가 되도록 $\lambda=(A,B,\pi)$ 를 조정한다. 이를 계산하는 해석적인 방법은 알려져 있지 않고 반복적으로 λ를 결정하는 방법으로 Baum-Welch의 재추정식이 있다[13].

비정상행위 판정에서는 이미 구축되어 있는 정상행위별 HMM에 사용자행위 시퀀스를 입력으로 넣어 각 정상행위에서 현재 행위가 생성되었을 확률을 구한다. 확률을 구하는 방법으로는 forward-backward procedure나 Viterbi 알고리즘을 사용할 수 있다[12]. 각 모델별로 구해진 확률은 판정모듈에 전달되어 비정상행위인지 판정된다.

4. 실험 및 결과

학습 데이터는 한달 동안 7명의 사용자가 발생시킨 정상행위 데이터를 사용하였다. 주 사용 프로그램은 메일 서버, 홈페이지 접속, 유닉스 명령어 실행, FTP 데이터 전송, 그리고 사용자가 작성한 프로그램이었다. 생성된 사용자 이벤트는 모두 767,237개였으며 테스트에서는 6명의 사용자가 참가하였고, 그 중 3명이 총 17개의 Exploit에 대해서 2~3번씩 침입을 시도하였다. 총 767,237개의 이벤트 중 권한이동과 관련된 이벤트의 수는 5,950개이다.

비정상행위 침입탐지를 위해 동등매칭과 다층퍼셉셉트론 방법에서 사용한 데이터는 정상행위와 대응되는 비정상행위 시퀀스를 랜덤 발생기로 만들어 사용하였다.

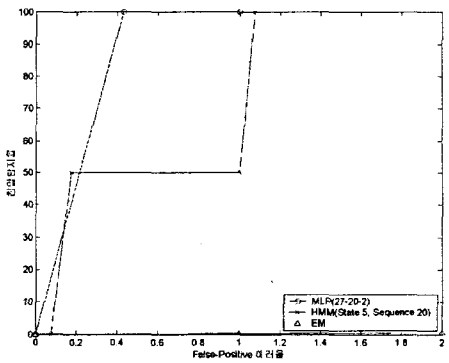

그림 3. EM, MLP, HMM 적용결과

그림 6은 각 실험에서 얻어진 침입탐지 결과를 ROC 곡선으로 보여주고 있다. 실험결과 EM방법은 매우 저조한 침입탐지율과 높은 false-positive 오류율을 보였다. 총 침입탐지율은 24.78%로서 시스템에서 발생하는 시스템 호출이 다양하게 발생되며 단순히 거리에 의해서는 시퀀스의 비교가 되지 않음을 알 수 있다. MLP 방법은 학습의 효과를 최대화하기 위해 은닉층의 노드수를 10, 15, 18, 20개로 각각 설정하고 학습세대수를 200부터 10000까지 조정하여 학습하였다. 침입 탐지를 위한 최상의 MLP는 27-20-2의 구조를 가지고 10000번 반복을 통해 얻어졌다. 침입판정율은 false-positive 오류율이 0.43%에서 100%의 침입탐지율을 보인다. HMM을 이용한 실험에서는 다른 방법들과 실험 데이터 적용에 동일성을 위해 시퀀스의 길이를 27로 고정하고 최적의 탐지를 위해 상태수를 5, 7, 10, 15로 변

경해가며 실험해 보았다. 실험결과 2.44%의 false-positive 오류율에서 100%의 침입탐지율을 보였다.

5. 결론

본 연구에서는 기존에 HMM모델을 이용할 경우 시간이 너무 많이 걸린다는 단점을 극복하고 좀더 좋은 탐지율을 얻기 위해 권한이동 침입탐지 시스템을 이용하여 다양한 탐지 모듈별 성능을 비교하였다. 시스템은 크게 전처리, 정상행위 모델링, 추론기법으로 나눌 수 있다. 전처리 부분은 SunOS에서 지원하는 BSM 모듈을 이용하여 Audit 데이터를 추출한다. 현재 BSM에서 사용되는 이벤트들은 시스템 콜, 시간, ruid, euid, rgid, egid, 등이다. 이러한 이벤트의 추출을 통하여 권한이동 순간을 비교하게 되는데 제안한 권한이동 모듈은 BSM 데이터에서 발생되는 EUID와 UID가 변경되었을 경우 그 시점을 기준으로 전에 사용되었던 일정양의 데이터 시퀀스를 가지고 평가를 하게 된다. 하지만 수집한 공격행위 분석 시 사용자의 변화뿐만 아니라 때로는 그룹 사용자가 변화하여 시스템을 침범 후 관리자 권한을 획득하는 경우가 있다. 이를 위해 본 연구에서는 사용자와 그룹 모두의 권한이동을 탐지하였다. 이렇게 추출된 일정한 크기의 시퀀스정보를 동등 매칭, 다층퍼셉트론, 은닉마르코프 모델을 통해 정상행위 학습을 한다. 동등 매칭은 저장된 시퀀스와 입력된 시퀀스간 거리를 비교하여 침입을 판정하므로 다양한 시퀀스에 대한 판정을 제대로 수행하지 못한다. 다층퍼셉트론을 이용한 침입탐지 결과는 적은 false-positive율을 보이면서 높은 탐지율을 나타내었다. 그러나 시퀀스가 가지고 있는 시간적 정보를 반영하지 못한다는 단점이 발생하게 된다. HMM에 의한 정상행위 학습은 시스템에서 생성된 시퀀스간 시간적 정보를 이용하여 일단 HMM의 정상행위가 모델링되면 사용자의 행위에 대한 평가 값을 얻을 수 있기 때문에 이러한 평가 값의 임계치를 정하여 침입여부를 결정하고, 침입탐지율도 매우 높게 나온다.

참고문헌

[1] 한국정보보호진흥원, "2001년 정보화역기능 실태조사", 2001.

[2] 한국정보보호진흥원, "99 국내외 해킹현황 분석", http://www.certcc.or.kr/statistics/hack/1999/99-hack.htm.

[3] T. Lane, "A Survey of Intursion Detection Techniques", *Computer & Security*, vol. 12, no. 4, June 1993

[4] 한국정보보호진흥원, "호스트기반 실시간 침입탐지시스템 개발을 위한 SunSHIELD Basic Security Module의 분석", 1998.

[5] Steven A. Hofmeyr, Stephanie Forrest, Anil Somayaji, "Intrusion Detection Using Sequences of System Calls", *Journal of Computer Security*, Vol 6. pp 151~181, 1998.

[6] Edward G. Amoroso, "Fundamentals of Computer Security Technology", *Prentice Hall*, 1994

[7] 박혁장, 정유석, 노영주, 조성배, "사용자 권한이동 이벤트 모델링 기반 침입탐지시스템의 체계적인 평가", 한국정보과학회, 제 28권 2호, pp. 661~663, October 2001.

[8] T. Lane C.E. broadly, "An Application of Machine Learning to Anomaly Detection" , *Proc, of NISSC '97*, pp. 366~380, 1997.

[9] A. K. Ghosh, A. Schwartzbard and M. Schatz, "Learning Program Behavior Profiles for Intrusion Detection", *Proceeding of Workshop on Intrusion Detection and Network Monitoring*, April 1999.

[10] Richard P. Lippmann, " An Introduction to Computing with Neural Nets", *IEEE ASSP MAGAZINE*, pp. 4~21, April 1987.

[11] James Cannady, James Mahaffey, "The Application of Artificial Neural Networks to Misuse Detection", *Proceedings of the 1998 National Information Systems Security Conference (NISSC'98)*, pp. 443~456 , October 1998.

[12] L. R. Rabiner and B. H. Juang, " An Introduction of Hidden Markov Models", *IEEE ASSP Magazine*, pp. 4~16, January 1986.

[13] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257~286, Febebuary 1989.