침입탐지시스템에서 긍정적 결함을 최소화하기 위한 학습 방법

정종근, 김철원

Learning Method for minimize false positive in IDS

Jong-Geun Jeong , Chul-Won Kim

요 약

시스템에서 사용 패턴의 다양화 때문에 비정상 행위 탐지 IDS를 구현하는 것은 오용탐지 IDS를 구현하는 것보다 많은 어려움이 있다. 따라서 상용화되어 있는 대부분의 IDS는 오용 탐지 방법에 의한 것이다. 그러나 이러한 오용 탐지 방법에 의한 IDS는 변형된 침입 패턴이 발생할 경우 탐지해내 지 못한다는 단점을 가지고 있다. 본 논문에서는 감사데이터간의 침입 관계를 가지고 침입을 탐지하기 위해 데이터 마이닝 기법을 적용한다. 분산되어 있는 IDS에서의 에이전트는 시스템을 감시할뿐만 아니라 로그데이터까지 수집할 수 있다. 침입탐지시스템의 핵심인 탐지정확도를 높이기 위해 긍정적 결함이 최소화 되어야 한다. 따라서 감사데이터 학습단계에서 변형된 침입 패턴을 예측하기 위해서 데이터 마이닝 알고리즘을 적용한다.

ABSTRACT

The implementation of abnormal behavior detection IDS is more difficult than the implementation of misuse behavior detection IDS because usage patterns are various. Therefore, most of commercial IDS is misuse behavior detection IDS. However, misuse behavior detection IDS cannot detect system intrusion in case of modified intrusion patterns occurs. In this paper, we apply data mining so as to detect intrusion with only audit data related in intrusion among many audit data. The agent in the distributed IDS can collect log data as well as monitoring target system. False positive should be minimized in order to make detection accuracy high, that is, core of intrusion detection system. So We apply data mining algorithm for prediction of modified intrusion pattern in the level of audit data learning.

Ⅰ. 서 론

IDS는 매우 다양한 모델이 있으나 이들의 기본 기능은 감사정보 수집, 수집된 정보의 분석 및 침입 판정, 보고 및 대응 행동으로 요약할 수 있다. 감사정보수집 기능은 감시대상 시스템 또는 네트워크로부터 보안 분석을 위한 감사정보를 수집하며, 정보 제공원에 따라 시스템의 로그파일, 시스템 호출 함수 등으로부터 정보를 수집하는 호스트 기반 IDS와 네트워크패킷으로부터 감사정보를 수집하는 네트워크기반 IDS로 분류한다[1]. 네트워크 기반 IDS와 호스트 기반 IDS는 여러 가지 면에서 차이를 가진다. 이러한 이유들 때문에 네트워크 기반 IDS의 도입이 좀더 증가 추세에 있으며 최근 두 모델을 혼합한 형태의 하이브리드형 IDS가 확산되고 있다.

침입판정을 위한 분석 기술은 오용탐지(Misuse Detection) 기법과 비정상탐지(Anomaly Detection) 기법으로 구분된다.

오용탐지기법은 일반적으로 침입으로 알려져

접수일자 : 2003. 5. 15

Matching) 기술을 사용하며 현재 많은 상용제품들이 오용탐지기법을 사용하고 있다.

비정상탐지기법은 정상적인 행위에 대한 프로파일을 생성하고 실제 수집되는 감사 데이터를 프로파일과 비교해 정상행위로부터 벗어나는 비정상행위를 탐지하는 기법이다. 새로운 침입 또는 오용의 탐지에 효율적이라는 장점이 있는 반면, 탐지비용이 높고 악의적인 목적으로 자신의 행위 패턴을 서서히 학습시키는 사용자에게는 취약하다[1,4]. 또한 데이터베이스의 정확도에 따라 정상행위를 침입으로 분류하는 긍정적 결함(False Positive) 오류를 범할 수도 있다. 비정상탐지 모델은 데닝의 모델이 기반을 이루고 있는데 현재 많이 적용되고 있는 탐지 모델로는 수량적 분석, 통계적 분석 그리고 신경망 기반 모델 등이 있다. 수량적 분석 모델은 탐지 규칙 또는 속성 값에 수치적인 값을 사용하여 침입 또는 오용을 탐지하는 방식으로써 대표적인 수량적 분석 모델에는 임계값에 기반한 탐지방식이 있으며 현재 많은 IDS가 임계값을 통한 침입탐지방식을 사용하고 있다. 그러나 임계값 기반 비정상탐지 방식은 침입 판정을 위한 정확한 임계값 설정의 어려움으로 인해 긍정적 결함이 증가한다는 문제점이 있다. 현재까지 제시된 침입탐지 시스템들은 몇 가지 문제점들을 공통적으로 가지고 있는데 이 중 가장 두드러진 문제점은 시스템 부하에 관한 것이다. 이를 해결하기 위해 별도의 침입탐지모듈에 의해 네트워크 전체가 분석되도록 하고 있다. 감사 흔적(Audit Trail)을 분석하기 위해서는 시스템 커널이 시스템 상에서 이루어지는 모든 행동들에 대해 감사 정보를 만들어 내야 하는 데, 그 양이 엄청나며 분석 작업에는 시스템의 디스크 용량이나 CPU Time의 엄청난 소모가 필요하다. 실제적으로 영국의 University College London(UCL)에서도 침입탐지 시스템을 개발하기 위해 기존의 신경망 기법과 유전자 알고리즘, 그리고 전문가 시스템 기술을 이용하였으나 규모문제(Scale)에 부딪쳐 사업이 중단되었다. 이는 소규모 시제품 시스템에서는 인공지능 기법이나 분류기법, 유전자 알고리즘이 효과적으로 적용되지만 실제 네트워크 시스템에서는 그 규모 문제 때문에 운영이 어려워지기 때문이다.

따라서, 본 논문에서는 가능한 시스템의 부하를 최소화하기 위해 감사 데이터 표준화 방안을 모색하였고, 막대한 감사 데이터의 양을 효율적으로 축약하는 데이터 마이닝 기법을 적용하였다. 데이터 마이닝 기법은 다량의 패턴 데이터로부터 예측 가능한 패턴 데이터를 추출하여 변형된 침입을 탐지하고자 한다.

II. 침입탐지시스템 기술분석

침입 탐지 시스템을 구현하는 방법은 다음과 같이 크게 세 가지로 분류할 수 있다.

- 실시간 침입 감시 및 분석 기술
- 실시간 패킷 수집 및 분석 기술
- 사후 감사 분석에 의한 분석 기술

실시간 침입 감시 기술은 허가 받지 않은 파일에 대한 임의적 접근이나 변경, 로그인(login) 프로그램의 변경 등을 탐지한다. 실시간 침입 탐지를 위한 효과적인 방법은 네트워크를 구성하는 여러 가지 시스템과 장치에서 발생하는 불법적인 행위들을 실시간적으로 모니터링 하고 조치를 취해야 한다. 대부분의 행위 모니터링은 운영체제(OS)에서 제공해 주는 감사 데이터를 활용한다. 반면에 다각도에서 탐지해 내기 위해서는 Webserver, Router, Firewall, TCP/UDP port의 활성화 등에 의한 감사 자료들을 이용해야 한다.

실시간 침입 감시는 침입자가 대부분 관리자 권한을 획득하려고 하기 때문에 이러한 행위가 감지되면 즉각적인 조치를 취하게 함으로써 시스템의 피해를 줄일 수 있다.

실시간 침입 탐지 기술은 [그림1]과 같이 단일 호스트 침입 탐지와 다중 호스트 침입 탐지로 나눌 수 있는데 단일 호스트 침입 탐지는 오직 한 시스템에서만 작동하므로 오늘날과 같은 멀티 플랫폼 환경에는 적합하지 않다[8].

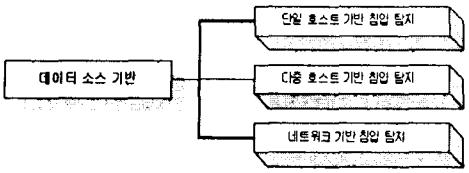

그림 4. 침입탐지시스템 분류

다중 호스트 침입 탐지 방법은 전체 네트워크와 시스템을 에이전트로 인식하여 분산된 환경에서 감사 자료를 수집, 분석하여 침입을 탐지한다. 실시간 패킷 수집 기술은 네트웍 침입 탐지에 이용하기 위해 각 시스템을 통해 지나가는 패킷들을 수집·분석하여 침입을 탐지하는 기술로 네트워크 기반 침입 탐지 시스템에서 이용한다. 사후 감사 분석 기술은 호스트 기반이나 네트워크 기반의 침입 탐지 시스템에서 발생하는 각종 로그 데이터나 침입 흔적 등을 실시간에 처리하지 않고 추후에 검사하여 이에 대한 분석이나 역 추적 등에서 이용되는 기술이다.

III. 데이터 마이닝을 이용한 감사데이터 학습

3.1 연관규칙을 이용한 감사데이터 분류

데이터 마이닝의 연관규칙 탐사 알고리즘 중 가장 대표적인 방법이 Apriori 알고리즘이다. Apriori 알고리즘은 여러 논문에서 연구되어 다양한 분야에 응용되고 있다. Apriori 알고리즘은 데이터베이스에서 후보 항목 집합을 구성하고, 구성된 후보 항목 집합에서 빈발 항목 집합을 탐사하는 과정으로 수행된다. Apriori 알고리즘은 후보 항목 생성시 모든 데이터베이스에서의 데이터 항목에 대한 생성이 아닌, 전 단계의 빈발 항목 집합을 대상으로 후보 항목을 생성한다. Apriori 알고리즘은 전 단계에서의 빈발 항목 집합에서 현재 단계의 후보 항목 집합을 구성 한 다음 데이터베이스의 스캔을 통해 후보 항목 집합의 지지도를 계산한다. 그리고, 사용자가 정의한 최소 지지도를 기초로 하여 현재 단계의 빈발 항목 집합을 구성한다. Aprirori 알고리즘의 단계의 진행은 데이터 항목의 증가에 따라 반복적으로 진행된다. k단계에서의 Apriori의 빈발 항목 탐사는 k-1 단계의 빈발 항목 집합으로부터 생성된 k-후보 항목 집합에 대하여 각각의 지지도를 계산한 후 이들 중에서 지지도를 만족하는 항목의 탐사를 통해 이루어진다. Apriori는 더 이상의 후보 항목을 생성할 수 없을 때까지 반복되어 빈발 항목을 탐사하며, 빈발항목 집합의 생성 알고리즘은 [그림 2]와 같다.

[그림 3]의 알고리즘과 같이 후보 항목 집합의 생성은 전 단계의 빈발 항목 집합의 조인 연산(Join operation)과 전지 과정(Prune process)을 통해 이루어진다. 조인 연산은 두 집합의 곱집합을 구하는 것과 같으며 전지 과정은 조인을 통해 생성된 후보 항목 집합의 부분 집합이 전 단계의 빈발 항목 집합의 원소가 아닌 경우, 그 항목을 삭제하는 과정이다. 그 이유는 전 단계에서 빈발하지 못하는 항목은 다음 단계에서도 빈발하지 못하기 때문이다. 전지 과정은 불필요한 후보 항목의 수를 줄여 데이터베이스를 읽는 횟수를 감소시키기 위하여 추가된 과정이다.

```
L1 = {large 1-itemsets}
for (k=2;Lk-1≠Ø;k++) do begin
  Ck = apriori-gen(Lk-1); //새로운 후보항목 집 합
   forall transactions t∈D do begin
  Ct = subset(Ck,t); //후보항목이 빈발항목 집합에
                      포함
       forall candidates c∈Ct do
          c.count++;
  end
Lk = {c∈Ck | c.count ≥ Smin}//최소지지도를 만족
end
Answer = ∪Lk;
```

그림 2. 빈발 항목 집합 생성 알고리즘

```
Algorithm Apriori-gen
insert into Ck  // 필요한 항목 추가
select a.item1, a.item2, ···,a.itemk-1, b.itemk-1
from Lk-1a,  Lk-1b
where  a.item1 = b.item1, ···,a.itemk-2;
                = b.itemk-2, a.itemk-1 < b.itemk-1
// 생성된 항목이 전단계의 빈발항목원소가 아닌 경우 삭제
for all itemset c∈Ck do
  for all (k-1)-subsets s of c do
  if (s∉Lk-1) then
     delete c from Ck
```

그림 3 조인연산과 전지과정의 알고리즘

3.2 침입패턴 분류

비정상 탐지를 위한 기계학습 방법의 어려움은 알려져 있지 않은 패턴과 알려진 패턴의 한계를 정하는 것이다. 학습 데이터에 있어서 비정상 패턴에 대한 별다른 예를 가지고 있지 않은 상태에서는 기계 학습 알고리즘은 훈련 데이터에 있는 알려진 패턴에 대한 한계를 구분할 수 없다[5,6]. 일반적으로 비정상과 오용 탐지를 구분하기란 쉬운 일이 아니다. 비정상 탐지는 전형적으로 비통제된 학습 방법을 사용하는 반면에 오용 탐지에서는 통제된 분류 방법을 사용한다[2,3,7].

따라서 변형된 패턴의 공격이 발생할 경우 이를 탐지해 내지 못하므로 변형된 새로운 유형의 공격이 발생할 경우, 이 공격 패턴을 즉시 학습시킴으로써 새로운 공격에 대응하고자 한다. 이를 위해 새로운 공격 패턴이 발생할 경우 이미 분류되어 있는 침입 패턴 집합에 계속적으로 추가시킨다. [그림 4]의 알고리즘에서와 같이 H2는 새로운 침입 패턴과 정상 데이터로부터 학습된 추가된 분류자이며 알고리즘에서 결정 규칙은 출력을 위해서 평가된다.

H_1은 존재하는 침입 탐지 시스템 모델이고 H2는 최근에 발견된 새로운 침입 패턴을 위해 훈련된 새로운 모델이다. H1에서는 정상과 비정상 패턴만을 확인하고 새로운 침입을 확인할 수 없기 때문에 대부분의 패턴들은 비정상과 오용으로 분류한다. 그러나 H2는 새로운 침입과 정상 데이터로 분류한다. 이때 새로운 침입 패턴의 양이 적기 때문에 H2는 다른 데이터로부터 침입 패턴을 쉽게 분류할 수 있다.

```
Intrusion_learning()
{
if (H1(x)-normal) ∨ (H1(x)-anomaly) then
  //정상패턴과 비정상 패턴 분류
  if H2(x)-normal
    then output ← H1(x)(normal or anomaly)
  //존재하는 침입 패턴 모델
    else output ← new_intrusion
else output ← H1(x)
}
```

그림 4. 침입 탐지 분류 알고리즘

3.3 에이전트 학습

본 논문에서는 이러한 에이전트 시스템의 특성에 기존의 학습 방법 대신 데이터 마이닝 학습 방법을 제안하였다. 데이터 마이닝 학습 방법과 기계 학습은 과거의 경험을 바탕으로 학습한다는 점은 동일하지만 다양한 패턴으로의 변화와 새로운 패턴의 추측 면에서는 데이터 마이닝 방법이 훨씬 뛰어나다. 따라서 기능은 기존의 에이전트 시스템과 동일하지만 학습 방법에서는 데이터 마이닝 방법을 채택하였다.

에이전트에 대한 학습 모듈은 [그림 5]와 같다.

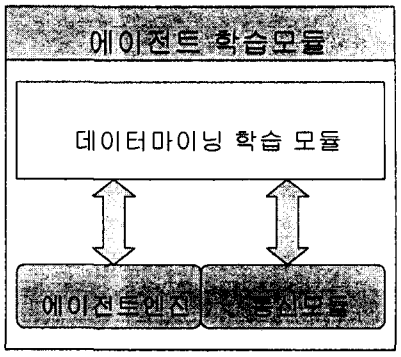

그림 5. 에이전트에서의 학습모듈

[그림 5]에서처럼 기존의 에이전트 엔진과 통신 모듈로 구성되어있고 데이터마이닝 학습 모듈을 에이전트에 탑재하였다. 에이전트는 수집하여 변환된 감사데이터를 학습모듈에서 학습시켜 직접 탐지에 이용한다. 침입 탐지 시 발생하는 긍정적 결함을 최소화하기 위해 임계값(threshold)를 학

습 단계에서 조정하여 시스템의 유연성을 크게하였다. 다음 [그림 6]은 에이전트에 감사데이터 학습 결과를 보여주고 있다.

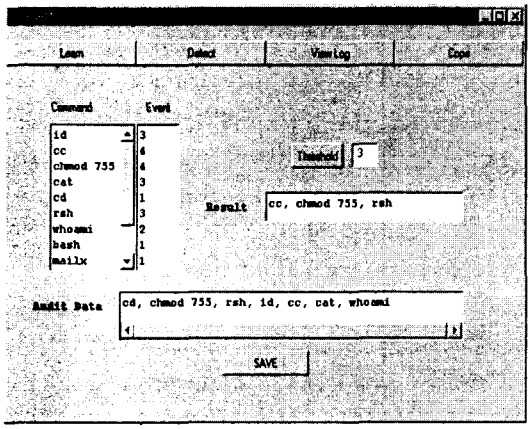

그림 6. 감사데이터 학습화면

3.4 감사데이터 표준화

감사 데이터 수집 에이전트에서 수집된 감사 데이터는 축약하여 표준 감사 데이터형식(SADF: Standard Audit Data Form)으로 변환시킨다. 변환된 표준 감사 데이터는 에이전트의 자율적인 침입 탐지를 위해 에이전트에 저장되고 일부는 침입 탐지 서버 시스템에 전송되어 분산된 공격을 탐지하는데 이용한다.

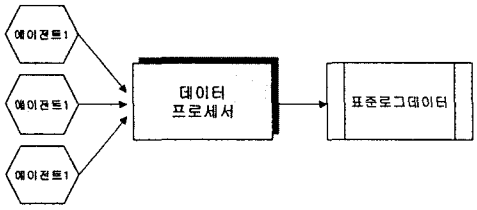

그림 7. 감사데이터 표준화 과정

침입 탐지 시스템을 구축하기 위해서 로그 프로세서에서는 비정상 행위 시나리오를 작성하여 각 시나리오에 해당하는 로그 데이터를 수집, 분석하게 된다.

여기에 사용된 호스트는 Linux를 기본으로 하지만 다중 호스트일 경우를 고려하였을 때 분석을 위한 로그 데이터의 형식이 OS마다 다르기 때문에 각각의 호스트에 로그 프로세서에서 로그감사 데이터를 표준화하는 방안을 찾아 침입 탐지 호스트의 부하를 최대한 감소시킨다. [그림 7]은 에이전트에 의해 수집된 로그 데이터를 데이터프로세서에 의한 감사 데이터 표준화 과정을 나타낸 것이다.

초기의 침입탐지에 대한 연구는 주로 감사 데이터의 분석에 초점이 맞추어졌다[9]. J. Anderson은 일괄처리 형식으로 설계된 추적 데이터의 분석 방법을 제시하였다. 이러한 방법은 사후에 추적 감사하는 방법으로 침입이 일어난 후에 실시하는 오프라인 분석 과정을 자동화하는 작업이다. 시스템 내에서 발생하는 로그 데이터는 방대한 양이어서 수작업에 의한 데이터의 수집 및 분석은 불가능하며 자동화된 추적 방법이 필수적이다. 자동화된 감사추적 기법의 발전적 형태인 침입탐지 시스템에서도 방대한 양의 로그 데이터들로부터 분석 및 침입 탐지가 가능할 수 있도록 하기위해 의미있는 정보로의 전환 및 축약시키는 단계가 필요하다[8].

Matt Bishop이 제안한 감사 데이터의 표준화 연구는 다수 호스트에서 생성한 다양한 형식의 로그정보를 표준 형식으로 변환하는 것이다[10,11]. 이전까지 연구되어온 침입탐지 시스템의 감사 데이터 기법은 시스템 의존적인 특성을 지니고 있어 이종의 환경을 지원하기에 적합하지 않다[4].

여기는 여러 호스트들의 다양한 로그 정보의 표준 형태로의 변환은 분석 호스트에서 별도의 처리 과정을 필요로 하지 않는다. 즉, 분석 호스트에서 재조정하는 과정을 수행하지 않고 감사 데이터 분석을 수행할 수 있다. 이와 같은 연구 결과로 분석 호스트의 작업 부하를 줄이고 이종의 환경을 지원할 수 있다.

```
struct std_audit_data {
        unsigned long     tseq;
        char              hostname[32];
        char              remotehost[32];
        char              ttyname[16];
        char              cmd[18];
        char              jobname[16];
        char              dellog;
        char              errlogin;
        char              errorflagd;
        time_t            timestamp;
        long              syscall;
        long              errno;
        char              port;
        long              pid;
```

그림 8. 감사 데이터 표준 형식

그러나, 방대한 양의 감사 데이터를 표준 형태로 변환하기 위해서는 시간이 많이 소요된다는 문제점이 있기 때문에 이 단계는 오프라인(off-line)으로 처리 하였다. 본 연구를 위해 Solaris 2.6과 SunOS 5.6의 로그 데이터를 데이터 프로세서를 통해서 표준화된 감사 데이터로 형성하였다. 데이터 프로세서를 통해 변환된 표준화된 감사 데이터의 구조는 [그림 8]과 같다.

IV. 구현 및 성능 평가

4.1 침입 탐지 시스템 구현

각 호스트에서 수집된 표준화된 로그 데이터는 이미 침입 탐지 시스템 데이터 베이스에 저장되어 있는 침입 패턴과의 매칭을 통해 침입을 판단한다. 이때 데이터베이스에 저장되어 있는 감사 데이터는 지속적인 학습을 통해 새로운 유형의 침입 패턴을 계속 갱신(update)한다. 또한 미리 정의해 놓은 규칙들로부터 변형된 침입 패턴을 예측·생성한다.

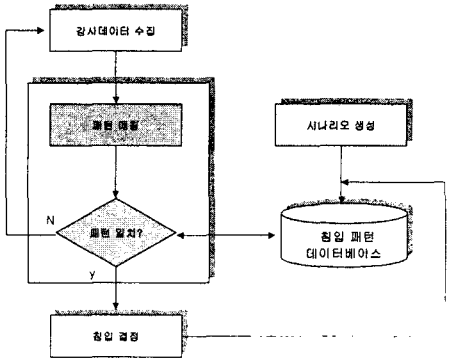

그림 9. 탐지 모듈 구성도
Fig 9. Strusture of detection module

[그림 9]는 침입 탐지 모듈에 대한 구성도이고, [그림 10]은 침입 탐지 알고리즘을 나타낸다. 에이전트에서 수집한 사용자의 로그 데이터와 침입 패턴(PT)과 비교하여 기대치 이상이거나 일치하는 침입 유형을 찾아서 일치하면 침입 상태를 보고한다. 시나리오에는 없지만 새로운 침입 패턴이라고 판명 될 때는 감사 데이터 DB에 저장한다.

```
Intrusion_Detection()
{
 event=Associate;
  //시나리오에 따라 사용자 로그데이터 수집
 forall intrusion p ∈ PT
  {
    CMP - compare(PT(p), event);
     //사용자의 로그 데이터와  침입 패턴 비교
    if (CMP > MIN(E)) then report-CMP
   //임계값 이상을 탐지
  }
  warning    Intrusion_detection(report)
   //침입 상태 보고
  store - new pattern
}
```

그림 10. 침입 탐지 알고리즘

[그림 11]은 에이전트에서 수집한 사용자들의 로그데이터와 침입 패턴 감사데이터를 비교하여 패턴의 일치여부를 백분율(%)로 표시하였다. 이 탐지 화면에는 기본적으로 사용자들의 IP주소, 계정, 시간, 공격 호스트 IP등이 표시되어 관리자가

쉽게 확인할 수 있다.

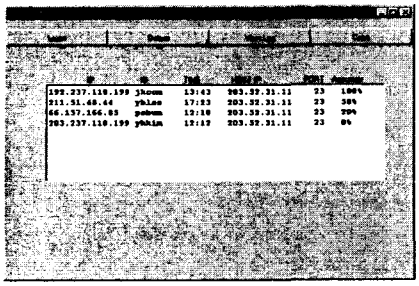

그림 11. 사용자별 침입 상황

4.2 성능 평가

침입 탐지 시스템의 핵심이 탐지의 정확도와 높은 탐지율이라면 가장 큰 문제점은 탐지 오판율을 최소화시키는 일이다. 침입 탐지 오판의 대부분은 긍정적 결함(false positive)과 부정적 결함(false negative)으로써, 이와같은 결함들을 최소화시키는 것이 오판율을 줄이는 것이다.

본 논문에서는 긍정적 결함을 최소화하는 것에 초점을 두었다. 긍정적 결함의 발생원인은 침입 패턴을 감사 데이터화하는 과정에서 침입 패턴에 대한 감사 데이터 범위를 결정하는 과정에서 발생한다. 이를 해결하기 위해서 본 논문에서는 감사 데이터를 학습하는 과정에서 데이터 마이닝 기법을 적용하여 하나의 침입 패턴에서 발생할 수 있는 여러 가지 변형 형태에 대한 예측 학습이 가능하도록 하여 긍정적 결함의 발생을 최소화하였다. [그림12]는 임계값(threshold)이 증가함에 따라 긍정적 결함이 감소하는 실험 결과를 보여주고 있다.

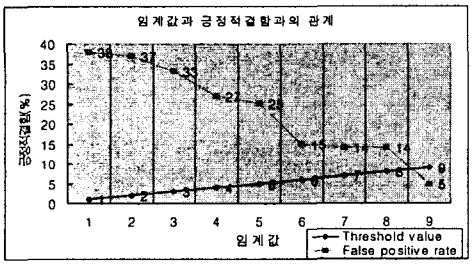

그림 12. 임계값과 긍정적 결함과의 관계

[그림 12]에서와 같이 임계값을 증가시킬수록 긍정적 결함의 비율이 작아지는 것들 볼 수 있다. 하지만 9이상의 임계값을 주었을때는 긍정적 결함의 비율의 오차가 별 차이가 나타나지 않는다. 이것은 본 침입 탐지 시스템에서도 완벽하게 긍정적 결함을 없앨 수는 없다는 것이다. 그 이유는 감사데이터 학습과정에서 정상인 침입 패턴의 일부가 학습되어지기 때문에 완전하게 긍정적 결함을 제거하는 것은 불가능하였다.

V. 결론 및 향후 연구 방향

본 논문에서는 침입 탐지 시스템에 데이터마이닝 학습 기법을 도입하여 다량의 데이터 축약과 변형된 침입 패턴을 탐지할 수 있게 하였다. 감시 시스템에서 발생하는 로그 데이터는 감사 데이터를 표준화하여 이종의 시스템에서 생성되는 다른 형태의 로그 데이터를 단일화된 형태로 변형하여 침입 탐지 시 시스템의 부하를 최소화하였다. 또한 독립적으로 침입을 탐지할 수 있는 에이전트를 채용함으로써 분산환경에서 적합하도록 설계하였다. 에이전트는 중앙의 침입 탐지 호스트와 계속 통신하여 새로운 감사 데이터를 제공받으며 에이전트가 수집한 새로운 감사 데이터는 침입 탐지 호스트나 다른 에이전트와 서로 정보를 공유하게 된다.

특히 본 논문에서 제안한 시스템과 현재 사용되고 있는 다른 침입 탐지 방법들과 비교할 때, 탐지의 정확도를 높였고, 오판율을 줄이기 위해 임계값을 상황에 따라 조절하여 긍정적 결함을 최소화하였다. 본 시스템은 어떤 감사 데이터를 학습시키냐에 따라서 침입 탐지 범위가 결정된다. 따라서 다양한 감사 데이터 학습이나 감사 데이터 양에 따라 탐지 능력을 향상시킬 수 있다.향후 연구 방향으로는 본 논문에서는 감사 데이터 학습 단계를 오프라인(offline)으로 처리하여 전체적인 시스템의 부하를 최소화하였으나, 온라인(online) 상태에서 수행하여 자동화된 침입 탐지 시스템을 구축하는 연구가 필요하다. 또한 감사 데이터 학습과정에서 최소 임계값을 결정하는 문제가 크게 대두되었다. 임계값을 크게 하면 수집된 데이터들에서 정확한 감사집합을 구하지 못해 부정적 결함

(False negative)이 발생할 수 있다. 따라서 감사 데이터 학습 시 적절한 임계값 설정에 대한 연구가 필요하다.

참고 문헌

[1] R. Buschkes, M. Borning, and D. Kesdogan,Transaction based Anomaly Detection" Proc.of the Workshop on Intrusion Detection and Network monitoring, USENIX, Apr., 1999.

[2] Anup K. Ghosh, "Learning Program Behavior Profiles for Intrusion Detection", Proc. of the Workshop on Intrusion Detection and Network Monitoring, April., 1999.

[3] Samuel I. Schaen, "Network Auditing: Issues and Recommendations", IEEE 7th Computer Security Applications Conference, pp.66-79, Dec., 1991.

[4] T. Lane, "Filtering technique for rapid user classification", In Proceedings of the AAAI98/ICML98 Joint Workshop on AI Approaches to Time series Analysis, 1998.

[5] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "The KDD process of extracting useful knowledge from volumesof data", Communications of the ACM, 39(11):27-34, Nov., 1996.

[6] W. Lee, S. J. Stolfo and K. W. Mok, "Mining Audit data to build Intrusion Detection Models", In proceeding of the 4th International Conference on KnowledgeDiscovery and Data Mining, New York, NY, Aug., 1998.

[7] 정종근, 이윤배, "새로운 침입 패턴을 위한 데이터마이닝 침입탐지시스템 설계", 대한전자공학회 논문지, 제39권 TE편 제1호 pp. 77-87, 3, 2002

[8] 한국전자통신연구원, "인터넷보안 기술/ 시장보고서", 12, 2001.

[9] Abdelaziz Mounji, Baudouin Le Charlier,Denis Zampunieris and Naji Habra, "Distributed Audit Trail Analysis", Proc. 2000.

[10] P. Proctor, "Audit Reduction and Misuse Detection in Heterogeneous Environment; Framework and Application", Proc 10th Annual Computer Security Applications Conference, Dec., 1994.

[11] Cheri Dowell and Paul Ramstedt. "The Computer Watch data reduction tool", In Proceedings of the 13th National Computer Security Conference, PP.99-108, Washington DC, Oct., 1990.

저자 소개

정 종 근

1995년 조선대학교 전자계산학과 졸업(이학사)
1997년 조선대학교 대학원 전자계산학과 졸업(이학석사), 2002년 8월 조선대학교 대학원 전자계산학과 졸업 (이학박사)
1999년 3월 - 2002년 동강대학 전자정보과 겸임교수
2003년 3월 - 2003년 현재 호남대학교 컴퓨터공학과 겸임교수
※ 관심분야 : 인공지능, 검색엔진, 데이터베이스, 정보보안, 전자상거래, 바이러스

김 철 원

1997년 광운대학교 컴퓨터공학과 (공학박사)
1988년 ~ 현재 호남대학교 컴퓨터 공학과 교수
※ 관심분야 : 정보보안, XML응용, 전자상거래 멀티미디어 정보 검색