침입탐지시스템의 성능향상을 위한 결정트리 기반 오경보 분류

(Classification of False Alarms based on the Decision Tree for Improving the Performance of Intrusion Detection Systems)

신 문 선 † 류 근 호 ††
(Moon Sun Shin) (Keun Ho Ryu)

요 약 네트워크 기반의 침입탐지시스템에서는 수집된 패킷데이타의 분석을 통해 침입인지 정상행위인지를 판단하여 경보를 발생 시키며 이런 경보데이타의 양은 기하급수적으로 증가하고 있다. 보안관리자는 이러한 대량의 경보데이타들을 분석하고 통합 관리하여 네트워크 보안레벨을 진단하거나 시간에 따른 적절한 대응을 하는데 유용하게 사용하여야 한다. 그러나 오경보의 비율이 너무 높아 경보 데이타들간의 상관관계 분석이나 고수준의 의미 분석에 어려움이 많으므로 분석결과에 대한 신뢰성이나 분석의 효율성이 낮아지는 문제점을 가진다. 이 논문에서는 데이타 마이닝의 분류 기법을 적용하여 오경보율을 최소화하는 방법을 제안한다. 결정트리기반의 분류 기법을 오경보 분류 모델로 적용하여 오경보들 중 실제는 공격이 아님에도 불구하고 공격이라 판단된 오경보를 정상으로 분류할 수 있는 경보 데이타 분류 모델을 설계하고 구현한다. 구현된 경보데이타 분류 모델은 오경보율을 최소화하므로 경보데이타의 분석 및 통합을 통해 경보메시지의 축약 및 침입탐지시스템의 탐지율을 높이는데 활용될 수 있다.

키워드 : 침입탐지시스템, 경보데이타, 오경보, 분류, 결정트리

Abstract Network-based IDS(Intrusion Detection System) gathers network packet data and analyzes them into attack or normal. They raise alarm when possible intrusion happens. But they often output a large amount of low-level or incomplete alert information. Consequently, a large amount of incomplete alert information that can be unmanageable and also be mixed with false alerts can prevent intrusion response systems and security administrator from adequately understanding and analyzing the state of network security, and initiating appropriate response in a timely fashion. So it is important for the security administrator to reduce the redundancy of alerts, integrate and correlate security alerts, construct attack scenarios and present high-level aggregated information. False alarm rate is the ratio between the number of normal connections that are incorrectly misclassified as attacks and the total number of normal connections. In this paper we propose a false alarm classification model to reduce the false alarm rate using classification analysis of data mining techniques. The proposed model can classify the alarms from the intrusion detection systems into false alert or true attack. Our approach is useful to reduce false alerts and to improve the detection rate of network-based intrusion detection systems.

Key words : Intrusion Detection System, Alert/Alarm, False Alarm, Classification, Decision Tree

· 이 논문 또는 저서는 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(지방연구중심대학육성사업/충북BIT연구중심대학육성사업단)

† 정 회 원 : 건국대학교 컴퓨터시스템 교수
msshin@kku.ac.kr
†† 종신회원 : 충북대학교 전기전자 컴퓨터공학부 교수
khryu@dblab.chungbuk.ac.kr
(Corresponding author임)

논문접수 : 2004년 2월 5일
심사완료 : 2007년 10월 22일

정보과학회논문지 : 데이타베이스 제34권 제6호(2007.12)

1. 서 론

최근 인터넷을 이용한 공격은 단순 시스템 공격에서 네트워크 공격으로 변하고 있다. 따라서 정보 보호를 위해서 여러 가지 형태의 보안 메커니즘들이 이용되고 있다. 방화벽, 침입 탐지 시스템, 안티 바이러스프로그램, 인증서비스 등이 대표적인 정보보호 메커니즘들이다. 그런데 이러한 정보보호 메커니즘들은 많은 양의 경보데이타들을 발생시키고 있으며 보안관리자는 이러한 대량의 경보데이타들을 분석하고 통합 관리하여 네트워크의 보안 레벨을 진단하거나 시간 내에 공격에 대한 적절한 대응을 하는데 유용하게 사용하여야 한다. 특히 침입탐지시스템이 정보보호 메커니즘으로 많이 사용되고 있으나 오경보의 비율이 너무 높아 경보 데이타들간의 상관관계 분석이나 고수준의 의미분석에 어려움이 많다. 이것은 분석결과에 대한 신뢰성이나 분석의 효율성이 낮아지는 문제점을 가지게 된다.

따라서 이 논문에서는 데이타 마이닝의 분류 기법 중 결정트리에 기반한 침입탐지시스템의 오경보율을 최소화하는 방법을 제안한다. 결정트리기반의 분류 기법을 오경보 분류 모델에 적용하여 오경보들 중 실제는 공격이 아님에도 불구하고 공격이라 판단된 오경보를 정상으로 정확히 분류 할 수 있는 경보 데이타 분류 모델을 설계하고 구현한다. 그리고 결정트리 기법을 사용하는 경우 트리의 깊이에 따라 정확도와 비용간의 상충적 문제가 발생한다. 이 문제를 해결하기 위해서 첫째 여러 속성 중에서 가장 유용하게 분류해 낼 수 있는 속성을 선택하는 방법이고, 두 번째는 결정트리 분할을 위해 사용되는 노드의 속성 값들에 대한 변환이다. 연속적인 속성 값들을 이산적인 형태로 변환하여야하는데 이 논문에서는 분류 속성 선택을 위해 연관규칙 기반의 빈발항목과 통계적 상관관계 분석을 이용한 접근방안을 제시한다.

구현된 경보데이타 분류 모델은 오경보율을 최소화하므로 경보데이타의 분석 및 통합을 통해 경보메시지의 축약 및 침입탐지시스템의 탐지율을 높이는데 활용될 수 있다.

논문의 구성은 다음과 같다. 2장에서는 관련 연구로써 침입탐지시스템의 오경보 관련 기존 연구를 분석하며, 3장에서는 결정트리기반 오경보 분류 모델을 제안한다. 4장에서는 제안된 모델의 실험평가를 위한 프로토타입의 구현에 대하여 기술하며 5장에서는 실험 및 평가를 통해 제안된 모델의 유용성을 분석하고 6장에서 결론 및 향후 연구에 대하여 기술한다.

2. 관련 연구

네트워크의 규모가 커지고 공격수법들이 다양해지고 지능화 되어감에 따라 기존 침입탐지 시스템[1]은 대응에 한계를 드러내고 있고 이에 따라 경보데이타를 분석하여 능동적인 대응을 하기 위한 여러 방법들이 시도되고 있다. 경보데이타 상관관계 분석을 위해 시그너쳐 분석, 페트리네트, 통계적 방법 그리고 데이타 마이닝 기법 적용 등의 다양한 연구[2]가 있었으나 여전히 제한적인 적용에 그치고 있으며 또한 오경보에 대한 해결책을 제시하지는 못하고 있다.

미국방성의 지원으로 콜럼비아 대학에서 진행된 침입탐지시스템 프로젝트[3]는 이상탐지를 하는 침입탐지시스템을 적절한 비용수준에서 구축할 수 있도록 하며 또한 데이타 마이닝 기법을 적용하여 침입탐지시스템의 효율성을 높이고자 하였다. 그 시도로서 침입탐지시스템의 실시간 대응능력을 높이고 새로운 공격패턴을 찾아내기 위해서 감사데이타의 속성들 간의 상관관계 분석을 통하여 감사데이타 속성들의 수를 감소시켜 빠른 대응을 하고자 하였으며 또한 빈발패턴을 탐사하여 알려지지 않은 공격패턴을 유추하는데 데이타 마이닝 기법을 적용[4]하였다. 뿐만 아니라 마이닝을 통한 지속적인 규칙생성과 학습으로 자동화된 침입탐지 갱신모델을 구축할 수 있도록 하는 연구[5]가 진행 되었다.

마이트레사에서는 최근 관심 있는 결과들에 대한 가중치를 높이고 그렇지 않은 결과들의 가중치를 낮추어 유사한 공격 이벤트들을 그룹화 하는 방법이 연구되고 있는데 이는 분류 방법을 이용하여 관심항목과 비 관심항목을 구분하여 기술하고, 클러스터링 기법을 이용하여 이벤트들을 구분한 뒤 유사패턴을 찾아내어 이에 대한 가중치를 높임으로서 데이타량의 감소가능하다.

미네소타 대학에서 연구 중인 마인즈(MINnesota INtrusion Detection System) 프로젝트는 희귀 클래스로부터 학습하여 희귀클래스 예측 모델을 생성하고, 이상/아웃라이어 탐지를 하며 연관규칙 패턴 분석 방식을 사용하여 공격을 특성화 하는 방식을 연구[6]하고 있다.

그리고 과다 경보를 방지하기 위한 메세지 축약의 방법으로써 경보상관관계분석 방법[7]이 연구되고 있는데 이 분야를 크게 세 가지로 나누어 보면 확률적인 방법으로 경보데이타들 간의 유사성을 평가하는 확률기반 상관관계분석[8], 미리 정의된 상황(상황레벨: 7개)그룹별로 유사성을 평가하는 경보축약 및 상관관계분석[9], 마지막으로 경보메시지 전후의 연계 관계를 이용하여 유사성을 평가하는 전제조건기반 상관관계분석[10]이 있다. 이렇게 침입탐지시스템의 성능향상을 위한 다각화된 연구가 진행되고 있으나 아직 상용화 단계는 아니며 자동적인 침입탐지모델 구축을 위한 프레임워크를 제공하는 기반을 마련하고 있는 수준이다. 침입탐지시스템의 성능향상을 위한 기존의 연구들은 다음과 같은 문제점

을 가지고 있다.

- 인터넷 침입에 대응하기 위한 완벽한 침입탐지 시스템은 없다. 그 이유는 인터넷의 발달과 아울러 계속해서 새로운 형태의 침입 유형이 나타나고 있으며, 침입탐지시스템은 대부분 학습기반 혹은 시그네쳐 기반으로 새로운 혹은 변형된 공격을 탐지하는 탐지율은 지극히 낮다.
- 침입탐지시스템에서 공격이라고 판단되면 경보를 보내며 대량의 경보 데이타들중 잘못된 경보들의 비율이 아주 높다.
- 침입탐지 시스템의 표준화된 포맷이 없기 때문에 같은 성질의 경보 데이타도 각각의 탐지 센서에 따라 다른 경보데이타로 판단되어질 수 있고, 이것 역시 다량의 중복되는 데이타를 발생하게 된다.
- 이렇게 대량의 데이타들이 생성된다는 사실이 침입탐지 시스템의 부하를 가져오고, 성능저하를 가져오는 서비스 거부 공격이 될 수 있다는 점이다.

이러한 문제점들 중 침입탐지시스템의 성능저하의 원인이 되는 높은 오경보 문제를 해결하기 위해 불필요한 다량의 경보 데이타를 감소시키고 데이타 자체만으로는 무의미한 정보인 경보데이타를 분석하기 위해서 통계적 기법이나 데이타 마이닝 기법을 이용하여 유용한 정보를 찾아내는 작업[12]이 필요하다. 따라서 이 논문에서는 오경보 데이타들을 필터링하고 잘못 예측된 경보들 중 오경보를 분류하기 위해서 데이타 마이닝 기법의 분류기법[13]을 이용하여 경보데이타 분류 모델을 설계하고 구현한다. 분류 기법은 훈련데이타를 기반으로 학습을 수행하여 실험데이타들의 특성에 맞는 분류 모델을 구축하며 구축된 모델을 이용하여 실제데이타를 분류할 수 있게 한다. 오경보 분류 모델을 구축하기 위해서는 분류모델의 정확도를 보장하는 속성들이 선택하여야 하며 속성 선택을 위한 방법은 연관규칙 기반의 속성선택을 제안하였으며 이 방법은 경보데이타의 특성상 연관된 속성을 선택하여 공격별 유용한 속성들을 오경보 분류모델에 특성으로 추출해주는 역할을 하게 된다.

3. 결정트리 기반 오경보 분류

오경보는 침입탐지시스템의 예측 모델에서는 침입이지만 실제로는 정상 행위인 것을 의미한다. 즉, 공격의 판단은 표 1과 같이 생각해 볼 수 있다. 실제 침입과 침입이 아닌 정상행위에 대해 4가지로 판단 할 수가 있는데 1) 정상행위를 정상행위로 판단하는 경우(TN), 2) 침입행위를 침입으로 정확히 탐지하는 경우(TP), 3) 정상행위를 침입으로 오인하는 경우(FP), 4) 침입을 정상행위로 오인하는 경우(FN)이고 이중 오경보라는 것은 3)과 4)의 경우를 말한다. 경보들 중 이러한 오경보는 네트워크 전반에 걸친 보안 서비스의 질을 하락시키는 원인이 된다.

그러므로 침입탐지시스템의 탐지율 향상을 위해서는 이러한 오경보율 최소화를 위한 방안에 대한 연구가 필수적이다. 대부분 실제 공격 정보 속에 포함되어 있는 발생된 경보의 예를 표 2에서 보여주고 있다. 스노트에서 발생시킨 ICMP redirect host라는 경보이다. 이 경보는 네트워크 호스트가 호스트 다이어그램을 위해 ICMP가 방향을 바꾸고자 할 때 발생되는데 이 행위는 기본적인 행위로서 침입이라고 볼 수 없다. 그러나 침입탐지시스템에서는 이 행위를 공격으로 간주하여 표 2와 같은 경보를 발생시키는 것이다. 이러한 경보의 모든 속성을 데이타베이스에 저장하게 되지만 실질적으로 결정트리 기반 오경보 분류 모델을 생성하기 위해서는 특정 공격과 관련된 속성들을 추출하는 것이 결정트리의 정확도 보장을 위해 필요하다.

표 2 발생된 경보 예

```
[**] [1:472:1] ICMP redirect host [**]
[Classification: Potentially Bad Traffic] [Priority: 2]
11/11-15:21:29.856720 210.125.146.126 -> 210.125.145.192
ICMP TTL:63 TOS:0x0 ID:55550 IpLen:20 DgmLen:56
Type:5  Code:1  REDIRECT HOST NEW GW: 210.125.146.1
[Xref => http://www.whitehats.com/info/IDS135]
[Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-1999-0265]
```

따라서 오경보율을 최소화시키기 위해 이 논문에서는 데이타 마이닝 기법을 적용한다. 마이닝 기법은 알려지지 않은 지식의 추출뿐만 아니라 데이타의 분류 및 예측에도 활용되어지는 기술로 많은 응용분야에서 이를 적용 및 활용하고 있다. 기존의 마이닝 알고리즘중 결정트리기반의 분류알고리즘[13]을 확장 적용하여 연관규칙 기반의 속성 선택을 하는 경보 분류 모델을 제안한다. 제안된 경보데이타 분류모델을 구축하여 침입탐지시스템에서 경보를 발생시키기 전에 오경보를 걸러내는 역할을 하므로 오경보율의 최소화를 유도할 수 있다. 이는 결과적으로 많은 양의 경보로 인한 경보과다로 인한 서

표 1 공격 평가 표준 메트릭

Standard metrics		Predicted Connection Label	
		Normal	Intrusions
Actual Connection Label	Normal	True Negative(TN)	False Positive(FP)
	Intrusions	False Negative(FN)	True Positive(TP)

비스 거부 공격을 방지하는 역할도 하게 된다. 그림 1은 결정트리 기반 오경보 분류 프레임워크를 보여준다.

그림 1에서 볼 수 있는 것처럼 침입탐지시스템의 전반부 혹은 후반부에서 경보 데이타들중 오경보들 중 특히 거짓 공격인 경보들을 분류하여 필터링을 수행함으로써 침입탐지시스템의 오경보율을 감소시킬 수 있다.

이는 능동적인 보안이라는 측면에서 침입탐지시스템의 자동화된 침입탐지모델 구축에 활용될 수 있을 뿐 아니라 기존 침입탐지 시스템의 문제점으로 드러난 거짓 공격 경보의 양을 감소시켜 양질의 네트워크 보안 서비스 제공의 기반을 마련할 수 있을 것이다.

분류 기법은 데이타를 각 클래스가 갖는 특징에 근거하여 미리 정의된 클래스 별로 분류하는 것으로서 과거에 알고 있는 데이타베이스 정보로부터 새로운 레코드를 자동적으로 분류해 낼 수 있는 분류 규칙을 생성하는 것이다. 그래서 데이타베이스 레코드들이 주어지면, 클래스 레이블을 가지고 각각의 레코드에 대해서 분류를 하게 된다. 따라서 침입 탐지 시스템에서 경보 데이타의 오경보율을 최소화 시키는데 이 기법이 적합하다.

연관 규칙을 기반으로 하여 추출된 속성들로 저장된 데이타베이스로부터 속성들을 추출하여 속성 목록 데이타구조에 저장을 하게 된다. 트리를 구축하기 전에 먼저 트리의 노드를 구성하기 위해 각 속성별로 정보이득을 계산하여 정보 이득값이 가장 큰 속성을 트리의 루트로 놓고 트리를 구축하게 된다. 구축된 트리에서 아웃라이어를 제거하기 위해서 트리정제 과정을 병행하여 트리를 구축하게 된다. 구축된 트리로부터 분류 규칙을 추출하기 위해서 IF-THEN 절을 이용하여 규칙을 추출하게 된다. 구축된 분류 모델을 가지고 테스트 데이타로 실험을 하여 오경보 분류 모델을 최종적으로 생성한다.

거짓공격 오경보가 생기게 되는 원인으로는 탐지해야 할 패킷이 너무 방대하여 미처 탐지하지 못하고 놓치는 패킷이 생기는 경우와 패턴 매칭 방식을 이용하기 때문에 침입탐지시스템의 시그네처와 일치하지 않는 변형된 형태의 패턴을 탐지하지 못하는 경우이다. 거짓공격 오경보는 경보 발생 후에 관리자의 분석을 통해 판단하거나, 혹은 실제 시스템이 공격을 당한 후에야 판단할 수 있는 데이타이다. 따라서 이 논문에서는 거짓공격 오경보를 분석하기 위해 침입으로 검증된 달파 데이타의 티시피덤프 원시 데이타를 침입 탐지 시스템에 통과시켜 시스템 통과 전 공격의 개수와 통과 후 공격의 개수를 비교하여 두 가지 데이타의 차이를 거짓 공격 오경보로 정의하여 데이타베이스에 저장하였다. 저장된 거짓공격 오경보 데이타는 데이타 마이닝 기법을 통해 분류 모델 구축에 트레이닝 데이타셋으로 사용할 수 있다. 그렇지만 추출된 거짓공격 오경보는 원시 데이타이기 때문에 관계형 데이타베이스에 저장하기 위해서는 또 한번의 속성 추출을 위한 전처리 과정을 수행하여야 한다.

전 처리 프로세서에서 변환된 데이타는 타임스탬프에 의해 순서화 되어 있으며 네트워크 접속 정보를 포함하고 있으며 텍스트 형태로 이루어져 있고 프로토콜별로 분류하여 저장한다. 단, 여기에서는 필요속성 및 불필요속성의 선택에 대한 부분은 고려하지 않는다. 속성 추출과 속성 구축에 대한 부분은 미리 학습된 방법을 통해서 이루어진다.

저장된 거짓 공격 오경보 데이타를 이용하여 오경보 분류 모델을 구축하기 위한 절차는 다음과 같이 이루어진다.

단계 1: 속성 선택

단계 2: 훈련 데이타를 이용한 분류규칙 생성

단계 3: 실험 데이타를 이용한 분류 모델의 정확도 평가

단계 1은 의사결정트리의 뿌리마디부터 끝마디까지 사용할 속성들을 결정하는 부분이다. 속성을 선택하는 방법에는 여러 가지가 있을 수 있으나 이 논문에서는 데이타 마이닝 기법중 연관규칙을 이용한다.

연관 규칙이란 어떤 사건이 발생하면 다른 사건도 따라 발생하는 사건 사이의 강한 연관성을 의미하는 것으로 항목 집합으로 표현된 트랜잭션에서 각 항목간의 연관성을 반영하는 규칙이다. A⇒B의 형태를 갖는 패턴으로서, 이러한 규칙이 갖는 의미는 A 항목집합이 나타날 때는 B 항목집합도 동반하여 나타나는 경향이 있다는 뜻이다. 연관 규칙 탐사는 빈발 항목 집합을 결정하는 단계와 연관 규칙을 생성하는 단계로 구성된다. 빈발 항목 집합을 결정하는 단계는 미리 결정된 최소 지지도 이상의 트랜잭션 지지도를 가지는 항목 집합들의 모든 집합들인 빈발 항목 집합들을 찾아내는 단계이다. 연관 규칙을 생성하는 단계에서는 찾아낸 모든 빈발 항목 집합에 대해 그것의 부분집합 A가 support(A)에 대한

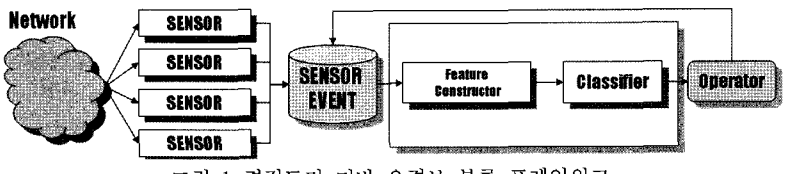

그림 1 결정트리 기반 오경보 분류 프레임워크

support(L)의 비율이 적어도 최소 신뢰도 이상이면 (support(L) / support(A) ≥ minimum confidence), A ⇒ (L-A)의 형태의 규칙을 생성한다. 이 때, 이 규칙의 지지도는 support(L)이고 신뢰도는 support(L) / Support(A)이다. 따라서 이 연관규칙을 이용하여 속성간의 상관관계를 분석함으로써 각각의 패킷들 간에 강한 연관성을 가진 속성들을 지지도와 신뢰도를 기반으로 선택할 수 있는 것이다. 이 논문에서는 마디가 될 속성을 선택하여 속성목록에 저장한다.

단계 2는 단계 1에서 저장된 속성목록의 속성 값들을 마디로 하여 분류규칙을 생성하는 과정이다. 먼저 첫 번째로 연속적인 값을 일정한 범위를 가지는 이산적인 값으로 변환하는 작업을 수행한다. 의사결정트리에서 하위 노드로 가지를 생성할 때 연속적인 값을 그대로 사용한다면 불필요한 개수의 가지가 생성되게 된다. 그래서 일정한 범위를 정해 이산적인 값으로 변환하는 것이다. 두 번째로 뿌리 노드 결정 작업을 수행한다. (1) 각 속성들의 정보값을 구하고 (2) 불순도(순수도)를 측정한다. 불순도 함수로는 엔트로피 지수를 이용하였다.

$$Information = -\log_2 p(p = \text{속성의 갯수}) \quad (1)$$

$$Entropy(S) \equiv -p_{FP}\log_2 p_{FP} - p_{TP}\log_2 p_{TP} \quad (2)$$

각각의 마디들이 가지는 내부 항목, 즉 속성들이 가지고 있는 이산적인 값들에 대해서도 불순도를 (1), (2)와 같은 방법으로 측정한다. 측정된 각각의 엔트로피 값을 이용해 최종적으로 정보이득율을 계산하게 되는데 (3)과 같은 방법을 이용하여 계산한다.

$$Gain(S,A) = Entropy - \Sigma\frac{|S_v|}{|S|}Entropy(S_v) \quad (3)$$

(S: 마디가 가지는 클래스 라벨의 갯수, S_v: 가지가 갖고 있는 클래스 라벨의 개수)

각 속성들에 대해 정보이득율을 계산한 후 가장 높은 속성을 뿌리 마디로 결정한다. 세 번째는 중간마디를 결정하는 과정인데 부모노드의 결정과 마찬가지로 (1)에서 (3)까지의 항목을 각각의 가지에 대해 수행한다. 재귀적인 방법으로 더 이상 분할이 일어나지 않을 때까지, 즉 끝마디까지 마디가 분할이 이루어지면 의사결정 트리가 종료된다. 트리의 분할이 종료되면 부모마디부터 끝마디까지 이어지는 규칙들이 생성되고 이 규칙은 규칙 데이타베이스에 삽입하여 단계 2를 마치게 된다.

단계 3은 훈련 데이타를 이용해 생성한 분류 규칙의 정확도를 평가하는 단계로써 훈련데이타의 일부분을 추출하거나 또는 소량의 실험용 데이타를 이용하여 수행한다. 이를 통해 관리자는 생성된 분류 규칙의 정확도 및 과적용 등을 분석하고 필요한 경우 가지치기를 통해 의사결정트리의 정확도를 높인다. 그림 2는 거짓공격 오경보 데이타를 분류하기 위한 과정을 도식화 한 것이다. 마지막 단계에서는 분류할 데이타의 집합을 구축된 분류 모델에 적용시키는 과정으로써 데이타베이스에 저장된 분류 규칙들과의 패턴 매칭을 통해 데이타를 분류한다.

• 실험 데이타를 이용한 분류 모델의 정확도 평가

훈련 데이타를 이용해 생성한 분류 규칙의 정확도를 평가하는 단계로써 훈련데이타의 일부분을 추출하거나 또는 소량의 실험용 데이타를 이용하여 수행한다. 이를 통해 관리자는 생성된 분류 규칙의 정확도 및 과적용 등을 분석하고 필요한 경우 가지치기를 통해 의사결정트리의 정확도를 높인다.

• 실제 테이터를 적용한 데이타 분류

이 단계에서는 분류할 데이타의 집합을 구축된 분류 모델에 적용시키는 과정으로써 데이타베이스에 저장된 분류 규칙들과의 패턴 매칭을 통해 데이타를 분류한다. 결정트리의 규칙 집합은 IF~THEN 형태로 출력된다. 그림 3은 결정 트리 생성 알고리즘이다.

```
Algorithm DCTree()
Input  : the training samples (samples), represented by discrete-valued attributes, attribute list
Attribute List : the set of candidate attributes
Method :
  Call DecideNode()
    create a node N;
    if samples are all of the same class, C then return N as a leaf node labeled with the class C;
    if attribute-list is empty then return N as a leaf node labeled with the most common class
       in samples;
    select test-attribute, the attribute among attribute-list with the highest information gain;
    label node N with test-attribute;
    for each known value a_i of test-attribute grow a branch from node N for the condition test-
       attribute = a_i;
    let s_i be the set of samples in samples for which test-attribute = a_i;
    if s_i is empty then attach a leaf labeled with the most common class in samples;
    else attach the node returned by generate_decision_node ;
Output : Decision Tree
```

그림 3 결정 트리 생성 알고리즘

데이타 분류 클래스는 결정트리 생성 클래스에서 생성된 분류규칙들을 기반으로 실제 데이타를 입력 값으로 받아서 데이타를 분류하는 클래스이다. 데이타베이스에 저장된 각각의 투플에 대해 결정트리 생성 클래스가 만든 규칙들과의 매칭 작업을 수행하여 데이타를 분류한다. 분류하지 못하는 데이타의 발생을 방지하기 위해 하나의 데이타에 대해 분류 규칙 개수만큼 반복적으로 매칭 작업을 수행한다. 분류를 하기 위해 데이타베이스에 저장되어 있는 테이블을 선택하고 분류를 수행하게 되면 규칙집합이 저장되어 있는 규칙 데이타베이스를 불러 각각의 투플들이 이 규칙과 일치하는지를 조건식에 의해 비교한 후 결과값을 출력한다.

데이타 분류 클래스의 알고리즘은 그림 4에서 기술하였다. 이 알고리즘은 실험 데이타를 이용한 결정트리의 정확도 검증 및 실제 데이타를 이용한 오경보의 분류에 사용된다.

Algorithm *False Alert Classifier()*
Input : **the test data, represented by discrete-valued attributes, attribute list**
Attribute List : **the set of candidate attributes**
Method :
Call Classifier()
1. Load the Decision Tree
2. Partition into classification rules of IF-Then structure
3. Counting the number of classification rule
4. Decision Class Compare given input data to a number of classification rules
5. 4 repeat until no more exist data
Output : **Classification Rule (True Positive/False Positive)**

그림 4 오경보 분류 알고리즘

4. 구 현

오경보 분류 시스템은 데이타 전처리 모듈, 속성 선택 모듈, 결정트리 생성 모듈, 데이타 분류 모듈로 구성하였다. 구현된 프로토타입에서, 데이타 전처리 모듈은 이진 데이타로 이루어진 정상행위 및 공격행위 데이타를 관계형 데이타베이스에 저장할 수 있는 형태로 가공한다.

속성 선택 모듈은 데이타베이스에 저장되어 있는 데이타의 속성들 간의 연관규칙 및 상관분석 기반으로 얻을 수 있는 속성들을 데이타베이스에 저장한다. 결정트리 생성 모듈, 데이타 분류 모듈은 실제 데이타를 분류한다. 결정트리 생성 모듈은 훈련 데이타를 입력으로 분류 모델을 생성하고 데이타 분류 모듈은 실험 데이타와 실제 데이타를 입력으로 하여 결정트리 생성 모듈에서 생성한 분류 규칙을 기반으로 데이타를 분류 한다. 최종 결과는 마찬가지로 관계형 데이타베이스에 저장되어 관리자에게 제공되고 이는 이후에 침입탐지 시스템에 적용되어 거짓공격 오경보 감소 모델로 사용하게 된다. 그림 5는 거짓공격 오경보 데이타를 분류하기 위한 시스

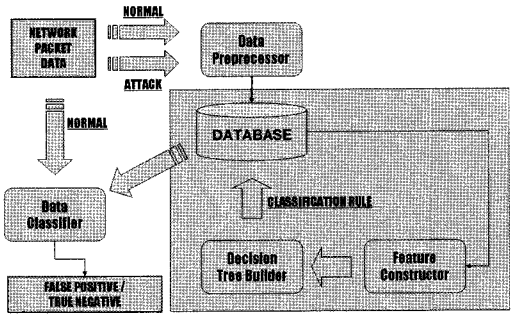

그림 5 오경보 분류 시스템의 구조

템의 구조로 각 구성 요소간의 상호 관계를 보여준다.

오경보 분류 시스템의 구현 환경은 다음과 같다. 프로그래밍 언어는 자바를 이용하였다. 자바는 운영체제에 구애를 받지 않고 컴파일할 수 있는 멀티 플랫폼 언어로 운영체제에 구애를 받는 프로그래밍 언어를 사용할 경우 이종의 운영체제에서 작동하는 침입탐지 시스템과의 통합에 어려움이 많기 때문에 다종의 침입탐지 시스템에 적용 가능한 언어인 자바를 이용하여 구현하였다. 데이타를 저장하는 관계형 데이타베이스로는 오라클 8i를 이용하였고, 분류시스템의 백그라운드에서 구동되는 침입탐지 시스템은 다종의 운영체제에서 사용 가능한 스노트를 사용하였다.

구현된 프로토타입은 3계층으로 구성된다. 결과를 입출력 할 수 있는 사용자 인터페이스 부분과 데이타베이스 접속, 결정트리 생성, 데이타 분류 클래스 등의 프로그램 부분과 데이타 저장소인 관계형 데이타베이스가 있다. 구현된 프로토타입에서는 분류모델 구축 및 데이타 분류 등과 관련하여 거의 모든 자료의 저장을 관계형 데이타베이스를 이용하였다.

데이타 분류 클래스는 결정트리 생성 클래스에서 생성된 분류규칙들을 기반으로 실제 데이타를 입력값으로 받아서 데이타를 분류하는 클래스이다. 데이타베이스에 저장된 각각의 투플에 대해 결정트리 생성 클래스가 만든 규칙들과의 매칭 작업을 수행하여 데이타를 분류한다. 분류하지 못하는 데이타의 발생을 방지하기 위해 하나에 데이타에 대해 분류 규칙 개수 만큼 반복적으로 매칭 작업을 수행한다. 그림 6은 구현된 프로토타입의 인터페이스 부분이다. 훈련데이타를 이용하여 오경보 분류 모델을 생성하고 실제 데이타로 분류를 수행하며 그 결과를 볼 수 있다.

5. 실험 및 평가

5.1 평가 항목

이 장에서는 구현된 분류 시스템을 통하여 제안한 기법의 유용성을 실험하여 평가한다. 이 논문의 목적은 결

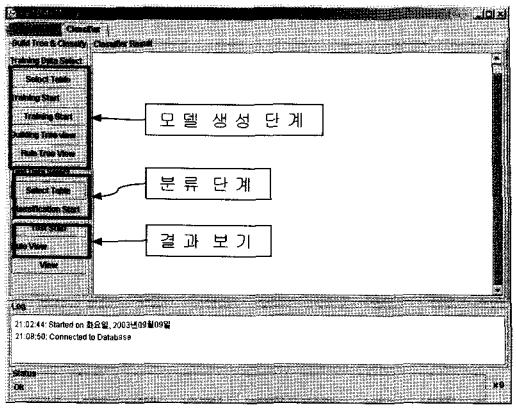

그림 6 오경보 분류 모델 프로토타입

정 트리를 이용하여 기존의 침입탐지 시스템에서 발생하는 거짓공격 오경보 데이타를 감소시켜 침입탐지 시스템의 성능향상을 가져오는 것이므로 분류 시스템의 위치는 침입탐지 시스템이 네트워크 패킷을 추출하기 이전의 위치에서 동작하며 분류시스템이 정상행위로 판별한 패킷을 침입탐지 시스템에 통과시킴으로써 탐지율을 측정한다. 또한 이전에 언급했던 것처럼 훈련 및 테스트 데이타로써 원시 패킷 데이타를 사용하였고, 탐지의 정확도를 높이기 위해 공격의 카테고리는 서비스 거부 공격으로 제한하였다. 따라서 제안한 결정트리 기법을 다음의 평가 항목에 따라 수행 성능을 분석하였다.

- 속성 선택 방법의 차이에 따른 노드의 변화
- 속성 선택 방법의 차이에 따른 분류 모델의 정확도
- False Positive 감소에 따른 침입탐지 시스템의 성능평가

5.2 실험 데이타

평가를 위해 이 논문에서 사용한 실험 데이타는 1998년 달파 데이타 집합이다[16]. 이 실험 데이타는 총 7주간의 네트워크 패킷으로 구성된 티시피덤프 데이타로 이를 훈련 데이타로 사용하여 오경보 분류모델을 생성하였다. 이 데이타 집합은 약 5,000,000개의 데이타 인스턴스로 구성되어 있으며 네트워크 환경 상에서 가능한 다양한 형태의 침입을 포함하고 있다. 이 중 공격으로 명시된 데이타 중 서비스 거부 공격들만을 따로 추출하였으며, 정상행위 데이타의 추출은 공격데이타의 전체 크기와 동일한 양의 패킷을 추출하였다. 분류모델의 구축을 위해 추출된 데이타 중 1-4주까지의 데이타를 훈련 데이타로 이용하였으며, 탐지율 측정을 위해 5-7주까지의 데이타를 실험 데이타로 사용하였다. 또한 결정트리의 가지 수를 줄이기 위해서는 속성들의 값을 이산적인 형태로 변환시켜주어야 할 필요가 있다. 특정 값들이 가중치가 있는 속성(목적지/근원지 포트번호)에 대해서는 가중치가 있는 값들은 변환하지 않고 그 이외의 값들에 대해서만 변환을 수행하였고, 가중치가 없는 연속적인 값들로 이루어진 속성들은 속성의 크기에 따라 변환하였으며, 7개 이하의 특정 값들로 이루어진 속성에 대해서는 변환을 수행하지 않았다.

5.3 평가

이 절에서는 5.1절에 기술한 평가 항목에 대한 실험을 통해 제안한 오경보 감소 기법의 유용성에 대해 평가한다.

실험 1. 속성 선택 방법에 따른 트리 노드의 변화

결정트리는 데이타를 정확히 분류 할 수 있으면서 크기(깊이, 노드 수 등)가 작을수록 바람직하다. 따라서 여러 개의 속성 중에서 결정 트리를 형성함에 있어 불필요한 속성을 제거하고 필요한 속성을 선택하여야 한다. 속성을 선택하는 첫 번째 방법은 통계적인 방법 중의 하나인 상관관계 분석을 통한 방법이다. 표 3은 각 속성에 대한 상관계수표이다. 이 표에서 볼 수 있듯이 각각의 속성들이 -0.535부터 1까지의 상관관계를 가지고 있음을 알 수 있고, 이들 중 비교적 강한 상관관계를 보이는 속성들은 FSI(Fragment Size), ILN(Ip Length), IOF(Ip Offset), FOS(Fragment Offset), SEQ(Sequence Number)이다.

다음은 연관규칙 기반의 속성 선택 방법으로써 지지도의 변화를 조건으로 빈발한 후보항목을 찾는 방법을 이용한다. 지지도의 변화는 1을 최대값으로 0.5에서 0.95까지로 하였으며, 전체 데이타에 대해 지지도를 0.5%씩 증가시켜 수행하여 선택되는 후보항목의 최대 길이는 표와 같고 이를 그래프로 도식화 하면 그림 7과 같다.

그림 7에서처럼 지지도가 0.5일 때는 최대 길이가 8인 라지 항목이 생성되었고 지지도가 0.95였을 때는 최대 길이가 3인 빈발 라지 항목이 생성되었다. 따라서 지지도의 평균값인 0.75에 해당하는 4개의 속성을 결정트리의 노드로 결정하였고, 이 노드의 속성 값은 (1) TTL (Time to Live) (2) ACK(Acknowledgement), (3) TLN (Tcp Length), (4) FLG(Flag)이다. 위와 같은 두 가지 방법으로 선택되어진 속성들을 이용하여 결정트리의 노드로 설정하여 분류모델들을 구축한 후 다음 실험에서 분류모델의 정확도 측정을 위해 사용될 것이다.

실험 2. 속성 선택 방법의 차이에 따른 분류 모델의 정확도

이 실험에서는 실험 1에서 수행한 두 가지의 속성 선택 방법을 이용하여 결정트리의 노드를 결정한 후 두 개의 분류모델을 생성하여 5-7주까지의 데이타를 사용하여 생성된 분류 모델의 정확도를 자체 평가한다. 상관관계 분석을 통하여 얻어진 속성으로 생성된 결정트리의 규칙은 그림 8과 같으며 그림 9에서는 연관규칙 기반의 속성선택의 결과 생성된 결정트리를 보여준다.

표 3 상관 계수표

		SPO	DPO	TTL	ILN	FOS	FSI	IOF	SEQ	ACK	WIN	TLN	URP	FLG
SPO	Pearson 상관계수	1,000	-,168**	,331**	,180**	-,024**	,180**	-,104**	-,094**	,232**	-,225**	-,019**	-,003	-,023**
	유의확률 (양쪽)	.	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,077	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
DPO	Pearson 상관계수	-,168**	1,000	,005**	-,033**	-,016**	-,033**	,079**	,126**	-,224**	,041**	-,022**	-,001	,138**
	유의확률 (양쪽)	,000	.	,009	,000	,000	,000	,000	,000	,000	,000	,000	,485	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
TTL	Pearson 상관계수	,331**	,005**	1,000	-,220**	-,010**	-,220**	,186**	-,352**	,190**	-,742**	-,032**	,004	,149**
	유의확률 (양쪽)	,000	,009	.	,000	,000	,000	,000	,000	,000	,000	,000	,051	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
ILN	Pearson 상관계수	,180**	-,033**	-,220**	1,000	-,028**	1,000**	-,535**	,000	,100**	,273**	-,105**	-,004**	-,066**
	유의확률 (양쪽)	,000	,000	,000	.	,000	,000	,000	,872	,000	,000	,000	,039	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
FOS	Pearson 상관계수	-,024**	-,016**	-,010**	-,028**	1,000	-,028**	,598**	-,046**	-,044**	-,107**	-,513**	,000	-,122**
	유의확률 (양쪽)	,000	,000	,000	,000	.	,000	,000	,000	,000	,000	,000	,937	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
FSI	Pearson 상관계수	,180**	-,033**	-,220**	1,000**	-,028**	1,000	-,535**	,000	,100**	,273**	-,105**	-,004**	-,066**
	유의확률 (양쪽)	,000	,000	,000	,000	,000	.	,000	,869	,000	,000	,000	,039	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
IOF	Pearson 상관계수	-,104**	,079**	,186**	-,535**	,598**	-,535**	1,000	-,040**	-,024**	-,106**	-,446**	,003	,205**
	유의확률 (양쪽)	,000	,000	,000	,000	,000	,000	.	,000	,000	,000	,000	,096	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
SEQ	Pearson 상관계수	-,094**	,126**	-,352**	,000	-,046**	,000	-,040**	1,000	-,148**	,289**	,085**	,001	,019**
	유의확률 (양쪽)	,000	,000	,000	,872	,000	,869	,000	.	,000	,000	,000	,750	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
ACK	Pearson 상관계수	,232**	-,224**	,190**	,100**	-,044**	,100**	-,024**	-,148**	1,000	-,308**	-,080**	-,002	,156**
	유의확률 (양쪽)	,000	,000	,000	,000	,000	,000	,000	,000	.	,000	,000	,274	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
WIN	Pearson 상관계수	-,225**	,041**	-,742**	,273**	-,107**	,273**	-,106**	,289**	-,005**	1,000	-,147**	-,002	,252**
	유의확률 (양쪽)	,000	,000	,000	,000	,000	,000	,000	,000	,008	.	,000	,193	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
TLN	Pearson 상관계수	-,019**	-,022**	-,032**	-,105**	-,513**	-,105**	-,446**	,085**	-,080**	-,147**	1,000	-,001	-,220**
	유의확률 (양쪽)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.	,759	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
URP	Pearson 상관계수	-,003	-,001	,004	-,004*	,000	-,004*	,003	,001	-,002	-,002	-,001	1,000	,032**
	유의확률 (양쪽)	,077	,485	,051	,039	,937	,039	,096	,750	,274	,193	,759	.	,000
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732
FLG	Pearson 상관계수	-,023**	,138**	,149**	-,066**	-,122**	-,066**	,205**	,019**	,156**	,252**	-,220**	,032**	1,000
	유의확률 (양쪽)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	.
	N	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732	272732

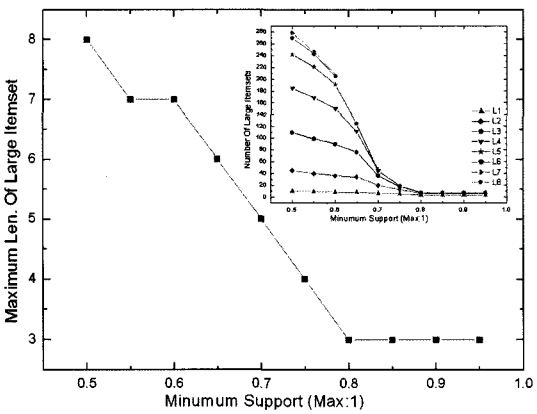

그림 7 지지도에 따른 라지 항목의 최대 길이

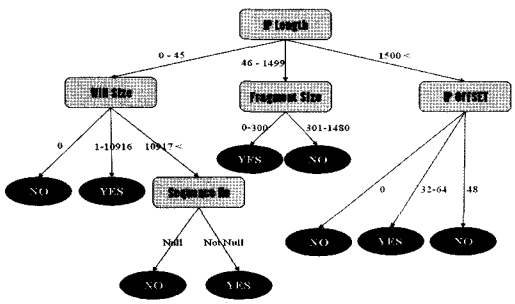

그림 8 상관분석 기반의 결정트리

각각의 결정트리에 5주 100,000개, 6주 46,650개. 7주 55,150개의 데이타를 입력값으로 분류한 결과인 정상패킷의 탐지율은 상관분석 기반의 결정트리의 경우 29.79%,

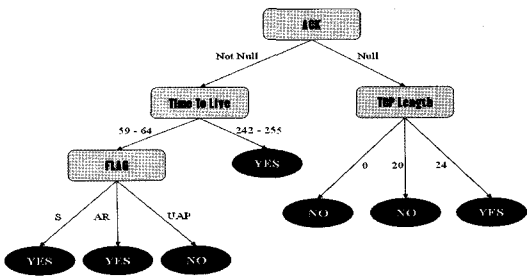

그림 9 연관규칙 기반의 결정트리

연관규칙 기반의 경우는 37.37%의 평균값을 보였다. 각 주간 데이타의 탐지율은 그림 10에서와 같이 나타나는데 전반적으로 연관규칙 기반의 결정 트리의 탐지율이 상관분석 기반의 결정트리에 비해서 높다는 걸 알 수 있다.

또한 전체적인 탐지율이 50%미만의 낮은 탐지율을 보이고 있는 데에서 착안하여 탐지율을 높일 수 있는 방법으로 두 개의 결정트리를 결합한 모델을 사용하여 실험해 보았다. 하나의 트리를 통과하여 분류된 패킷 데이타를 대상으로 나머지 결정트리에 한번 더 통과시키는 방법을 이용하여 측정하였는데 결과는 평균 48.32%의 탐지율을 보였다. 예상대로 결합모델 쪽이 다른 두 개의 결정트리 보다는 높은 탐지율을 보였지만 두 개의 결정트리 탐지율 합에는 미치지 못하는 결과가 나타났다.

이는 이미 첫 번째 트리에서 분류된 패킷 데이타와 두 번째에서 분류된 데이타간의 중복에 의한 것인데 이를 통해 중복되지 않는 형태의 속성들을 갖는 결정트리를 구성한다면 좀 더 높은 탐지율을 보일 수 있다는 가

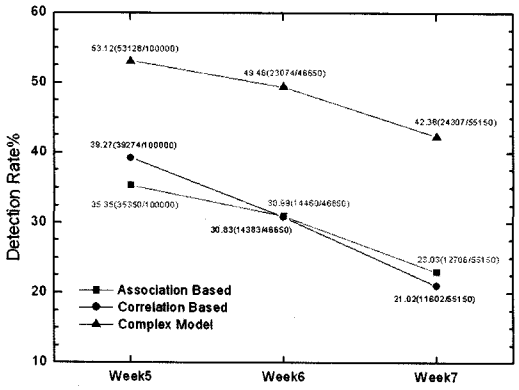

그림 10 분류모델별 정상패킷 탐지율

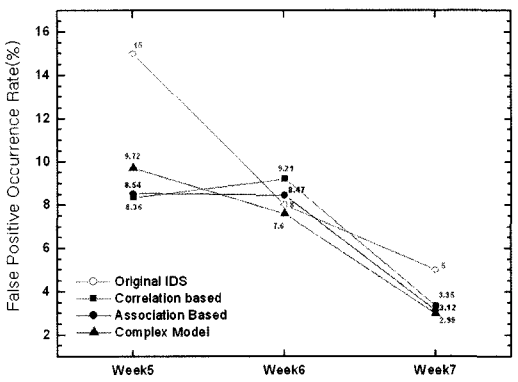

그림 11 분류모델별 오경보 발생 비율

능성을 발견할 수 있다.

실험 3. 오경보 감소에 따른 침입탐지 시스템의 성능 평가

이 실험은 최종 단계로써 결정트리 모델을 통과한 패킷 데이타가 침입탐지 시스템의 거짓공격 오경보 생성에 어느 정도의 유용함을 가지는지를 평가하는 실험이다. 침입탐지 시스템으로는 스노트1.8.6을 이용하였으며 입력 데이타는 실험 2와 동일한 데이타를 사용하였다. 그림 11은 결정트리를 통과하기 이전과 이후의 오경보 탐지율의 변화이다. 전반적으로 결정트리 기반 분류 모델을 통과한 이후의 오경보 발생율이 낮게 나왔으나 6주차 데이타의 경우 결합모델을 제외한 두 개의 모델은 그렇지 못했다. 특징적인 결과를 분석하였을 때 5 주차에서는 침입탐지 시스템이 공격으로 오인하는 데이타의 대부분이 근원지 주소와 목적지 주소가 동일한 랜드공격이기 때문에 침입탐지 시스템의 오경보 생성율이 다른 주차에 비해 높았다. 그 이유는 랜드 공격이란 몇몇 시스템에서 TCP/IP 구현상의 문제로 인하여 수신하는 SYN 제어패킷의 출발지주소와 목적지 주소가 해당 패킷을 수신하는 시스템의 IP주소를 가지는 경우 이를 제대로 처리하지 못하고 멈추게 되는 현상이기 때문이다.

즉 침입탐지 시스템은 근원지/목적지 주소가 동일하면 무조건 랜드공격으로 탐지하는 반면 훈련 데이타의 속성 중에는 근원지/목적지 주소가 선택되어 있지 않았다. 더구나 위와 같은 속성을 제외한 다른 속성들만으로도 결정트리 기반 분류 모델은 침입탐지 시스템에 비해 훨씬 정확한 분류를 할 수 있었다.

6. 결 론

네트워크 기반 침입탐지 시스템은 패킷데이타를 분석하여 침입의 행위를 판단하여 관리자에게 경보를 전달하며 이 경보의 양은 네트워크의 광역화와 해킹 기법의 발달로 인해 기하급수적으로 증가하고 있다. 경보데이타의 증가로 생겨나는 문제점은 잘못된 경보의 발생을 초래하는데 이중 거짓 공격 경보 데이타가 잘못된 경보의 대부분을 차지하고 있다. 따라서 침입탐지 시스템에서 다량으로 발생하여 시스템의 부하를 가져오는 오경보를 감소시킴으로써 침입탐지 시스템의 성능 향상을 위한 방안으로 이 논문에서는 데이타 마이닝 기법중 결정트리를 이용한 오경보 분류 기법에 대해서 제안하였다.

기본적인 속성들로 이미 테이블 화 되어진 데이타를 사용하며 모든 공격에 대해 프로파일을 생성하여 분류하는 기존 데이타 집합의 문제점을 보완하기 위한 방법으로 바이너리 형태로 이루어진 네크워크 패킷 데이타에서 가시적으로 확인할 수 없는 속성값들을 추출하는 전처리 과정을 수행 한 후, 공격의 범위를 서비스 거부 공격으로 한정시켜 분류할 수 있는 분류모델을 제안하였다. 그리고 제안된 오경보 분류 모델을 구현하였고 이 모델의 검증을 위해 1998년 달파데이타를 이용하여 실험 평가를 수행하였다. 실험 결과 침입탐지 시스템에 적용할 경우 서비스 거부 공격에 대한 오경보 발생 비율이 감소됨을 확인할 수 있었다. 그러나 서비스 거부공격 이외의 공격에 대한 오경보를 해결하기 위해서는 각각의 공격별 분류 모델이 생성되어야 한다.

그리고 결정트리 기법을 사용하는 경우 트리의 깊이에 따라 정확도와 비용간의 상충적 문제가 발생한다. 이 문제를 해결하기 위하여 먼저 데이타 전처리 과정에서 연속적인 속성 값들을 이산적인 형태로 변환하였으며 여러 속성들중 분류 모델에 유용하게 사용될 분류 속성들을 선택하기 위해서 이 논문에서는 분류 속성 선택을 위해 연관규칙 기반의 빈발 항목과 통계적 상관관계 분석을 이용한 접근방안을 제시하였다. 그리고 연관규칙 기반의 속성 선택과 통계적 상관관계 분석을 이용한 속성 선택 방법 각각에 대하여 정확도 측정 실험을 하였으며 두 방법을 혼용한 경우와도 비교평가를 수행하여 세 방법중 혼용 방법이 유용하다는 것을 확인하였다.

이 논문에서는 속성 값들의 변환을 위해 관리자의 발견적인 방법을 이용하였지만 알려지지 않은 데이타의 값들을 변환할 때는 발견적인 방법은 유용하지 못하다는 단점이 있다. 따라서 발견적인 방법이 아닌 자동화된 형태로 향상시켜 경보 분석 및 통합관리에 활용할 수 있는 연구가 향후 계속되어야 한다.

참 고 문 헌

[1] D. Anderson, T. Frivold, and A. Valdes, "Next Generation Intrusion Detection Expert System (NIDES)," Technical Report SRI-CSL-95-07, 1995.

[2] R.G. Bace, "Intrusion Detection," Macmillan Technology, 2000.

[3] W. Lee, S. J. Stolfo, "Data Mining Approaches for Intrusion Detection," In Proceedings of the 7th USENIX Security Symposium, 1998.

[4] W. Lee, Salvatore J. Stolfo and K. W. Mok, "Mining Audit Data to Build Introduction Detection Models," In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, 1998.

[5] W. Lee and S. J. Stolfo, "A Data Mining Framework for Building Intrusion Detection Models," Columbia University, 2001.

[6] M.V. Joshi, R.C. Agarwal, V. Kumar, "Mining Needles in a Haystack: Classifying Rare Classes via Two-Phase Rule Induction," ACM SIGMOD 2001.

[7] E. Bloedon, et al., "Data Mining for Network Intrusion Detection: How to get Started," 2001.

[8] A. Valdes and K. Skinner, "Probabilistic Alert Correlation," In Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection, pages 54-68, 2001.

[9] H. Debar and A.Wespi, "Aggregation and Correlation of Intrusion-Detection Alerts," In Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection, pages 85-103, 2001.

[10] P. Ning and Y. Cui., "An Intrusion Alert Correlator based on Prerequisites of Intrusions," Technical Report TR-2002-01, Department of Computer Science, 2002.

[11] Cuppens, F. Miege, A., "Alert Correlation in a Cooperative Intrusion Detection Framework," In Proceedings of IEEE Symposium on Security and Privacy, 2002.

[12] M. Klemettinen. "A Knowledge Discovery Methodology for Telecommunication Network Alarm Data," PhD thesis, University of Helsinky, 1999.

[13] J. Ross Quinlan, "C4.5: Programs for Machine Learning and Neural Networks," 1993.

[14] Snort. Open-source Network Intrusion Detection System. http://www.snort.org

[15] http://www.tcpdump.org Tcpdump/Libpcap

[16] http://ideval.ll.mit.edu Lincoln Lab MIT. DARPA 2000 Datasets

[17] C. Kruegel, T. Toth, "Using Decision ITrees to Improve Signature-based Intrusion Detection," In Proceedings of the 4th nternational Symposium on Recent Advances in Intrusion Detection, 2003.

[18] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical Automated Detection of Stealthy Portscans," In Proceedings of ACM Computer and Communications Security IDS Workshop, 2000.

[19] M.S. Shin, H.S.Moon, K.H. Ryu, J.O.Kim and K.Y.Kim, "Applying Data Mining Techniques to Analyze Alert Data," APWeb'03, LNCS 2642, pp.193-200, SpringerVerlag, 2003.

[20] Moon Sun Shin, Keun Ho Ryu, "Data Mining Methods for Alert Correlation Analysis," IJCIS, Vol.4, No.4, pp.225-234, 2003.

[21] M.S. Shin, K.J. Jeong, "Alert Data Mining Framework for Intrusion Detection System," WISA'05, LNCS3786, SpringerVerlag, 2005.

[22] 신문선, 문호성, 류근호, 장종수, "클러스터링기법을 이용한 침입탐지시스템의 경보상관관계분석", 정보처리학회 논문지C 제10-C권, 제6호, pp.665-674, 2003.

[23] 신문선, 김은희, 문호성, 류근호, 김기영, "데이타마이닝기법을 이용한 경보 분석기구현", 정보과학회 논문지, 제31권, 제1호, pp.1-12, 2004.

신 문 선

1988년 충북대학교 전산통계학과(학사) 1997년 충북대학교 전자계산학과(석사) 2004년 충북대학교 전자계산학과(이학박사). 2005년~2005년 8월 가림정보기술 선임연구원. 2005년 9월~현재 건국대학교 컴퓨터시스템전공 강의교수. 관심분야는 시공간데이타베이스, 데이타마이닝, 데이타베이스보안, USN보안

류 근 호

1976년 숭실대학교 전산학과(이학사). 1980년 연세대학교 공학대학원 전산전공(공학석사). 1988년 연세대학교 대학원 전산전공(공학박사). 1976년~1986년 육군군수지원사 전산실(ROTC 장교), 한국전자통신 연구원(연구원), 한국방송통신대 전산학과(조교수) 근무. 1989년~1991년 Univ. of Arizona Research Staff(TempIS 연구원, Temporal DB). 1986년~현재 충북대학교 전기전자 및 컴퓨터공학부 교수. 관심분야는 시간 데이타베이스, 시공간 데이타베이스, Temporal GIS, 지식기반 정보검색 시스템, 유비쿼터스컴퓨팅 및 스트림데이타 처리, 데이타마이닝, 데이타베이스 보안, 바이오인포메틱스