침입탐지시스템의 비대칭 오류비용을 이용한 데이터마이닝의 적용전략

홍태호[a], 김진완[b]

[a]부산대학교 상과대학 경영학부
부산시 금정구 장전동 산30, 609-735
Tel: +82-51-510-2531, Fax: +82-51-581-3144, E-mail:hongth@pusan.ac.kr

[b]동부산대학교 상과대학 경영학과
부산시 금정구 장전동 산30, 609-735
Tel: +82-51-510-2531, Fax: +82-51-581-3144, E-mail: kimjw@pusan.ac.kr

Abstract

최근 들어 네트워크 침입탐지시스템은 정보시스템 보안에서 매우 중요하게 인식되고 있다. 네트워크침입시스템에 데이터마이닝 기법들을 활용하는 연구들이 활발하게 그동안 활발하게 진행되어 왔다. 하지만 단순한 데이터마이닝 기법의 적용만으로는 침입탐지시스템의 효과를 극대화 할 수 없다. 침입탐지시스템은 오류의 종류에 따라 조직에 미치는 영향이 매우 상이한 특징을 갖는다. 따라서 본 연구에서는 침입탐지시스템의 오류의 특징에 따른 각기 다른 데이터마이닝 기법을 적용하는 방안을 제시하였다. 또한 국내에서 사용된 실제 네트워크를 통한 침입공격에 관한 데이터를 수집하고, 신경망, 귀납적 학습법, 러프집합을 적용하여 국내 데이터 특성을 고려한 네트워크 침입탐지모형을 제시하였다.

Keywords:

침입탐지시스템, 데이터마이닝, 보안

서론

오늘날 인터넷 사용의 폭발적 증가로 대부분의 시스템이 타 시스템과 네트워크로 연결되어 있는 상황 하에서 악의적인 해킹 또는 네트워크 침입은 그 시스템을 운영하는 조직뿐만 아니라 네트워크에 연결되어 있는 타 기관들에게도 치명적인 손해를 입힐 수 있는 구조를 갖고 있다. 따라서 정보시스템의 네트워크 환경의 급속한 발달로 인한 역기능을 줄이기 위한 네트워크 침입탐지시스템의 필요성이 강조되고 있다. 네트워크 환경 하에서 정보시스템을 운영하는 회사, 학교, 정부 기관 등은 침입탐지시스템을 운영하여 외부의 해커 또는 공격자로부터 정보자원을 보호하고 있다.

기존의 네트워크 침입탐지시스템에 사용되는 침입탐지모형은 전문가들의 지식을 이용한 네트워크 침입자 또는 해커의 행위를

탐지하여 비정상적인 정보시스템에 대한 접근을 제한하는 형태가 일반적이다. Zhu(2001) 등은 네트워크 침입탐지분야에서 데이터마이닝 기법이 매우 우수한 성과를 보이는 것으로 보고했다. 데이터마이닝 기법을 이용한 침입탐지모형은 기존의 데이터를 이용하여 새로운 패턴을 발견할 수 있다는 장점이 있다.

그러나 침입탐지시스템의 오류에 따른 조직에 미치는 영향은 서로 상이하다. 이러한 오류의 특성에 데이터마이닝 기법의 적용을 위한 방법이 필요하며, 국내 네트워크 환경에 적합한 침입탐지시스템의 개발이 또한 필요하다. 따라서 본 연구에서는 최근 매우 중요하게 인식되고 있는 정보시스템 보안의 한 분야인 네트워크 침입탐지시스템에 데이터마이닝 기법을 적용하여 국내 특성에 맞는 침입탐지시스템을 개발하고자 한다.

이론적 배경

침입탐지시스템

침입이란 컴퓨터 시스템의 자원들에 대한 허가되지 않은 접속 또는 사용을 말한다(Esmaili et al., 1996). 정보자원을 보호하기 위한 침입탐지시스템(Intrusion detection system)에 대한 정의는 다음과 같이 다양하게 되고 있다. 침입탐지시스템은 목표 시스템에 대한 허가되지 않았거나 변칙적인 활동들을 탐지하고, 식별하고, 대응하는 기능을 가진 소프트웨어이다(Richards, 1999). 침입탐지시스템의 목적은 실시간으로 또는 일괄처리 방식으로 보안 위반을 탐지하기 위한 메커니즘을 제공하는 것이다(Debar et al., 1992). 위반은 시스템을 파괴하려고 시도하는 외부인들에 의해 일어나거나, 권한을 오용하기 위해 시도하는 내부인들에 의해 일어난다(Weber, 1999). 침입탐지시스템은 다양한 시스템과 네트워크 자원들로부터 정보를 수집한 뒤에 침입과 오용에 관한 신호를 보내기 위해 정보를 분석한다(Lippmann and Cunningham, 2000). 침입탐지시스템에 의해 수행될 수 있는 주요한 기능으로는 사용자와 시스템 활동을 모니터링하고 분석하며, 중요한 시스템과 데이터 파일들의 무결성을 평가하며, 알려진 공격들을 반영한 활동 패턴들을 인식하며, 탐지된 활동에 자동적으로 대응하며, 그리고 탐지 프로세스의 결과를 보고한다. 이러한 침입탐지시스템은 전자상거래, 교육기관, 은행 등의 금융기관, 일반회사의 인트라넷 등에 다양하게 적용될 수 있다.

표 1- 침입탐지시스템의 적용영역 및 기대효과

적용영역	기대효과
전자상거래 영역	데이터 무결성(Integrity)와 기밀성(Confidentiality) 확보 사용자 프라이버시 보호
은행, 증권사 등의 금융기관	쇼핑몰 등의 시스템 보안성 향상
교육기관	고객 DB 등의 보호
기업 인트라넷	시스템에 대한 비권한자 활동(unauthorized activity) 보호 등

침입탐지는 탐지 방법에 따라서 크게 오용 탐지(misuse detection)와 비정상적 탐지(anomaly detection)의 두 가지 범주로 구분할 수 있다(Zue et al., 2001). 첫째, 오용 탐지는 잘 알려진 공격들의 증적이나 패턴을 검색하는 방법으로 탐지한다. 오직 특징적인 증적을 남긴 잘 알려진 공격들만이 이 방법으로 탐지될 수 있다. 둘째, 비정상적 탐지는 정상적 사용자나 시스템 행동의 모델을

사용하고, 악의적인 가능성이 있는 경우에는 정상적인 사용과 편차가 발생하는 지를 탐지한다. 정상적 사용자나 시스템 행동에 관한 이 모델은 일반적으로 사용자 또는 시스템 프로파일로서 알려져 있다. 비정상적 탐지의 주요한 강점은 사전에 알려지지 않은 공격들을 탐지하기 위한 능력이 있다는 것이다.

침입탐지시스템은 분석하는 감사 데이터 출처의 종류에 따라서도 분류되어질 수 있다(Joo et al., 2003). 대부분의 침입탐지시스템은 공격들을 인지하거나 피하기 위한 접근법으로 네트워크 기반 침입탐지 또는 호스트 기반 침입탐지로 구분된다. 침입탐지시스템이 네트워크 트래픽에서 패턴을 찾는 경우에는 네트워크 기반 침입탐지로 분류된다. 침입탐지시스템이 로그 파일에서 공격 흔적을 찾는 경우에는 호스트 기반 침입탐지로 분류된다.

침입탐지시스템에 대한 최근의 많은 접근법들은 데이터마이닝 기술들을 활용하고 있다(Lam et al., 1996). 이러한 접근법들은 시스템에 의해 수집된 감사 추적의 대형 데이터 셋에 데이터마이닝 기술들을 적용하는 방법으로 탐지 모델을 구축한다 (Helman and Liepins, 1993). 데이터마이닝 기반 침입탐지시스템은 시스템의 일부분을 감시하는 감지장치(sensor)로부터 데이터를 수집한다. 감지장치들은 네트워크 활동, 사용자 프로세스에 의해 사용되는 시스템 호출, 그리고 파일 시스템 접속을 감시한다. 감지장치들은 탐지를 위해 사용되어질 수 있는 형식화된 데이터를 만들기 위해 수집된 원본 데이터로부터 예측적인 특징들을 추출한다. 감지장치에 의해 수집된 데이터들은 탐지기술을 사용하는 탐지기에 의해 평가되어진다. [표-2]는 침입탐지시스템을 위한 데이터마이닝 응용 연구를 보여주고 있다.

표 2- 침입탐지시스템에서의 데이터마이닝 응용사례

분류기준	침입탐지시스템 유형	데이터마이닝 기법
탐지 방법	오용탐지	CBR(Esmaili et al., 1996) NN(Endler, 1998) NN(Cannady, 1999) GA(Balajinath and Raghavan, 2001)
	비정상적 탐지	NN(Debar et al., 1992) NN(Bonifacio et al., 1998) NN(Endler, 1998) GA(Balajinath and Raghavan, 2001)
감사 자료 출처	네트워크 기반 탐지	NN(Kumar and Venkataram, 1997) NN(Endler, 1998) NN(Bonifacio et al., 1998) GA(Sinclair et al., 1999) NN(Lippmann and Cunningham, 2000)
	호스트 기반 탐지	CBR(Esmaili et al., 1998) NN(Heatley and Otto, 1998) GA(Balajinath and Raghavan, 2001)

지식채굴 프로세스

지식채굴은 데이터에서 유효하고, 새롭고, 유용한 그리고 궁극적으로 이해할 수 있는 패턴을 확인하는 과정이다(Fayyad et al., 1996). 지식채굴 과정에 대한 완전한 방법론은 존재하지는 않지만 지식채굴은 <그림 1>과 같은 단계를 거쳐 발견된다고 할 수 있다.

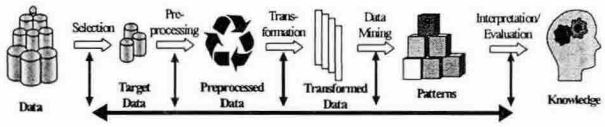

그림 1 - 지식채굴 프로세스의 개요

지식채굴을 위해서는 먼저 데이터와 업무를 파악하고 분석을 위한 적당한 환경으로

옮겨서 데이터를 사용되어질 형태로 통합시키고 검토하게 된다. 이때 데이터에서 이상치와 에러를 제거하여 데이터를 정리하고 적용하는 기법에 따른 모형을 세우고 이를 검증하기 위한 기본 가정을 세운다. 이 가정을 검증하기 위해 데이터마이닝 기법을 적용하여 패턴이나 지식을 채굴해 나간다. 여기서 채굴된 지식의 유용성을 검증하고, 새로 발견된 지식의 사용을 위해 해석을 하게 된다. 지식채굴을 수행하기 위해서는 통계적 방법론과 인공지능기법을 이용하는 방법 등이 사용되고 있는데, 본 연구에서는 인공지능기법들 중에서 신경망, 귀납적 학습법, 러프집합을 적용하도록 한다.

(1) 인공신경망

일반적인 인공신경망은 다층퍼셉트론 (Multi layer perceptron)이라 불리우며, 다층퍼셉트론은 입력층과 출력층 사이에 하나 이상의 중간층이 존재하는 신경망을 지칭하는 것이다. 이 때, 입력층과 출력층 사이의 중간층을 은닉층(hidden layer)이라 하며 network는 입력층, 은닉층, 출력층으로 연결되어 있다. 다층퍼셉트론에서의 가중치는 지속적으로 전체 신경망이 만족할 만한 목표에 도달할 때까지 변하게 된다. 즉 인공신경망을 통해 계산된 출력값과 목표출력값(output)을 비교하여 그 차이(오차합수)를 최소화시킬 수 있도록 지속적으로 가중치를 조정하는 것이다.

이러한 신경망의 가중치의 조절은 역전파 알고리즘에 의한 학습과정을 통해 이루어진다. 역전파 학습 알고리즘의 기본 원리는 다음과 같다. 입력층의 각 유니트에 입력패턴을 주면, 이 신호는 각 유니트에서 변환되어 중간층에 전달되고 최후에 출력층에서 신호를 출력하게 된다. 이 출력값과 기대값을 비교하여 차이를 줄여 나가는 방향으로 연결강도를 조절하고, 상위층에서 역전파하여 하위층에서는 이를 근거로 다시 자기층의 연결강도를 조정해 나가게 된다. 인공신경망에 대한 자세한 내용은 Rumelhart & McClelland(1986)을 참고한다.

(2) 러프집합(Rough Set)

러프집합 이론은 1982년에 Pawlak에 의해서 부정확한 것, 모호한 것 그리고 부확실한 것에 대한 새로운 수학적 접근법으로 제안되었다(Pawlak, 1999). 러프집합 이론은 세상에 모든 개체들은 그들이 가진 어떤 정보로서 집합을 지울 수 있다는 가정에서 설립되었다. 동일한 정보에 의해 특징지워지는 개체들은 그들의 정보로 인해 동일한 것으로 취급된다. 이 방법에서 생성된 동질성 관계(indiscernibility Relationship)가 러프집합 이론의 수학적 기초가 된다.

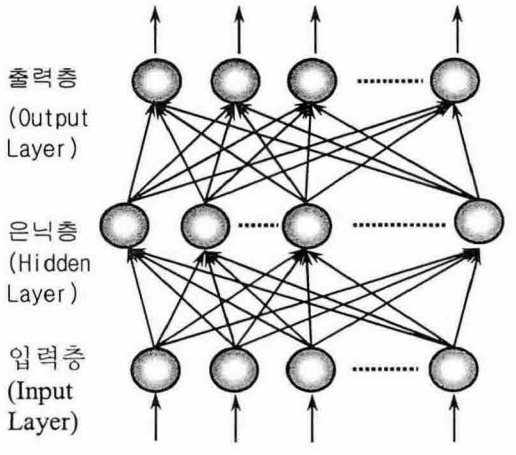

그림 2-인공신경망 모형

어떤 동질성을 가진 집합을 기본집합(Elementary Set)이라 하고, 이들이 어떤 전체집합에 대한 지식의 기본 단위가 된다. 어떤 기본집합의 합집합이 되는 집합을 일반집합(Crisp Set)이라 하고 그 외의 경우를 러프집합(Rough Set)이라고 한다. 결과적으로

각 러프집합은 개체들이 집합의 구성요소로써 명확하게 분류되지 않는 경우(상위근사) 또는 완전하게 분류되는 경우(하위근사) 등과 같은 경계영역개체(boundary-line cases)를 가진다.

러프집합 이론은 지식표현시스템(knowledge representation systems), 동질성 관계(indiscernible relations), 집합의 근사(approximation of set), 속성의 종속성(dependency of attributes), 속성의 축소(reduction of attributes), 그리고 의사결정 규칙(decision rules)으로 특징지워질 수 있다(Slowinski and Zopounidis, 1995; Pawlak, 1999; Zue, et al., 2001).

러프집합의 주요한 장점은 데이터에 대한 사전 또는 추가적인 정보가 요구되지 않는다는 것이다. 이 방법론은 데이터에 숨겨진 중요한 요인을 발견할 수 있고, 의사결정 규칙을 자연어를 통해 표현할 수 있다. 또한 거대한 양의 정성적 및 정량적 데이터 모두를 다루는 능력을 제공하고, 비선형적이거나 비연속적인 기능적 관계를 모델화하는 능력은 복잡하고 다차원적인 패턴을 위한 강력한 방법론을 제공한다. 따라서 러프집합은 지식 획득, 예상과 예측 모델링, 그리고 의사결정 지원에서 성공적으로 적용되었다(Slowinski and Zopounidis, 1995; Pawlak, 1999; Zue, et al., 2001).

연구모형

본 연구에서 제안하는 연구모형은 <그림 2>와 같이 인공신경망, 귀납적 학습법, 러프집합을 이용한 침입탐지모형을 개발하였다. 각 모델의 성과로 정상적인 사용자를 잘못 탐지하여 사용제한을 시키고 이에 대한 대응책을 실행하게 하여 일어나는 기회비용인 False positive error와 악의적인 침입자를 정상 사용자로 분류하여 시스템에 대한 접근을 허용함으로써 발생되는 정보 시스템 자산의 피해인 False negative error를 사용하였다. 추가적으로 기업의 상황에 맞는 침입탐지 전략 선정을 위해서 분류기준값(Treshold)을 0.3~0.7로 조정하면서 False positive error와 False negative error 의 변화에 따른 성과분석을 제시하였다. 이를 통해 보안담당자는 기업 환경에 적합한 최적의 비용으로 침입탐지 전략을 설정할 수 있다.

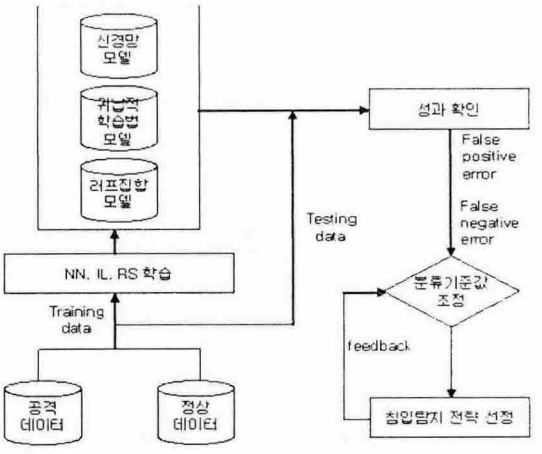

그림 3 - 지능형 침입탐지시스템 연구모형

실험 설계

실험에 사용된 데이터는 통합 보안 서비스를 제공하는 Cyber-PATROL사의 IDS 센스로부터 무작위로 정상 패턴과 공격 패턴을 포함한 500개의 사례를 수집하였다. 이 표본은 다시 Training data와 Testing data로 분할하였다. 침입탐지를 위한 신경망, 귀납적 학습방법, 러프집합의 입력 변수들은 문헌 연구 및 전문가와의 인터뷰를 통해 <표 3>과 같이 선정하였다.

표 3 - 모델에 사용된 변수

변수	설명
Event Data	기록된 사건의 날짜와 시간
Protocol ID	사건과 관련된 프로토콜
Source Port	출처의 포트 숫자
Destination Port	도착지의 포트 숫자
Source IP Address	출처의 IP 주소
Destination IP Address	도착지의 IP 주소

표 4 - 신경망, 귀납적 학습법, 러프집합에 따른 성과표(오류 %)

모델	데이터셋	S1[1]	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg.
NN[2]	학습용	13.78	15.11	15.33	15.56	16.22	15.33	14.22	15.56	15.56	15.56	15.38
	검증용	18.00	26.00	6.00	20.00	14.00	20.00	14.00	22.00	22.00	10.00	15.38
IL[3]	학습용	16.00	14.22	17.11	17.33	12.89	16.44	15.56	16.22	15.56	16.89	15.82
	검증용	18.00	26.00	8.00	6.00	18.00	14.00	22.00	16.00	22.00	10.00	16.00
RS[4]	학습용	15.56	14.22	16.00	16.22	14.67	16.22	14.67	15.11	15.33	16.22	15.42
	검증용	14.00	26.00	8.00	6.00	22.00	12.00	20.00	24.00	22.00	10.00	16.40

1) S1 – S10, 10-fold 데이터 셋, 2)신경망, 3)귀납적 학습방법, 4)러프집합

본 연구에서는 침입탐지를 위한 데이터마이닝 기법으로 인공신경망, 귀납적 학습법, 러프집합의 3가지 실험을 진행하였다. 3가지 실험은 모두 10-fold cross validation으로 수행하기 위해서 무작위로 서로 다른 데이터를 생성하였다. k-fold cross-validation은 전체 표본을 k개의 테스트 집합으로 분할하여 실험하는 것이다.

데이터마이닝 기법간 성과

침입탐지를 위한 3가지 데이터마이닝 기법을 10-fold cross validation으로 실험한 결과는 <표 4>과 같다. 각 데이터 셋마다 성과가 조금씩 차이를 보이고 있지만 전체 평균으로 보면 3가지 기법은 성과에 큰 차이가 없는 것으로 보인다. 3가지 실험에 대해 변화의 유의성을 실제 통계적으로 검증하기 위해 멕네마르 검정(McNemar test)을 수행하였다. 멕네마르 검정은 사전사후(before and after) 형태의 실험에서 변화의 유의성을 검정하는 비모수적 방법이다. 이러한 변화의 유의성을 검정하기 위해 자료는 동일한 개체에서 두 가지 처리를 적용하여 나타나며, 2×2 분할표로 표현이 된다. 침입탐지를 데이터마이닝 기법간의 성과를 검정한 결과는 <표 5>과 같다. 멕네마르 검정의 결과를 살펴보면 모두가 유의수준 0.05에서 유의하지 않으므로 신경망, 귀납적 학습법, 러프집합의 성과차이는 있다고 할 수 없다.

표 5 - 멕네마르 검정 결과표

	IL		RS	
NN	통계량 (S)	0.1111	통계량 (S)	0.6923
	Pr > S	0.7389	Pr > S	0.4054
IL	.		통계량 (S)	0.2857
			Pr > S	0.5930

오류분류에 따른 성과변화 분석

일반적으로 데이터마이닝 기법에서 이진 의사결정을 위한 분류기준값(threshold)으로 0.5를 사용하고 있다. 그러나 이러한 획일적인 분류기준값보다는 각 분류기준값별로 성과 분석을 통해 최대 이익을 가져올 수 있는 특정 분류기준값을 발견할 수 있다. 분류기준값의 변화에 따른 False positive error(FPE)와 False negative error(FNE) 결과의 변화는 <표 4>와 같다.

표 6 - 신경망과 러프집합의 성과변화비교

분류	NN			RS			맥네마르(p)
기준값	FPE	FNE	전체	FPE	FNE	전체	
0.3	26.00	6.40	16.20	29.60	5.60	17.60	0.07*
0.4	25.20	6.80	16.00	27.20	6.00	16.60	0.37
0.5	24.40	7.20	15.80	26.80	6.00	16.40	0.41
0.6	22.40	8.40	15.40	26.80	7.20	17.00	0.07*
0.7	20.40	24.00	22.20	22.80	20.40	21.60	0.75

* 유의수준 10%

위 결과에서 귀납적 학습방법은 제외하였다. 그 이유는 귀납적 학습방법의 경우에 예측값들이 0 과 1 에 가깝게 동일하므로 분류기준값을 0.3~07 로 조정하더라도 결과가 모두 똑같이 나타나기 때문이다. 신경망과 러프집합의 분류기준값에 따른 멕네마르 검정 결과를 살펴보면 0.3 과 0.6 이 유의수준 10%에서 통계적으로 유의한 결과를 나타내고 있다. 이것은 신경망과 러프집합의 성과에 차이가 있음을 보여주는 것이다. 분류기준값 0.3 과 0.6 에서 신경망이 러프집합 보다는 False positive error 를 더욱 잘 탐지하고 있기 때문에 False negative error 에서 러프집합에 비해 성과가 조금 낮지만 전체 성과는 높은 것으로 나타나고 있다. 이렇게 분류기준값을 통한 변화를 제공함으로써 보안담당자들은 자사의 환경에 맞추어 False positive error 와 False negative error 에 대한 최적의 비율로 유연하게 침입탐지 전략을 설정할 수 있을 것이다.

결론

오늘날 정보시스템 네트워크 환경의 급속한 발달로 인한 역기능을 줄이기 위한 네트워크 침입탐지시스템(Intrusion Detection System : IDS)의 필요성이 강조되고 있다. 기존의 네트워크 침입탐지시스템에 사용되는 침입탐지모형은 전문가들의 지식을 이용한 네트워크 침입자 또는 해커의 행위를 탐지하여 비정상적인 정보시스템에

대한 접근을 제한하는 형태가 일반적이다. 그러나 이러한 네트워크 침입탐지시스템은 규칙기반(Rule-based)으로 구성되어 나날이 발전되고 있는 네트워크 공격기술에 적절한 대응력이 부족한 실정이다.

그러나 네트워크 침입탐지분야에서 데이터마이닝 기법은 매우 우수한 성과를 보이는 것으로 보고했다. 데이터마이닝 기법을 이용한 침입탐지모형은 기존의 데이터를 이용하여 새로운 패턴을 발견할 수 있다는 장점이 있다. 따라서 본 연구에서는 국내에서 사용된 실제 네트워크를 통한 침입공격에 관한 데이터를 수집하고, 3가지 데이터마이닝 방법론(신경망, 귀납적 학습법, 러프집합)을 적용하여 국내 데이터 특성을 고려한 네트워크 침입탐지모형을 설계하였다.

침입탐지를 위한 3가지 데이터마이닝 기법을 10-fold cross validation으로 실험한 결과로 도출된 성과는 유사한 것으로 나타났다. 그러나 일반적으로 데이터마이닝 기법에서 이진 의사결정(binary decision)을 위한 분류기준값(threshold)으로 0.5를 사용하고 있기 때문에 이러한 획일적인 분류기준값보다는 각 분류기준값별로 False positive error, False negative error 그리고 성과 분석을 통해 최대 이익을 가져올 수 있는 특정 분류기준값을 발견할 수 있게 된다.

본 연구에서 설계한 지능형 침입탐지시스템은 분류기준값을 통한 변화를 제공함으로써 기업의 보안담당자들이 자사의 환경에 맞추어 False positive error와 False negative error에 대한 최적의 비율로 유연하게 침입탐지 전략을 설정할 수 있도록 의사결정을 지원할 것이다.

References

[1] 박기남, 이훈영, 박상국. (2000). 러프집합을 이용한 통합형 채권등급 평가모형 구축에 관한 연구. 한국경영과학회지, 제25권, 제3호, 125-135.

[2] Balajinath, B., Raghavan, S.V. (2001). Intrusion detection through behavior model. Computer Communications, 24, 1202-1212.

[3] Barber, R. (2001). The Evolution of Intrusion Detection Systems-The Next Step. Computer&Security, 20, 132-145.

[4] Debar, H., Becker, M., & Siboni, D. (1992). A neural network component for an intrusion detection system. IEEE Computer Society Symposium Research in Security and Privacy, 240-250.

[5] Fayyad, U.M., Piatesky-Shapiro, G., & Smith, P. (1996). The KDD processes for extracting useful knowledge and learning from volumes of data. Communications of the ACM, 39(11), 27-34.

[6] Joo, D., Hong, T., & Han, I. (2003). The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors. Expert Systems with Applications, 25, 69-75.

[7] Lippmann, R. P., Cunningham, R. K. (2000). Improving intrusion detection performance using keyword selection and neural network. Computer Networks, 34, 597-603.

[8] Pawlak, Z. (1999). Rough set approach to knowledge-based decision support. European Journal of Operational Research, 48-57.

[9] Zue, D., Premkumar, G., Zhang, X., & Hsien Chu. (2001). Data Mining for Network Intrusion Detection : A Comparison of Alternative Methods. Decision Sciences, 32(4), 635-659.