실시간 침입탐지를 위한 자기 조직화 지도(SOM)기반 트래픽 속성 상관관계 메커니즘

황 경 애[†] · 오 하 영[††] · 임 지 영[†††] · 채 기 준[††††] · 나 중 찬[†††††]

요 약

네트워크 기반의 공격은 그 위험성과 피해의 규모가 크기 때문에 공격 초기에 빨리 탐지하는 것이 중요하다. 그러나 지도학습 데이터 마이닝을 이용한 네트워크상의 비정상 트래픽을 탐지하는 방법은 방대한 양의 데이터 전처리와 관리자의 분석이 요구되며 관리자의 분석이 정확하다는 보장이 없을 뿐만 아니라 각 네트워크의 실시간 특성을 고려하지 못하기 때문에 탐지의 어려움이 크다. 본 논문에서는 실시간 침입 탐지와 점진적 학습을 위해 비지도학습의 데이터마이닝 기법중 하나인 자기 조직화 지도를 기반으로 트래픽 속성 상관관계 메커니즘을 제안한다. 이는 세 단계로 이루어진다. 첫 번째 단계는 초기 학습이 이루어지는 단계로 비지도 학습을 통하여 성격이 비슷한 트래픽끼리 클러스터링 한 맵을 생성시킨다. 두 번째 단계는 맵의 각 클러스터가 정상과 비정상 트래픽의 클러스터로 구분되기 위해 각 공격별로 추출된 규칙(rule)을 적용하여 맵을 분석한다. 이 규칙은 지도 학습을 통한 규칙 기반의 방법으로, 각 데이터 항목마다 SOM을 이용한 속성별 맵의 상관관계(correlation) 분석을 통해 생성되었다. 마지막으로 분석된 맵을 이용하여 실시간 탐지와 함께 점진적 학습이 이루어지게 된다. 여러 실험을 통하여 비지도 학습과 지도 학습을 결합한 SOM 기반 트래픽 속성 상관관계 메커니즘이 지도 학습에 비해 실시간 탐지에 우수함을 증명하였다.

키워드 : 네트워크, 침입 탐지, 자기 조직화 지도, 클러스터링, 비지도 학습, 지도학습

Traffic Attributes Correlation Mechanism based on Self-Organizing Maps for Real-Time Intrusion Detection

Kyoungae Hwang[†] · Hayoung Oh[††] · Jiyoung Lim[†††] · Kijoon Chae[††††] · Jungchan Nah[†††††]

ABSTRACT

Since the Network based attack is extensive in the real state of damage, It is very important to detect intrusion quickly at the beginning. But the intrusion detection using supervised learning needs either the preprocessing enormous data or the manager's analysis. Also it has two difficulties to detect abnormal traffic that the manager's analysis might be incorrect and would miss the real time detection. In this paper, we propose a traffic attributes correlation analysis mechanism based on self-organizing maps(SOM) for the real-time intrusion detection. The proposed mechanism has three steps. First, with unsupervised learning build a map cluster composed of similar traffic. Second, label each map cluster to divide the map into normal traffic and abnormal traffic. In this step there is a rule which is created through the correlation analysis with SOM. At last, the mechanism would the process real-time detecting and updating gradually. During a lot of experiments the proposed mechanism has good performance in real-time intrusion to combine of unsupervised learning and supervised learning than that of supervised learning.

Key Words : Network, Intrusion Detection, Self-organizing Maps, Clustering, Unsupervised Learning, Supervised Learning

1. 서 론

초기 침입 탐지 시스템들은 이미 알려진 공격에 대한 정보를 수동적으로 시스템에 인코딩하여 침입 여부를 판단하는 방법으로 규칙의 생성 및 확장이 매우 어렵고 그 효율성도 매우 떨어지는 방법이다. 따라서 인공지능, 기계 학습 및 데이터 마이닝 기법들을 침입 탐지에 적용하는 연구가 늘어나는 추세이나 아직까지 많은 연구가 분류(classification) 방법을 포함한 지도 학습(supervised learning) 알고리즘에 근간을 두고 있어 다음과 같은 문제들을 가지고 있다. 먼저 학습과 침입 탐지 과정이 확연히 구분되어 있고 탐지 과정 전에 충분한 학습 과정이 이루어져야 하므로 안정적인 성능이 나오

※ 본 논문은 정보통신부 정보통신연구진흥원에서 지원한 ITRC 프로그램 및 한국전자통신연구원 정보보호연구단 위탁연구과제의 연구결과입니다.
† 정 회 원 : 삼성전자
†† 준 회 원 : 이화여자대학교 컴퓨터학과 석사과정
††† 정 회 원 : 한국성서대학교 정보과학부 전임강사
†††† 종신회원 : 이화여자대학교 컴퓨터학과 교수
††††† 정 회 원 : ETRI 능동보안기술연구팀 팀장
논문접수 : 2005년 5월 27일, 심사완료 : 2005년 9월 1일

기까지 많은 비용이 든다. 그리고 학습을 위해 많은 양의 분류된 데이터(labeled data)를 필요로 하는데, 이러한 방대한 양의 학습 데이터를 수집하고 분류하는 것은 매우 어려운 일이며 학습 데이터의 질에 의해 탐지 성능이 크게 좌우된다[1][2][3]. 또한 실시간으로 네트워크 성격을 반영하는 점진적 학습의 수행이 불가능하고 학습된 데이터 이외의 새로운 침입 유형에 대한 탐지가 어렵다. 대안으로 비지도 학습(unsupervised learning)의 데이터 마이닝 기법을 이용한 연구가 진행되었으나 비지도 학습만 사용했을 때 발생하는 문제점은 학습 시 입력 데이터에 대해 어떠한 정보도 주지 않으므로 그 결과에 대한 해석이 힘들다는 것이다.

이 모든 것을 고려하여 본 논문에서는 비지도 학습인 자기조직화 지도(SOM)와 지도 학습인 상관관계를 결합하여 점진적 학습 및 실시간 탐지가 가능한 SOM기반 트래픽 속성 상관관계 메커니즘을 제안하고 이를 비정상 실시간 침입탐지에 활용한다. 즉 비지도 학습 SOM은 점진적 학습과 실시간 탐지가 가능하지만 학습결과 지도 해석이 힘들다는 문제점이 있기 때문에 이를 해결하기 위해 분류되어 있는 데이터를 사용하여 속성간의 상관관계를 분석하여 규칙을 생성하고 이를 기반으로 결과에 대한 정보를 알아낸다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어서 2장에서는 관련 연구를 살펴본다. 3장에서는 SOM기반 트래픽 속성 상관관계 분석에 사용할 실험 데이터와 SOM을 제공하는 도구에 대해 살펴본다. 4장에서는 실시간 침입탐지를 위한 SOM기반 트래픽 속성 상관관계 메커니즘을 제안하고, 5장에서는 제안한 탐지 메커니즘을 다양한 측면에서 실험한 내용과 결과를 기술한다. 마지막으로 6장에서는 본 연구의 결론 및 향후 연구 계획에 대해서 기술한다.

2. 관련 연구

최근 네트워크 기술 및 서비스 발달과 함께 이에 대한 공격 또한 다변화되고 있다. 이들 공격에 대한 예방책이 최우선이지만, 완벽한 예방이 불가능하므로 공격 발생 시 실시간으로 공격을 탐지하여 차단하는 방법이 필수적이다. 이러한 취지에 맞춰 기존에 제안된 탐지 기법으로는 근원지에서의 공격 탐지 기법, 통계적 기법을 이용한 공격 탐지 기법, 데이터 마이닝 기술을 이용한 공격 탐지 기법 등이 있다.

근원지에서의 공격 탐지 기법으로 로스앤젤레스 소재 캘리포니아 대학교에서 개발한 'D-WARD'(DDoS netWork Attack Recognition and Defense)라는 기법은 보안 소프트웨어를 네트워크 게이트웨이에 설치하고, 게이트웨이를 통해 밖으로 나가는 트래픽에 대한 감시를 집중하는 방안이다[10]. 그러나 짧게 반복되는 공격에 대해서는 이전 공격을 메모리에 저장해 놓지 않기 때문에 매번 계산해야 하는 점이 비효율적이고, TCP와 ICMP 트래픽과 달리 역방향 트래픽이 없는 UDP 트래픽의 경우, 공격을 판단하는 기준 변수가 달라 탐지에 한계가 있다.

통계적 방법을 이용한 공격 탐지 연구에서는 공격 도구에 의해 생성된 공격 트래픽들은 정상 트래픽과 구별되는 특징을 갖고 있으며, 통계적인 기준을 이용하여 중심 라우터에서 정상과 공격 트래픽을 구별할 수 있다고 가정하였다[11]. 그러나 갈수록 공격 도구가 지능화 되면서 스푸핑의 랜덤 정도를 조절 가능하게 되고, 이로 인해 정상과 공격의 소스 주소 분포를 구분 짓는 것이 어려워지고 있다. 그리고 다양한 유형의 공격들이 존재하기 때문에 단순히 소스 주소만을 모니터링 하는 것은 모든 공격 유형을 탐지해 내는데 충분하지 못하다.

데이터 마이닝 기법은 일반적인 침입탐지를 위하여 다양한 연구가 진행되어 왔다. Wenke Lee[12]는 침입탐지를 위하여 sendmail과 tcpdump 데이터를 이용해 frequent episode 분류 기법과 연관규칙 기법을 통해 침입 모델을 생성하여 이를 시험하였고, 데이터 마이닝 적용 시 많은 모델링 시간이 걸리는 점을 고려하여 학습(learning) 에이전트와 탐지 에이전트로 구성된 침입 탐지 구조를 제안하였다.

의사결정트리, 신경망 모형의 지도학습 데이터 마이닝 기법을 이용한 공격 탐지 방법은 정상 데이터와 각 공격별 입력 데이터들을 초기 입력 자료로 사용하여 속성 추출 단계와 분류기 생성 단계를 거치는 공격 탐지 모델이다[4].

하지만 이런 기존의 지도 학습 데이터 마이닝을 기반으로 하는 탐지 기법들은 트래픽의 패턴을 학습하기 위한 트레이닝 단계에서 각 트래픽 데이터를 정상과 비정상으로 분류한 레이블링된 데이터가 필요하다. 그러나 관리자가 모든 트래픽을 정상과 비정상으로 구분해서 레이블링 해주는 것이 정확하지 않을 뿐더러 방대한 양의 데이터를 분석해야 하는 엄청난 작업량을 요구하므로 레이블링된 데이터를 수집하기가 쉽지 않다. 또한 트레이닝 단계와 탐지 단계가 확연히 구분되어 있어서 학습의 점진적 갱신이 이루어지지 않아 현 상태의 네트워크 특성이 정확히 반영된 탐지가 이루어지지 않다. 마지막으로 트레이닝 단계에서 학습 데이터에 포함되지 않은 새로운 유형의 공격이 발생하였을 때에는 탐지가 불가능하다. 이러한 이유로 실제 비정상 트래픽 탐지에서 비효율적이고 실시간 탐지가 불가능하다는 단점이 있다. 이외에도 침입탐지의 두 분류인 이상탐지와 오용탐지(Misuse Detection)를 위해 신경망을 사용한 연구[13]와 새로운 공격을 인식할 수 있도록 신경망 기반의 침입탐지 시스템[14]이 제안되었다.

3. 실험 데이터 및 SOM을 제공하는 도구

3.1 실험 데이터

본 논문에서는 1998 DARPA Intrusion Detection Evaluation Program에 의해 표준 데이터 집합을 얻기 위하여 미국 군사 네트워크상에서 시뮬레이션을 통해 만들어진 KDD Cup 1999 Data 데이터 집합 중 일부인 KDD_TND 와 KDD_TD를 실험데이터로 이용하였다. KDD 데이터는 크게 단일 TCP 연결의 Basic 속성, Content 속성 그리고 2초간의 타임 윈도우에 의해 계산된 Traffic 속성인 세 개의 카테고리로 구분되어진다[5].

이 논문에서는 실시간 탐지를 위해 네트워크 연결 기반의 공격 탐지에 초점을 맞추었으므로 Basic 속성(1번-9번)과 Traffic 속성(10번-28번)을 사용하여 총 28개, 3개의 기호 형(symbolic) 속성과 25개의 숫자 형(numeric) 속성을 사용한다. 이는 <표 1>과 같다.

〈표 1〉 데이터 속성 설명

번호	속성 항목	설 명	자료형
1	duration	연결 지속 시간	numeric
2	protocol	프로토콜 (tcp, udp, icmp)	symbolic
3	service	서비스 종류(http, ftp etc.)	symbolic
4	flag	정상 또는 에러 플래그	symbolic
5	src_byte	소스로부터의 데이터 크기	numeric
6	dst_byte	목적지로부터의 데이터 크기	numeric
7	land	1:소스 주소=목적지 주소, 0	numeric
8	wrong_fragment	"wrong" fragment 개수	numeric
9	urgent	urgent 패킷 개수	numeric
10	count	같은 호스트 상에서 2초간 접속 시도 회수	numeric
11	srv_count	같은 호스트 상에서 2초간 서비스 요구 회수	numeric
12	serror_rate	"SYN" 에러율	numeric
13	srv_serror_rate	서비스 "SYN" 에러율	numeric
14	rerror_rate	"REJ" 에러율	numeric
15	srv_rerror_rate	서비스 "REJ" 에러율	numeric
16	same_srv_rate	접속 중 같은 서비스 요청 비율	numeric
17	diff_srv_rate	접속 중 다른 서비스 요청 비율	numeric
18	srv_diff_host_rate	다른 호스트 접속 비율	numeric
19	dst_host_count	목적지 호스트 회수	numeric
20	dst_host_srv_count	목적지 서비스 회수	numeric
21	dst_host_same_srv_rate	목적지 호스트상 같은 서비스 회수	numeric
22	dst_host_diff_srv_rate	목적지 호스트상 다른 서비스 회수	numeric
23	dst_host_same_src_port_rate	목적지 호스트상 같은 소스 포트 회수	numeric
24	dst_host_srv_diff_host_rate	목적지 서비스상 다른 호스트 비율	numeric
25	dst_host_serror_rate	목적지 호스트 "SYN" 에러율	numeric
26	dst_host_srv_serror_rate	목적지 호스트 서비스 "SYN" 에러율	numeric
27	dst_host_rerror_rate	목적지 호스트 "REJ" 에러율	numeric
28	dst_host_srv_rerror_rate	목적지 호스트 서비스 "REJ" 에러율	numeric

3.2 SOM을 제공하는 도구

3.2.1 SAS Enterprise Miner

SAS Enterprise Miner는 대부분의 데이터 마이닝 기법들을 제공하는 고차원적인 데이터 마이닝 도구이다. 입력 데이터의 선정(sample), 탐색(explore), 변형(modify), 모델 생성(model), 평가(assess)라는 각 범주별로 여러 가지 방법들을 제공해준다.

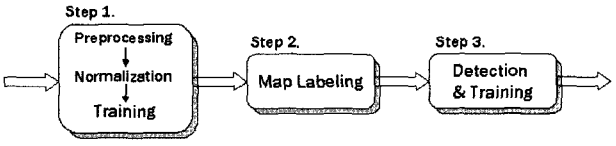
본 연구에서는 SOM/Kohonen 노드를 이용하여 실험하였고 사용자는 분석 방법의 설정, 자기 조직화 지도의 크기 설정, 군집수의 지정, 이웃 노드의 학습률, 학습 방법 선택 등의 옵션을 다양하게 지정할 수 있다[6].

3.2.2 Matlab 기반의 SOM Toolbox

SOM Toolbox는 Helsinki University of Technology에서 연구목적으로 SOM의 쉬운 사용을 위해 개발한 프리웨어의 SOM 프로그램 패키지로서 특히 시각화 측면에서 아주 뛰어나다[7]. 데이터의 전처리, 초기화와 학습, 맵 크기의 설정, 맵의 시각화 및 분석 등 다양한 기법을 제공하며 숫자형의 변수들만 적용된다는 특징이 있다.

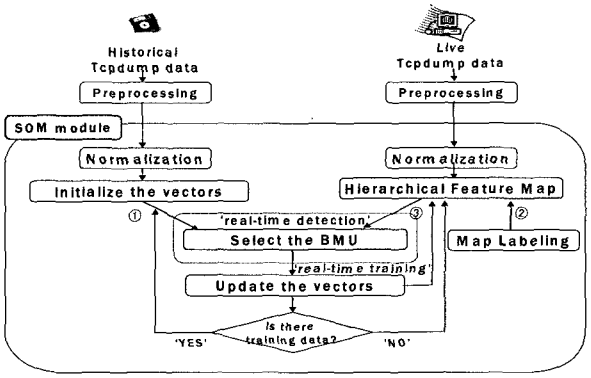
4. 자기 조직화지도(SOM)기반 트래픽 속성 상관관계 메커니즘

실시간 비정상 침입 탐지를 위해 앞에서 설명한 연결 기반의 트래픽 정보와 비지도 학습의 데이터 마이닝 기법인 SOM을 이용하여 트래픽 속성 상관관계를 분석해 보았다. 트래픽 속성 상관관계는 서로 다른 각 트래픽의 성격을 반영함으로 궁극적으로 실시간 침입탐지 메커니즘에 사용될 수 있다. 즉 정상 트래픽과 비정상 트래픽은 서로 다른 트래픽 속성 상관관계를 보이기 때문에 침입탐지가 가능하며 또한 다양한 종류의 비정상 트래픽들의 각 특징을 비교해보고 분석해 볼 수 있다. SOM기반 트래픽 속성 상관관계분석을 이용한 실시간 침입 탐지 메커니즘은 (그림 1)처럼 크게 3단계로 이루어진다. 첫 번째는 전처리와 정규화 된 실험데이터로 탐지에 필요한 맵을 생성하는 학습 단계, 두 번째는 학습된 맵에서 트래픽 속성 상관관계를 활용한 각 클러스터별 분류단계, 마지막으로 실시간 탐지와 점진적 학습이 이루어지는 단계이다.



(그림 1) 탐지 메커니즘

각 단계에서 동작하는 구체적인 탐지 모델은 (그림 2) 와 같다.



(그림 2) 제안한 공격 탐지 모델

4.1 학습과정(Training)

4.1.1 전처리 과정(Preprocessing)

학습 단계 전에 탐지에 효과적인 정보로 데이터를 변형해 주는 전처리 단계가 필요하다. 이 단계에서는 데이터가 앞에서 설명한 연결 기반의 28개 속성 항목을 갖는 중요 속성 추출 단계와 기호 형 데이터는 숫자 형 데이터로 변환되는 단계가 포함된다.

4.1.1.1 중요 속성 추출

탐지 성능은 입력 데이터의 어떠한 속성을 이용하여 탐지하는지에 따라서도 확연한 차이를 보이므로 탐지에 필요한 중요 속성 추출도 중요한 연구 부분이다. 기존 연구[4]에서는 데이터 마이닝의 의사결정트리를 이용하여 탐지에 필요한 중요 속성을 추출하였다. 그러나 이렇게 추출된 중요 속성은 학습한 데이터에 한해서만 효과적이고 학습 데이터가 아닌 다른 데이터에서는 그 성능이 떨어진다. 또한 포함된 공격 유형, 공격별 데이터 수, 전체 데이터의 개수에 따라 의사결정트리로 추출된 속성과 분류 기준이 되는 속성의 값이 다르기 때문에 일반적인 적용이 불가능하다는 것을 발견하였다.

따라서 본 연구에서는 KDD 데이터 집합 중 KDD_TND와 KDD_TD를 사용하여 Basic 속성을 입력 값으로 했을 때, Taffic 속성을 입력 값으로 했을 때, 그리고 이 두 종류의 속성을 모두 포함하여 실험을 했을 때 중에서 가장 성능이 좋은 경우를 알아보았다. 성능 평가는 침입 탐지 분야에서 기본적인 탐지 능력을 평가하기 위한 항목으로 탐지율과 False Positive, False Negative를 사용했다.

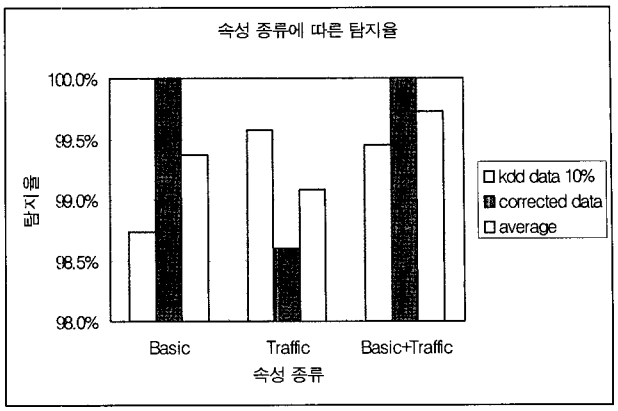
$$\text{탐지율} = \frac{\text{비정상으로 정확히 판정된 비정상 데이터의 개수}}{\text{전체 비정상 데이터 개수}} \times 100$$

$$\text{False Positive} = \frac{\text{비정상으로 오 판정된 정상 데이터의 개수}}{\text{전체 정상 데이터 개수}} \times 100$$

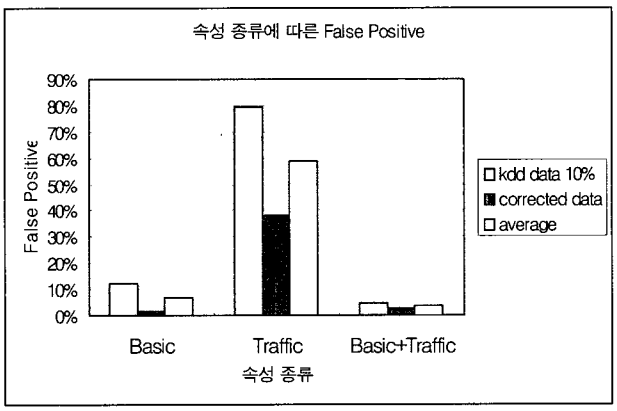
$$\text{False Negative} = \frac{\text{정상으로 오 판정된 비정상 데이터의 개수}}{\text{전체 비정상 데이터 개수}} \times 100$$

분석 결과 KDD_TND에서 탐지율은, Traffic 속성만 했을 경우 99.57%로 가장 높지만 이 경우 False Positive도 79.5%로 너무 높게 나타나므로 가장 성능이 좋다고 말할 수 없으며 그 다음, 탐지율이 좋게 나타난 Basic+Traffic 속성의 경우가 99.45%의 탐지율과 4.68%의 False Positive를 가지므로 Basic 속성을 사용했을 때보다 모든 면에서 더 나은 성능을 보였다. 이는 (그림 3), (그림 4), (그림 5)와 같다.

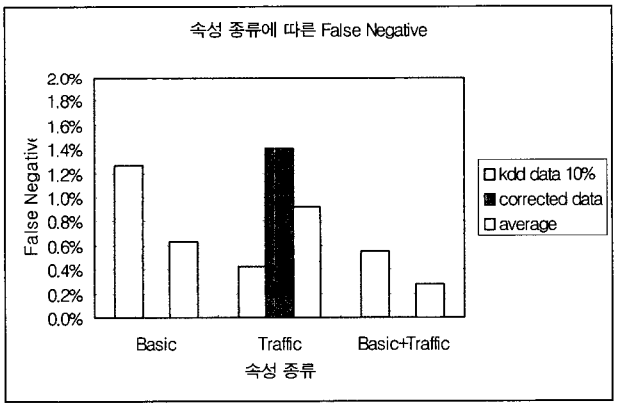
KDD_TD의 실험에서는 Basic 속성을 사용했을 때와 Basic+Traffic 속성을 사용했을 때 같은 100%의 탐지 율을 보였으나 Basic만 사용한 경우 False Positive가 1.34%로 Basic+Traffic 속성의 2.58%보다 낮게 나타나 성능이 더 좋은 것으로 판단된다.



(그림 3) 속성 종류에 따른 탐지율



(그림 4) 속성 종류에 따른 False Positive



(그림 5) 속성 종류에 따른 False Negative

이 두 결과의 평균값을 봤을 때 결과적으로 Traffic만 사용하면 너무 높은 False Positive를 갖게 됨으로 Basic 속성이 반드시 필요하고, Basic 속성과 Traffic 속성을 같이 사용하였을 때 가장 안정적이고 정확한 탐지가 가능한 것으로 보인다. 따라서 이 후 모든 실험에서는 Basic과 Traffic 속성을 모두 고려한 28개의 속성으로 모든 실험 및 분석을 하였다.

4.1.1.2 데이터 타입 변환

실험 데이터에서 기호 형 데이터를 갖는 속성은 3가지의

프로토콜(protocol)과 11개의 플래그(flag) 그리고 80여 가지의 다양한 값을 지닌 서비스(service)가 있다. SOM toolbox는 숫자 형 변수들만 적용되기 때문에 기호 형 데이터를 숫자 형 데이터로 변환해야한다.

4.1.2 정규화 과정(Normalization)

SOM에서는 각 속성 맵들의 형태를 결합하여 모든 속성이 고려된 U-matrix를 최종적으로 생성한다. 그런데 src_byte는 [0, 2194619]의 범위를, duration은 [0, 42448]의 범위를, 비

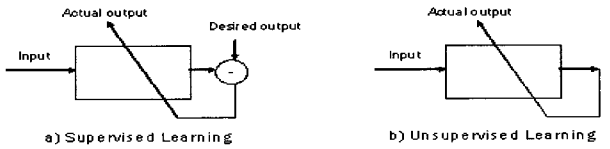
율에 관한 다른 속성 값은 [0, 1]의 범위를 갖는 등 각 속성마다 다양한 범위의 값을 갖는다. 이 상태에서는 다른 속성에 비해 범위가 큰 src_byte나 duration이 U-matrix를 생성하는데 많은 영향을 끼치게 되어 모든 속성을 평등하게 고려한 U-matrix가 생성되지 않으므로 모든 속성이 균등하게 U-matrix 생성에 반영될 수 있도록 데이터 값을 정규화 하는 과정이 필요하다.

정규화는 각 속성들마다 최소값을 0으로($V_{min}(x)\Rightarrow 0$), 최대값을 1로($V_{max}(x)\Rightarrow 1$) 변형하여 모든 값이 일정한 범위인 [0, 1]에 위치하도록 한다. 모든 속성 값이 0부터 1의 값을 갖도록 정규화 하는 식은 다음과 같다.

$$N_{i(x)}=(i(x)-V_{min}(x))/(V_{max}(x)-V_{min}(x))$$

4.1.3 SOM 알고리즘을 이용한 학습 과정(Training)

Kohonen에 의해서 개발된 자기 조직화 지도인 SOM은 신경망 기법을 사용하는 클러스터링의 모델이면서 비지도 학습을 사용한다는 것이 특징이다. (그림 6)은 지도학습과 비지도 학습의 차이를 보여준다. (그림 6)에서 b)의 비지도 학습은 a)의 지도 학습과 달리 주어진 입력에 대해 정확한 해답을 주지 않고 자기 스스로 학습하는 방법이다. 침입 탐지에 있어서는 미리 정상과 비정상으로 분류된 학습 데이터가 필요하지 않고, 분류되지 않은 학습 데이터를 넣어주면 비슷한 성격의 데이터끼리의 클러스터링을 통해 기계 스스로가 정상과 비정상 트래픽으로 분류해준다. 또한 입력 데이터와 가장 가까운 뉴런의 이웃 뉴런들도 비슷한 방향으로 함께 학습시키기 때문에 인접한 뉴런들은 비슷한 성격을 가질 것이라고 예측할 수 있다.



(그림 6) 지도학습과 비지도 학습

SOM의 학습 알고리즘은 5단계로 이루어지며[8], 이는 <표 2>와 같다.

이러한 과정을 통하여 KDD_TND 실험데이터로 (그림 7)과 같은 U-matrix가 형성되었으며 밝은 색은 이웃 클러스터와 구분 지어주는 경계선을 의미한다.

<표 2> SOM의 알고리즘

① 연결가중치의 초기화
$(w_j(n), n=0)$
② 입력 데이터와 모든 출력 뉴런들과의 거리를 계산
$i(x)=arg_{min}\| (x(n)-w_j(n) \|$
③ 최소거리를 가지는 승자뉴런(BMU)을 선택
④ 승자뉴런과 이웃한 출력뉴런에 연결된 가중치들을 갱신
$w_j(n+1)=w_j(n)+\alpha(n)\beta_{i(x)}(n,j)(x(n)-w_j(n))$
⑤ 조건이 만족할 때까지 2단계부터 4단계까지 반복
$w_j(n)$:각 뉴런의 초기 벡터값 $x(n)$:입력 데이터의 벡터값
$i(x)$:승자뉴런(BMU) $\alpha(n)$:학습률 $\beta_{i(x)}$:이웃노드 함수

4.2 분류 과정(Map Labeling)

(그림 7)에서 나타나듯이 학습 단계를 거친 후 생성된 맵은 입력 데이터에 대한 어떠한 정보도 주지 않기 때문에 SOM을 이용해 구분된 클러스터가 정상인지 혹은 비정상인지 사용자가 분간하기 어렵다. 이를 해결하기 위해 각 공격마다 속성별 맵의 유사도를 보고 28개 속성간의 상관관계를 분석하여 상관계수가 높은 속성 집합을 기반으로 규칙을 생성하여 맵의 클러스터 구분이 가능하게 한다.

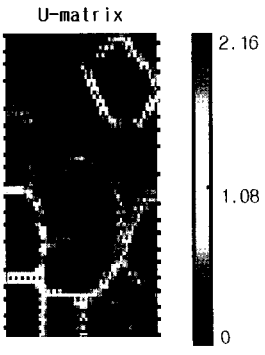
$$\{(x_1, y_1), \cdots, (x_n, y_n)\}$$

두 속성이 어느 정도 연관성이 있는지 상관관계를 분석하기 위해 피어슨(Pearson) 상관계수를 사용하며 주어진 데이터에 대한 상관계수의 공식은 다음과 같다[9].

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$$

$$r_p=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}}$$

속성별 상관관계를 분석하기 위해 이 실험 데이터에 포함된 정상 및 9가지의 공격별 데이터를 분류하여 각각 속성별 맵을 형성하였다. 상관관계에 대한 결과 및 규칙 생성은 다음과 같다.



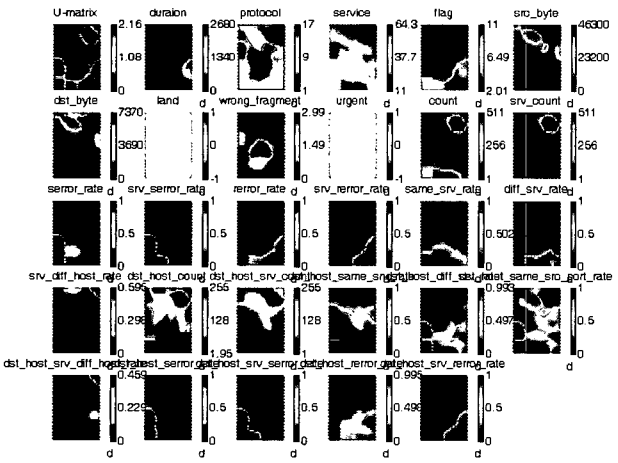
(그림 7) U-matrix

4.2.1 데이터 항목별 속성간의 상관관계를 이용한 맵 분류 규칙 형성

각 클러스터를 구분하는 맵 분류작업을 위해서 트래픽 항목마다 가지고 있는 속성 간 상관관계를 분석하였고 규칙을 생성하여 이를 기반으로 맵에서 위치를 판단하였다. 피어슨 상관계수를 계산하여 일정 수치 이상의 상관관계를 갖는 속성을 파악하고 다른 공격과 구별될 수 있는 상관관계 속성 집합을 추출하여 <표 3>과 같은 규칙을 생성하였다. 이렇게 상관계수의 수치가 높은 속성끼리는 (그림 8)에서 보듯이 SOM의 학습 후 생성된 속성별 맵도 유사한 것으로 나타났다. 규칙은 상관관계를 갖는 속성 집합(correlation)과 특정 속성의 수치 값(value)의 교집합으로 이루어지며 (+)는 양의 상관관계, (-)는 음의 상관관계, (∩)는 교집합을 의미하고 속성은 2장의 <표 1>에서 설명한 속성별 번호로 표기한다.

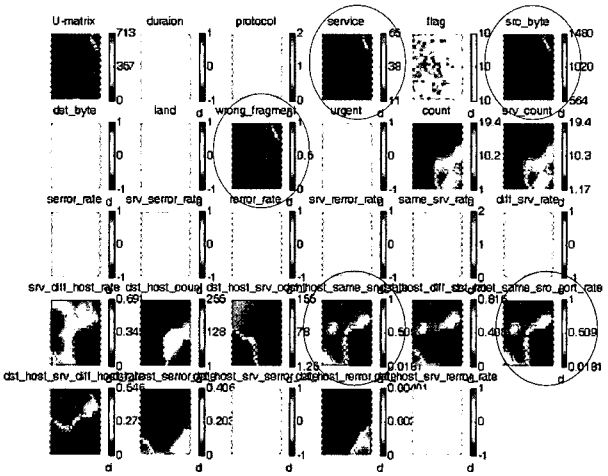
〈표 3〉 데이터 항목별 속성간의 상관관계를 이용한 규칙

Normal	*correlation*	=(*10*+*11*)∩(*12*+*13*)∩(*14*+*15*)∩(*16*-*17*)∩(*20*+*21*)
	value	=none
Neptune	*correlation*	=(*11*+*16*)
	value	=(*2*=6)∩{(*16*=low)∩(*17*=low)}
Smurf	*correlation*	=(*10*+*11*)∩(*19*+*20*)∩{(*21*+*23*)-*22*}
	value	=(*2*=1)∩{(*10*≥511)∩(*11*≥511)}
Teardrop	*correlation*	=(*20*+*21*+*23*)
	value	=(*2*=17)∩(*8*=3)
Back	*correlation*	=(*12*+*13*)∩(*25*+*26*)∩(*27*+*28*)
	value	={(*5*=high)∩(*6*=high)}
Pod	*correlation*	=(*21*+*23*)∩{*3*-(*5*+*8*)}
	value	=(*2*=1)
Ipsweep	*correlation*	=(*3*+*19*+*22*)-(*21*+*23*)
	value	=(*2*=1)
Nmap	*correlation*	={(*4*+*12*+*13*+*26*)-*21*}
	value	=none
Portsweep	*correlation*	=(*22*+*23*+*27*)∩(*15*+*28*)
	value	={(*15*=1)∩(*18*=1)}
Satan	*correlation*	=(*10*+*22*)∩(*14*-*23*)∩(*15*+*28*)
	value	=(*2*=6)



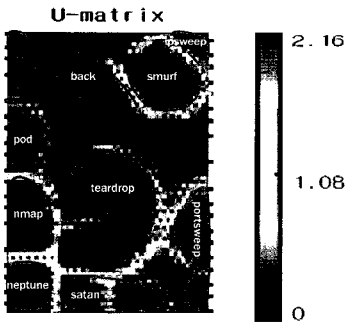
(그림 8) U-matrix와 속성별 맵

(그림 9)는 다른 공격에서 볼 수 없는 Pod 공격 데이터만의 특징을 보여준다. Pod 공격 데이터는 <표 3>의 Pod 속성간의 상관관계 Pod = (*21*+*23*)∩{*3*-(*5*+*8*)} 에서 알 수 있듯이 21번 속성과 23번 속성, 5번 속성과 8번 속성이 양의 상관관계를 가진다. 반면 양의 상관관계를 가지는 5번, 8번 속성과 3번 속성과는 음의 상관관계를 가짐을 알 수 있다.



(그림 9) Pod의 상관관계와 속성별 맵의 유사도

Pod와 같은 방식으로 다른 공격 데이터들도 속성 간 상관관계를 이용한 규칙으로 U-matrix 맵에서 위치 공격 트래픽의 클러스터 위치를 알게 되었다. 또한 공격트래픽의 위치가 아닌 곳은 정상 데이터의 클러스터라고 볼 수 있으며 (그림 10) U-matrix 맵에서 위치가 가까운 공격끼리는 좀 더 공통된 성향을 보이며 상관관계를 갖는 집합도 공통된 것들이 많다는 것을 발견할 수 있었다.



(그림 10) 규칙에 의한 맵 분류

4.3 실시간 탐지 및 점진적 학습 과정(Detection and Training)

학습을 통해 생성된 U-matrix를 이용한 실시간 트래픽 탐지는 SOM 알고리즘의 일부인 유클리디언 거리 측정을 이용하여 이루어진다. 탐지 알고리즘은 <표 4>와 같다.

〈표 4〉 유클리디언 거리측정을 이용한 침입 탐지 알고리즘

BMU=arg_{min} ‖ $(x(n)-w_j(n)$ ‖

If BMU ∈ *normal cluster*

then $x(n)$=*normal*

else $x(n)$=*abnormal*

탐지 후에 입력 데이터의 순차적인 학습을 통하여 승자 뉴런의 가중치와 인접한 이웃 뉴런의 가중치가 조정되면서 시간이 흐름에 따라 맵이 갱신되며, 이렇게 점진적 학습이 이루어지면서 실시간 네트워크 특성을 반영한 탐지가 이루어지게 된다. 이는 <표 5>와 같다.

〈표 5〉 실시간 학습 알고리즘

$$w_j(n+1)=w_j(n)+\alpha(n)\beta_{i(x)}(n,j)(x(n)-w_j(n))$$

5. 실험 및 결과

본 장에서는 지도학습의 의사결정트리와 신경망 기법과 비정상 트래픽 탐지를 위해 제안된 비지도 학습의 데이터 마이닝 기법 기반 트래픽 속성 상관관계 메커니즘 성능을 비교평가하고 검증하기 위한 실험 방법, 분석 및 실험 결과에 대하여 설명한다. 본 논문에서 제안한 메커니즘의 성능 평가를 위해 KDD Cup 1999 Data에서 다음 <표 6>과 같이 총 50,000개의 데이터를 추출하여 실험 데이터를 만들었다.

〈표 6〉 실험 데이터 구성

번호	데이터 항목	데이터 개수
1	Normal	5,000
2	Neptune	5,000
3	Smurf	5,000
4	Teardrop	5,000
5	Back	5,000
6	Pod	5,000
7	Ipsweep	5,000
8	Nmap	5,000
9	Portsweep	5,000
10	Satan	5,000
합계		50,000

5.1에서는 성능 평가 방법에 사용된 측정값을 정의하고, 5.2에서는 여러 가지의 실험을 통한 성능 평가를 바탕으로 효과적인 탐지 방법을 알아본다.

5.1 성능 평가 방법

$$\text{데이터 유형별 분류율}=\frac{\text{해당 유형으로 정확히 분류된 데이터의 개수}}{\text{해당 데이터 유형의 데이터 개수}}\times 100$$

패턴 인식 분야에서 일반적으로 사용되는 분류 평가 방법을 일부 재정의하여 성능 평가를 하였다. 분류 성능 테스트는 데이터 유형별 분류 능력을 평가하기 위한 항목으로 데이터 유형별 분류율과 데이터 항목별 분류율로 분류 정확률을 검증한다. 데이터 유형별 분류율은 데이터를 정상과 비정상의 두 가지 유형으로 볼 때, 해당 유형의 데이터 중 해당 클러스터로 정확히 분류된 데이터의 비율을 백분율로 나타낸 값이다. 데이터 항목별 분류율은 정상과 각 공격별 9개 항목으로 나눌 때, 전체 각 항목의 데이터에 대한 해당 항목으로 정확히 분류된 데이터의 비율을 백분율로 나타낸 값이다.

$$\text{데이터 항목별 분류율}=\frac{\text{해당 항목으로 정확히 분류된 데이터의 개수}}{\text{해당 데이터 항목의 데이터 개수}}\times 100$$

5.2 실험 내용 및 결과

5.2.1 실험 I - 분류 정확률

지도학습의 의사결정트리와 신경망 기법과 비정상 트래픽 탐지를 위해 제안된 비지도 학습의 데이터 마이닝 기법기반 트래픽 속성 상관관계 메커니즘의 성능을 비교하기 위해 KDD_TND 와 KDD_TD로 데이터 유형별 분류율과 데이터 항목별 분류율을 구해보았다. 실험 결과 비정상 트래픽 탐지를 위해 제안된 비지도 학습의 데이터 마이닝 기법기반 트래픽 속성 상관관계 메커니즘의 데이터 유형별 분류 정확률이 92.8%로 지도학습보다 높음을 알 수 있었고 데이터 항목별 분류율에서는 다른 공격에 비해 back 공격 분류 정확률이 가장 높고 지도학습에 비해 비지도학습의 데이터 항목별 분류 정확률이 높음을 알 수 있었다. 이는 <표 7>, <표 8>과 같다.

〈표 7〉 데이터 유형별 분류 정확률

데이터 유형별	분류 정확률	
정상/비정상	지도학습	비지도 학습
	71.5%	92.8%

〈표 8〉 데이터 항목별 분류 정확률

데이터 항목	분류 정확률	
	지도 학습	비지도 학습
neptune	64.1	84.5
smurf	49.5	59.6
back	88.9	100
teardrop	50.4	80.2
pod	62.5	87.2
ipsweep	49.9	68.6
portsweep	59.8	85.0
nmap	70.3	73.8
satan	59.2	82.8

5.2.2 실험 II - 모델링 시간 및 탐지 시간

지도학습의 의사결정트리와 신경망 기법과 제안된 비지도 학습의 데이터 마이닝 기법기반 트래픽 속성 상관관계 메커니즘의 실시간 탐지 가능성을 실험해 보기 위해 각 메커니즘을 모델링하는 시간과 탐지 시간을 구해보았다.

모델링 시간이란 지도 학습, 비지도 학습 각 메커니즘이 기존의 학습데이터로 충분한 학습을 시켜 공격 모델을 만드는 데까지 걸린 시간을 의미하며, 탐지 시간이란 각 공격 모델이 완성 된 후 실제로 새로운 데이터가 들어왔을 때 정확하게 탐지하는데 걸리는 시간을 의미한다. 실험 결과 <표 9>

와 같이 비지도 학습이 지도 학습에 비해 모델링시간과 탐지 시간이 빠름을 알 수 있었다.

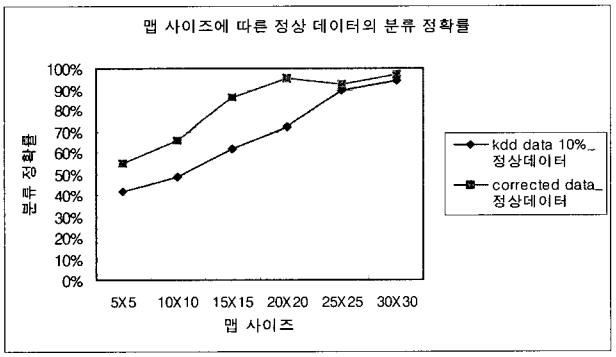
〈표 9〉 모델링 시간 및 탐지 시간

비교항목	분류 정확률	
	지도 학습	비지도 학습
모델링 시간	15.80초	12.40초
탐지 시간	1.90초	0.50초

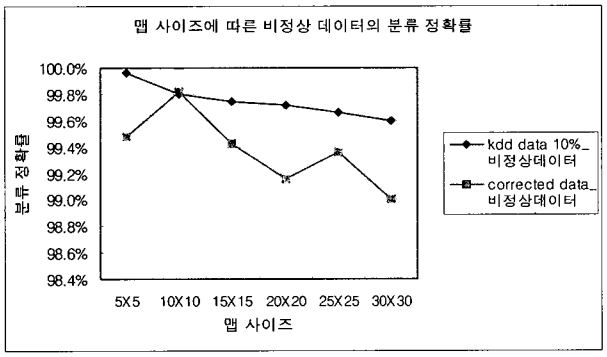
5.2.3 실험Ⅲ- 맵 사이즈에 따른 분류 정확률

SOM을 활용한 제안한 메커니즘은 맵 사이즈에 따라서 학습 시간과 트래픽의 분류 정확률이 달라질 수 있다. 따라서 KDD_TND 와 KDD_TD로 다양한 실험을 했다. 실험 결과 (그림 11)의 정상 트래픽에 대한 분류율을 살펴보면 맵의 사이즈가 작을 때에는 정상 트래픽의 분류 정확률이 현저히 낮은 값을 보이고 맵의 사이즈가 커질수록 큰 폭으로 상승되는 것을 볼 수 있다. 반면, (그림 12)에서 나타난 것처럼 비정상 트래픽에 대한 분류 정확률은 맵 사이즈가 커질수록 큰 차이는 아니지만 조금씩 낮아지는 것을 볼 수 있다.

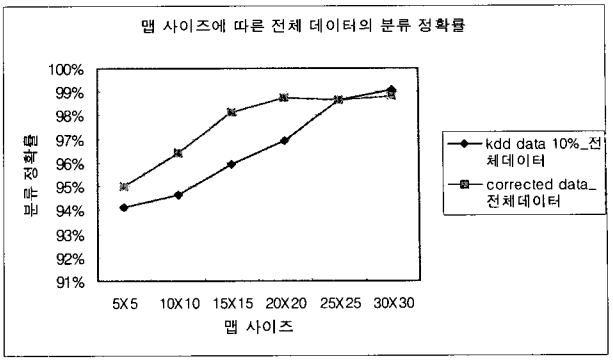
이를 통합하여 볼 때 (그림 13)에서 보여 지듯이 전체 데이터에 대한 분류 정확률은 맵의 사이즈가 커질수록 높아지는 것을 볼 수 있다. 따라서 비정상 데이터와 정상 데이터 모두가 적당한 분류 정확률을 가질 수 있도록 맵의 사이즈를 지정해주는 것이 필요하다. 또한 맵의 사이즈가 너무 커지면 학습 시간이 길어질 우려가 있으니 분류 정확률과 학습 시간을 모두 만족하는 적당한 맵 사이즈의 결정이 중요하다.



(그림 11) 맵 사이즈에 따른 정상 데이터의 분류 정확률



(그림 12) 맵 사이즈에 따른 비정상 데이터의 분류 정확률



(그림 13) 맵 사이즈에 따른 전체 데이터의 분류 정확률

5.2.4 실험Ⅳ - 계층적 맵에 따른 분류 정확률

SOM을 이용하여 학습 데이터를 분류하였을 때, 성격이 유사한 정상 트래픽과 비정상 트래픽이 함께 클러스터링 된 경우를 발견할 수 있었다. 이를 해결하기 위해 KDD_TND를 순서대로 약 65,000개를 추출한 실험 데이터 <표 10>을 사용하여 1차 맵을 형성하여 정상과 비정상 트래픽이 하나의 클러스터로 섞인 데이터는 클러스터별로 다시 한 번 SOM을 이용해 2차적인 클러스터링을 해주는 계층적 맵에 따른 데이터 유형별, 데이터 항목별 분류 정확률을 실험해 보았다. 이는 <표 11>, <표 12>와 같다.

〈표 10〉 실험 데이터

데이터 유형	공격 종류	데이터 수
DoS attack	neptune	11,955
	smurf	11,258
	back	2,002
	teardrop	99
	pod	20
	land	1
Probing attack	ipsweep	658
	portsweep	40
	nmap	130
	satan	2
Normal		39,297
Total		65,462

결과 1계층에서 정상과 비정상의 트래픽이 섞여있는 클러스터를 2계층의 맵까지 형성하였을 때 데이터 유형별 분류 정확률이 더 높은 것을 알 수 있었다.

〈표 11〉 계층 수에 따른 데이터 유형별 분류 정확률

맵 계층 수	분류 정확률
1계층	72.3%
2계층	98.4%

〈표 12〉 계층 수에 따른 데이터 항목별 분류 정확률

데이터 항목	1 계층	2 계층
neptune	67.7	99.8
smurf	99.6	100
back	99.6	99.9
teardrop	21.2	87.8
pod	0	80
land	0	0
ipsweep	48.6	96.3
portsweep	45	95
nmap	53.8	96.9
satan	0	0
normal	65.1	97.5
Total	72.3	98.4

6. 결 론

지도 학습기반 침입탐지의 문제점을 해결하고 점진적 학습 및 실시간 침입탐지를 위해 본 논문에서는 비지도 학습인 자기 조직화 지도(SOM)과 지도 학습인 상관관계를 결합한 SOM기반 트래픽 속성 상관관계 메커니즘을 제안하였다. 비지도 학습만 사용했을 때의 문제점을 해결하기 위해 9종의 공격별 속성간의 상관관계 분석으로 규칙 생성함으로써 지도 학습법을 도입했으며 연결기반의 데이터 속성을 사용하고 SOM 알고리즘 특성상 실시간 탐지 가능하다고 판단했다.

중요 속성을 추출하는 실험 결과, 안정적으로 Basic 속성과 Traffic 속성이 모두 필요하다는 것과 제안된 비지도 학습의 데이터 마이닝 기법 기반 트래픽 속성 상관관계 메커니즘의 성능이 데이터 유형별 분류 정확률 92.8%로 매우 우수하고 공격별로 탐지 정확성에 차이가 있지만 대체적으로 지도 학습보다 데이터 항목별 분류율도 높음을 보여주었다. 또한 제안한 비지도 학습 메커니즘이 지도 학습보다 모델링 시간과 탐지 시간이 더 빠르기에 실시간 탐지가 가능함을 알 수 있었으며 맵 사이즈가 클수록 계층적 맵을 형성할수록 정확한 분류율을 가진다는 것을 알 수 있었다.

향후 과제로는 다른 다양한 데이터로 상관관계 규칙에 기반 하여 맵 분류 시 성능을 측정하는 것과 좀 더 일반적인 상관관계 규칙을 생성해보는 연구가 필요하다.

참 고 문 헌

[1] Leonid Portnoy, "Intrusion detection with unlabeled data using clustering", Undergraduate Thesis, Columbia University, 2000.

[2] Jack Marin, Daniel Ragsdale, John Shurdu, "A Hybrid Approach to the Profile Creation and Intrusion Detection", Proceedings of DARPA Information Survivability Conference and Exposition, IEEE, 2001.

[3] Nong Ye, Xiangyang Li, "A Scalable Clustering Technique for Intrusion Signature Recognition", Proceedings of 2001 IEEE Workshop on Information Assurance and Security, 2001.

[4] 박정민, 나현정, 황경애, 채기준, "광역망에서의 DDoS 탐지 메커니즘에 관한 연구", 이화여자대학교, 2003년도 한국전자통신연구원 위탁과제.

[5] KDD Cup 1999 Data, Available in http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[6] 최종후, 한상태, 강현철, 김은석, 심미경, 이성건, "SAS Enterprise Miner 4.0을 이용한 데이터 마이닝 - 기능과 사용법", 자유아카데미.

[7] SOM Toolbox for Matlab, Available in http://www.cis.hut.fi/projects/somtoolbox/.

[8] 도용태, 김일곤, 김종완, 박정현, "인공지능 : 개념 및 응용", 사이텍미디어.

[9] Pearson Correlation Coefficient, Available in http://www.indstate.edu/nurs/mary/N322/pearsonr.html/

[10] Jelena Mirkovic, Gregory Prier, Peter Reiher, "Atta- cking DDoS at the Source," Proc. of ICNP 2002.

[11] Laura Feinstein, Dan Schnackenberg, Ravindra Balupari, Darrell Kindred, "Statistical Approaches to DDoS Attack Detection and Response," Proc. of The DARPA Information Survivability Conference and Exposition, 2003.

[12] Wenke Lee, Salvatore J. Stolfo, "Data Mining Approaches for Intrusion Detection," Proc. of the 7th USENIX Security Symposium, pp.79-94, Jan., 1998.

[13] Anup K. Ghosh, Aaron Schwartzbard, "A Study in using Neural Networks for Anomaly and Misuse Detection," Proc. of the 8th USENIX Security Symposium, Washington, D.C., USA, Aug., 1999.

[14] Susan C. Lee, David V. Heinbuch, "Training a Neural-Network Based Intrusion Detector to Recognize Novel Attacks," Proc. of the 2000 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY, 6-7 Jun., 2000.

황 경 애

e-mail : dewbelle@ewhain.net

2003년 서울여자대학교 멀티미디어통신공학과(학사)

2005년 이화여자대학교 컴퓨터학과(석사)

2005년~현재 삼성전자

관심분야 : 네트워크 보안, 침입 탐지 시스템, DDos

오 하 영

e-mail : hyoh@ewhain.net

2002년 덕성여자대학교 전산학과(학사)

2001년~2004년 신한금융지주회사 e-신한

2004년~현재 이화여자대학교 컴퓨터학과 석사과정

관심분야 : 네트워크 보안, DDos, 센서 네트워크, 홈 네트워크. 유비쿼터스 컴퓨팅

임 지 영

e-mail : jylim@bible.ac.kr

1994년 이화여자대학교 전자계산학과(학사)

1996년 이화여자대학교 전자계산학과(석사)

2001년 이화여자대학교 과학기술대학원(박사)

2001년~2003년 이화여자대학교 컴퓨터학과 대우 전임강사

2003년~현재 한국성서대학교 정보과학부 전임강사

관심분야 : 네트워크 보안, 애드 혹 네트워크, 센서 네트워크, 유비쿼터스 컴퓨팅, 네트워크 프로토콜 설계 및 성능분석

채 기 준

e-mail : kjchae@ewha.ac.kr

1982년 연세대학교 수학과(학사)

1984년 미국 Syracuse University 컴퓨터학과(석사)

1990년 미국 North Carolina State University 컴퓨터공학과(박사)

1990년~1992년 미국 해군사관학교 컴퓨터학과 조교수

1992년~현재 이화여자대학교 컴퓨터학과 교수

관심분야 : 네트워크 보안, 인터넷/무선통신망/고속통신망 프로토콜 설계 및 성능분석, 센서네트워크, 홈 네트워크, 유니쿼터스 컴퓨팅

나 중 찬

e-mail : njc@etri.re.kr

1986년 충남대학교 계산통계학과(이학사)

1989년 숭실대학교 전자계산학과(공학석사)

2004년 충남대학교 컴퓨터과학과(이학박사)

1989년~현재 ETRI 능동보안기술연구팀 팀장

관심분야 : 실시간시스템, 네트워크 관리, 네트워크 보안