Misuse IDS의 성능 향상을 위한 패킷 단위 기계학습 알고리즘의 결합 모형

원 일 용[†] · 송 두 헌[††] · 이 창 훈[†††]

요 약

전문가의 침입 분석 지식을 기반으로 한 Misuse IDS는 침입 탐지 비율은 우수하지만 과도한 오경보를 생성하여 관리 효율성이 낮다. 우리는 패킷 정보 중심의 사례 기반 학습을 Misuse IDS와 결합하여 그 행동 특성에 따라 오경보를 줄이는 모형을 제안하고 실험하였다. 또 기존의 IBL(Instance Based Learner)을 개선한 XIBL(Extended Instance Based Learner)을 이용하여 Snort의 alarm을 패킷 수준에서 역 추적 분석하여, 그 alarm이 실제로 보내질 가치가 있는지를 검사한다. 실험 결과 진성경보와 오경보 사이에는 XIBL의 행동상 분명한 차이가 드러나며, 네트워크 상의 공격이 비록 여러 패킷의 결합된 형태로 나타나지만, 개별 패킷에 대한 정상/비정상 의사 결정도 Misuse IDS와 결합하면 전체 시스템의 성능을 향상하는 데에 기여할 수 있음을 실증적으로 보여주었다.

A Hybrid Model of Network Intrusion Detection System : Applying Packet based Machine Learning Algorithm to Misuse IDS for Better Performance

Ill-Young Weon[†] · Doo Heon Song[††] · Chang-Hoon Lee[†††]

ABSTRACT

Misuse IDS is known to have an acceptable accuracy but suffers from high rates of false alarms. We show a behavior based alarm reduction with a memory-based machine learning technique. Our extended form of IBL (XIBL) examines SNORT alarm signals if that signal is worthy sending signals to security manager. An experiment shows that there exists an apparent difference between true alarms and false alarms with respect to XIBL behavior. This gives clear evidence that although an attack in the network consists of a sequence of packets, decisions over individual packet can be used in conjunction with misuse IDS for better performance.

키워드 : 침입탐지시스템(Intrusion Detection System), 사례기반학습(Instance Based Learning), 오경보(False Alarm), 학습(Machine Learning), 패킷 헤더 정보(Packet Header Information)

1. 서 론

현재 대부분 사용되고 있는 네트워크 기반 침입탐지 시스템은 그 탐지 기법이 네트워크와 해킹 전문가에 의해 만들어진 시그너처를 기반으로 하는 Misuse System이다. 이러한 Misuse 방식은 등록되어 있지 않은 새로운 유형의 공격이나, 이미 등록 되어 있지만 약간 변형된 공격 탐지에는 취약점이 있다. 준비된 시그너처에 잡히지 않는 공격이 전문가에 의하여 탐지되면, 로그 분석 및 각종 통계량에 의거 탐지된 공격의 특징을 추출하여 규칙을 만들게 되는데 이 과정은 많은 시간적 인적 자원이 소요된다[1].

전문가의 손길을 줄이면서 침입탐지 능력을 증가 시키기 위해 각종 인공지능 기법들이 침입탐지 영역에 도입 되고 있는데, 그 중 주목을 받고 있는 것은 Misuse 방식과는 달리 네트워크의 상태를 모델링 하고, 인공지능의 기계학습 기법을 이용하여 네트워크 상태에 대한 지식을 생성, 실제 환경에서 현재 네트워크의 정상여부를 판단하는 기계학습 또는 데이터 마이닝 기반의 침입 탐지 기법들이다[2-5]. 여기서 네트워크의 정상(normal)이라 함은 패킷의 흐름에 공격이 포함되지 않은 것을 의미하며, 비정상(abnormal)이라 함은 네트워크 흐름에 공격이 포함되어있거나 있을 것으로 의심되는 것을 의미한다. 이 방법의 장점은 원시 자료의 생성부터 학습, 판단까지 사람의 손이 거의 필요로 하지 않는다는 점이다. 이 방법의 문제점은 학습의 기반이 되는 attribute를 무엇으로 하느냐에 그 성능이 달려 있는데, 시그너처 방법에서 사용했던 통계량을 중심으로 attribute를 구성하면 새로운 공격에 대한 반응이 떨어지고, 패킷 헤더 정

† 준 회 원 : 건국대학교 대학원 컴퓨터공학과
†† 정 회 원 : 용인송담대학 컴퓨터 S/W과 교수
††† 종신회원 : 건국대학교 컴퓨터공학과 교수
논문접수 : 2003년 12월 11일, 심사완료 : 2004년 4월 13일

보에 의존할 경우 대부분의 공격이 여러 패킷의 sequence에 의한 결과로 나타나기 때문에 동일한 패킷이 선행 패킷의 상태에 따라 정상으로 또는 비정상으로 분류될 수 있는 내용 민감도(context sensitivity) 문제가 있어 Misuse 방식에 비해 비정상 인식 정도가 높지 않으며, 오히려 정상 상태의 인식 비율이 높은 편이다.

최근의 네트워크 침입탐지 시스템의 연구 경향 중 한 방향은 Misuse IDS와 학습 기반의 IDS의 서로 다른 특성 -Misuse IDS의 높은 공격 인식 비율과 학습 기반 IDS의 '정상' 모형화 기능-을 상호 보완적으로 적용하여 침입탐지 능력을 향상 시키는 것이다. 이 방향의 연구에서 특히 지적되고 있는 것은 대부분의 상용 IDS가 채택하고 있는 Misuse 방식이 공격 탐지 비율은 높지만 탐지 규칙의 증가에 비례하여 false positive비율이 기하 급수적으로 증가하여 관리자를 괴롭히고 있다는 점이다[4, 6]. 최근 연구에 따르면 하루에 발생한 alarm 중에 중요한 것은 5개에서 6개 정도 이고 나머지 수천 개는 이것 때문에 부수적으로 발생한 alert인 경우도 있음이 알려져 있다[7]. 더구나 대상 IDS의 알려진 공격 탐지 시그너처를 겨냥하여 고의적으로 침입을 계속적으로 탐지케 하는 패킷을 생성하는 것도 가능함이 밝혀져 오경보 제거 문제는 Misuse IDS 성능 향상의 중심 문제가 되어 있다[1] .

따라서 시그너처 기반 IDS의 공격 탐지 비율을 손상하지 않으면서 오경보를 줄여 관리의 효율성을 높이는 문제가 중요하게 다루어지는데, 이 문제를 해결하기 위해 두 방향의 연구가 진행되고 있다.

첫째 방법은 데이터 마이닝 기법을 이용하여 지식 기반의 선행 필터로 하여금 중요한 alarm만 취급하게 하는 것으로 실제 시스템에서 30% 정도의 오경보 감소가 보고되고 있다[7, 8]. 그러나 이 방법은 자주 발생하지 않는 공격을 무시하는 등 그 구조상 정보 손실을 피할 수 없어 비정상 인식도의 손상 및 새로운 공격에 대한 대처 능력이 미흡하다.

다른 방법은 Misuse IDS가 alarm을 발생시켰을 때 이것을 학습 또는 클러스터링 방법을 통해 관리자에게 통보할 가치가 있는가를 재차 판단하는 방법[9]으로 흔히 행동 기반 분석법이라고 불린다. 이 경우는 앞에서 언급했던 선행 필터 사용법 보다 시스템 부하는 높지만 정보 손실이 없다. 따라서 학습 또는 데이터 마이닝 기법이 적절하게 결합된다면 큰 효과를 거둘 수 있다.

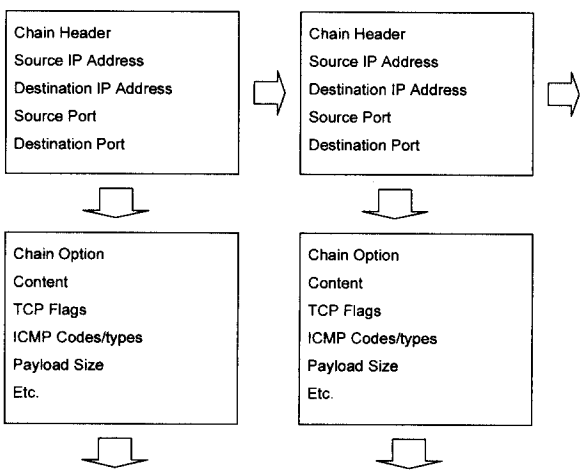
본 논문은 Misuse IDS와 학습 기반 IDS를 행동 기반 분석 기법으로 결합하는 모형에 관한 연구로 Misuse System으로는 Snort (1.8.1)[10]을 사용하였고, 학습 기법에서는 IBL (Instance Based Learning)[11]을 확장한 XIBL을 사용하였다. 본 논문에서 제시하는 결합 모델은 Snort에서 alarm이 발생하면 이 시점을 기준으로 일정 크기만큼 패킷을 역추적한다. 그리고 이 패킷들에 대한 IBL의 Alert 패턴을 검사하여 Snort의 Alert에 대한 신뢰도를 정밀하게 측정하는 방식이다. 이러한 모델은 Misuse IDS를 주 시스템으로 학습 기반 IDS를 보조적으로 사용하고자 하는 실제 환경의 요구사항과도 부합되며, 성능 또한 Misuse IDS를 단독으로 사용하는 것 보다 향상 될 수 있음을 DARPA 1998 data set [12] 분석을 통하여 실증하였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에 사용될 Misuse System인 Snort의 동작 원리에 대하여 설명하였고, 기계 학습의 대표적인 학습 알고리즘인 C4.5 및 IBL의 확장에 대한 내용을 기술하였으며, 제안된 모형에 적합한 학습 알고리즘에 대하여 설명하였다. 3장에서는 결합 모형의 의미 및 구조에 대하여 기술하고, 실험에 사용될 자료인 DARPA 1998 data set을 설명하고, 네트워크 모델링을 통한 학습 자료 생성에 대하여도 언급하였다. 4장에서는 제안된 모델의 유용성을 보이기 위해 오경보 및 진성경보 실험 및 그 결과를 분석하였다. 5장에서는 결론 및 향후 과제에 대하여 논의하였다.

2. 목표 시스템

2.1 Snort

Snort는 Packet Decoder, Detection Engine, Logging and Alerting 의 3가지 기본 모듈로 구성되어 있다. Snort는 최하부에 패킷 수집을 위해 libcap[13]을 사용한다. Packet Decoder에서는 발생한 패킷을 data link layer에서 application layer 까지 순차적으로 processing한다. Snort는 체인헤더와 체인옵션이라고 불리는 2차원 linked-list 형태로 탐지 규칙을 사용한다. 이 규칙들은 체인 헤더에 있는 공통적인 속성들과 체인 옵션에 포함되어 있는 탐지 옵션들로 표현된다.



(그림 1) Snort의 룰 체인

예를 들어 만약 여러 개의 CGI-BIN probe 탐지 규칙들

이 Snort의 탐지 엔진에 포함되어 있다면, 이 규칙들 모두는 일반적으로 공통의 소스와 목적지 IP주소, 포트를 공유한다. 탐지 속도를 증가시키기 위해, 이 모든 공통적인 요소들이 한 개의 체인 헤더에 표현되어 있고, 개별적 탐지 시그너처들은 체인의 옵션 구조에 포함되어 있다. 패킷이 발생 하면 규칙 체인들은 양쪽방향에서 각각의 패킷에 대하여 재귀적으로 검색된다. 탐지 엔진은 실행시기에 규칙 Parser에 의해 활성화된 체인 옵션들만을 검사한다. 탐지엔진에서 decoding된 패킷과 일치하는 첫 규칙은 규칙 선언에서 정의된 동작을 발생시킨다.

아래는 Snort에서 사용하는 룰의 한 예이다.

```
alert tcp $EXTERNAL_NET any → $HOME_NET 21
(msg : "FTP large PWD command" ; flow : to_server,established ;
content : "PWD" ; nocase ; dsize : 10 ; classtype : protocol-command-
decode ; sid : 1624 ; rev : 3 ; )
```

(그림 2) Snort의 룰

본 실험을 위해 사용된 snort의 버전은 1.8.1인데 최신 버전은 본 논문이 쓰여지는 시점을 기준으로 2.x에 이르고 있다. Darpa에서 사용할 수 있는 실험용 data는 1998, 1999, 2000년이 있는데, 본 실험에 적합한 것은 성격상 1998, 1999년 data라고 할 수 있다. Snort 2.x은 최근에 나온 것이므로 darpa에서 제공하는 1998, 1999, 2000년의 모든 공격들을 다 인식할 만큼 룰이 확장되어 있다. 따라서 1998, 1999년 data를 가지고 제안된 아이디어를 실험하기 위해서는 실험용 data와 비슷한 시기거나 더 앞선 rule로 운영되고 있는 snort버전이 필요하므로 구 버전(1.8.1)을 사용하였다. 또 구 버전의 룰 환경 설정은 snort.org에서 배포하는 기본 설정을 그대로 사용하였다. 제안된 논문의 핵심은 결합모델이며, misuse system의 한 예로 snort를 사용한 것이므로, snort가 어떤 공격은 인식하고 어떤 공격은 인지하지 못하게 룰 설정을 하였는지는 중요한 쟁점이 아니고, 각종 공격을 적당하게 인식하거나 인식하지 못하도록만 설정된 환경이라면, 그것이 snort 아닌 다른 어떠한 misuse IDS라도 실험에는 적합하다고 할 수 있다. 즉 misuse system과 학습기반 anomaly system의 결합이 초점이다.

2.2 C4.5

C4.5는 attribute를 기반으로 많은 사례를 통해서 이를 효율적으로 판단할 수 있는 attribute의 결정트리(Decision Tree)를 만든다. 결정 트리를 생성하기 위해 data set의 모든 속성 필드의 속성값에 따른 엔트로피를 구하고, 이때 계산된 엔트로피 중 가장 낮은 값을 가지는 항목을 root node로 두고, 이것을 기반으로 다른 항목들의 사례를 나누는 Divide and Conquer 기법을 사용한다. 이러한 과정을 반복하여 전체 트리를 구성하여 하나의 leaf가 하나의 규칙 또는 패턴을 형성하는 것이다.

C4.5는 다량의 데이터를 아주 짧은 시간 안에 비교적 정확한 규칙으로 생성하는 곳에는 성공하고 있다. 그러나 학습 중에 주어진 자료가 달라지면 매번 다른 트리를 형성하고 때로는 주어진 자료에 따라 매우 다른 형태의 트리가 형성되기도 한다. IDS 환경은 기본적으로 실시간 적이라는 점에서는 부합하지만 자료의 Dynamicity라는 점에서 C4.5가 이 분야에서는 매우 제한적으로 사용되어 왔다. 즉, 변화의 여지가 적은 규칙 군집의 검색 효율화[4]에서는 성능이 우수하였지만 패킷 정보를 직접 다루는 경우에는 전술한 문제 등으로 인해 그 효과가 실망스러웠다[2]. 따라서 C4.5를 IDS의 학습 알고리즘으로 사용하기에는 적당하지 않다고 판단된다.

2.3 XIBL

IBL은 training 자료에서 지식을 추출하는 방법이 C4.5처럼 attribute 중심의 개념적인 것이 아니고, 자료 자체 즉 인스턴스(Instance) 자체로 대변된다[11]. 이렇게 학습된 지식을 PCD(Partial Concept Description)라고 부르며, 이러한 PCD는 각각 특정 클래스를 대표하는 대표자의 성격을 갖는다. IBL의 클래스 결정 알고리즘은 단순히 새로 들어온 데이터와 가장 유사한 k개의 PCD 내의 인스턴스를 찾아 그들의 클래스 분포에 따라 입력된 인스턴스의 클래스를 결정한다. 이 알고리즘은 많은 문제에서 C4.5와 필적할 만한 정확도를 보이며 학습 속도는 C4.5 보다는 느리지만 학습된 결과에 따른 새 데이터의 검증 속도 및 자료 변화에 따른 안정성 등에서는 우수하다.

IBL 알고리즘의 두 가지 축은 PCD에 저장하는 인스턴스 결정법과 인스턴스 간 거리 계산법으로 말할 수 있다. Aha는 그의 박사학위 청구 논문에서 PCD 내 인스턴스 결정법에 대하여 네 가지 버전을 제시하였는데, 모든 인스턴스 저장법(IB1)은 다량 데이터에 대해서는 기억 장소 초과 문제를 갖는다. 따라서 유사한 인스턴스 그룹 중에서 대표성을 갖는 인스턴스만 저장하는 방법(IB2)이 제시되었으나 noise에 매우 취약하여 여기에 noise 제거 기능(IB3)과 불필요한 attribute의 간섭 문제를 약화시키는 attribute 가중치 알고리즘(IB4)이 개발되었다. 그러나 많은 실제 문제에서 IB3가 IB4보다 오히려 우수하여 IB3를 많이 쓰고 있다.

인스턴스 간 거리 계산은 실수형 자료는 Euclidean 또는 Manhattan Distance를, 이산형 자료사이의 거리는 attribute 값이 같으면 0, 다르면 1로 측정한다.

우리의 XIBL은 본래의 IBL에 아래의 몇 가지 기능을 부여하였다.

첫째, 인스턴스사이의 거리 계산에서 이산형 자료의 경우 IBL은 그 결과가 0 또는 1로 표시되기 때문에, 실수형 attribute가 정규화에 따라 0과 1 사이의 값을 갖는 것에 비해

실질적 가중치가 높다. 이러한 것을 반영하기 위해 XIBL에서는 이산형 데이터의 거리를 0과 1 사이의 연속 값으로 나타낼 수 있는 Value Difference Metric(VDM)[15]을 적용하였다.

둘째, IBL은 noise에 민감한데 이 문제를 해결하기 위해 수학적 통계에 근거를 두고 있는 Leave-one out[16] noise 필터를 적용하였다.

셋째, attribute 가중치 값을 지정하는 방법에서 IB4가 채택한 학습 중의 reward/penalty 부여법 대신 통계적으로 검증된 backward stepwise regression에 의한 attribute 가치 판단 기법을 적용하였다.

이렇게 생성된 XIBL의 성능은 C4.5보다 우수하고[17] 원본 IBL보다도 뛰어나며[3], PCD 내 인스턴스의 양은 학습 인스턴스의 8~10% 수준이며 다량의 데이터에 특히 안정성을 갖고 반응하였다[17]. 학습 알고리즘으로서의 XIBL에 대한 상세한 논의 및 특징 분석은 이 논문의 주제를 벗어나므로 여기에 언급하지 않았지만, 우리는 이미 그 내용을 논문[3]에서 밝혔다.

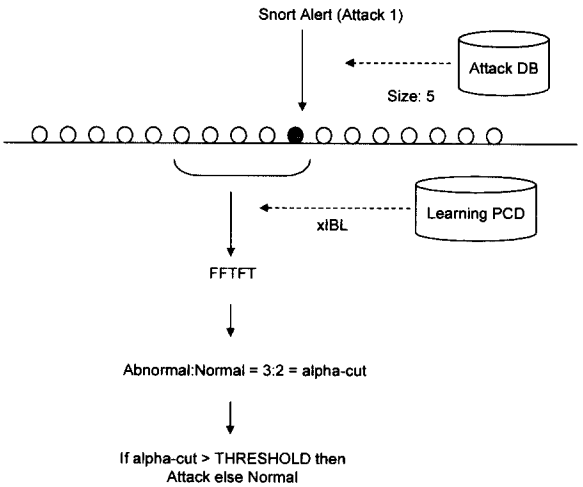
3. 결합 모형

3.1 구 조

Snort가 침입이라고 alarm을 했을 때, 이것은 2가지 종류로 구분해 볼 수 있는데, 첫째, 공격을 공격 이라고 한 경우와 둘째, 정상을 공격이라고 한 경우이다. Snort 자체로는 이것이 첫째 유형인지, 둘째 유형인지 분류하는 것이 불가능하다. 본 논문은 XIBL이 패킷 단위로 정상 및 비정상을 판단할 수 있다는 특성을 이용하여 snort에서의 alarm이 발생했을 때 그것이 어떤 유형인지를 XIBL이 판단 할 수 있는 모형을 제시하고자 한다.

네트워크 상에서 각각의 공격들은 패킷의 연속적인 집합으로 표시되는데, 이러한 공격들은 일정한 평균 개수의 패킷수를 보인다. 공격 타입이 알려지지 않은 경우는 알려진 공격 타입들에서 추출한 길이의 평균으로 적용한다. Snort에서 alarm이 발생하면 Snort의 로그를 통해서 공격 유형을 알 수 있고 이 유형에 따라 이 공격과 관련된 패킷의 평균 길이를 측정할 수 있다. Alarm이 발생한 시점을 기준으로 해당 공격의 패킷 길이만큼을 역추적하여 XIBL을 이용하여 각각의 패킷에 대한 특성을 분류한 후, 측정값에서 비정상 및 정상에 대한 비율을 측정한다. 이렇게 측정된 값이 특정 기준 이상을 넘어 서면 첫째 유형의 alarm이라고 판단하고, 그렇지 않으면 둘째 유형이라고 판단한다. 물론 이 방법이 유효하려면 같은 Snort Alarm이 진성경보인가 오경보인가에 따른 XIBL의 행동 특성이 명확히 구분되어야 하는데, 이 내용은 4장에서 실험으로 증명된다.

이 논의의 논거는 다음과 같은 논리에서 출발한다. 보통 공격은 그것이 충분한 수의 패킷동안 지속된다면, 같은 목적을 가진 하나 이상의 사용자의 행동 패턴이므로 패킷 단위의 정보 유형이 서로 유사할 것이다. 따라서 특히 이전에 탐지된 적이 있는 공격의 경우 XIBL의 학습 과정에서 그 공격의 특징적인 인스턴스가 PCD 내에 존재할 확률이 높다. 또 이전에 탐지된 적이 없는 공격의 경우도 상당 부분 '공격'의 특성을 공유하고 있다면 새롭게 탐지될 가능성이 높아진다. 오경보의 경우는 대부분의 시그너처가 통계량에 의존하므로 놓칠 수 밖에 없는 패킷 단위 정보 특성을 통하여 정상적인 패킷의 stream으로 판단될 가능성이 높다. (그림 3)은 이 구조를 보여주고 있다.



(그림 3) 결합 방법에 대한 구조

3.2 DARPA 1998년 Data Set

DARPA 1998년 Data에는 네트워크 침입 탐지를 위해 가상 환경을 구축하고 각종 시나리오에 따라 발생하는 망의 모든 패킷들을 packet dump한 파일이 제공된다. 제공되는 packet dump data는 Training Data set과 Test Data Set

〈표 1〉 공격 Type에 대한 설명

공 격 명	세 부 설 명
Back	Denial of service attack against apache webserver where a client requests a URL containing many backslashes.
ftp-write	Remote FTP user creates rhost file in world writable anonymous FTP directory and obtains local login.
Guest	Try to guess password via telnet for guest account.
Imap	Remote buffer overflow using imap port leads to root shell
Land	Denial of service where a remote host is sent a UDP packet with the same source and destination
Loadmodule	Non-stealthy loadmodule attack which resets IFS for a normal user and creates a root shell
...	...

으로 나누어 진다. Training data set은 전체 7주 분량이 있는데, 각 주는 월요일에서 금요일까지 5개의 data가 제공된다. Test data set은 2주 분이 제공되며 각 주의 구성은 Training data set과 동일하다. 제공되는 packet dump 자체는 단순하게 패킷들을 저장한 것이므로, 이것만으로 공격 및 정상 패킷을 구분하는 것이 불가능하다. 그러므로 각각의 packet dump에 대한 공격 및 정상 여부를 표시한 별도의 정보 파일이 제공된다.

1998년 data set에서 제공되는 공격의 종류는 대략 20종류 정도 되는데, 포함되어 있는 공격의 종류는 <표 1>과 같다.

3.3 학습자료를 위한 네트워크 상태 모델링

기계학습을 이용한 침입탐지에서 학습할 알고리즘을 선택하기 전에 먼저 선행되어야 하는 작업은 알고리즘이 학습할 학습 자료를 만드는 일이다. 네트워크의 상태를 모델링 하는 방법에는 여러 가지가 있으며, 크게 여러 개의 패킷들을 모아 특정 조건이 되면 원시 자료를 한 개 생성 하는 방법과 각 패킷마다 한 개의 학습 자료로 만드는 방법이 있다. 여기에 대해서는 많은 연구가 진행되고 있지만 아직 가장 효율적인 방법은 알려지지 않고 있다. 본 논문에서 네트워크의 상태를 모델링 하기 위해 사용한 방법은 패킷 하나에 학습 자료 하나를 생성하는 방법을 사용 하였다. 이러한 방법은 본 논문에서 제안 하는 패킷 단위의 행동모형을 적용하기에 가장 적합하기 때문이다.

〈표 2〉 자료 attribute 설명

Name	Type
Length	Real
TimeToLive	Real
ServiceType	{default, mincost, maxrelia, maxthrou, maxdelay, unknown}
Protocol	{icmp, igmp, tcp, egp, udp, ipv6, ospf, elsepro}
SrcAddType	{homenet, plumnet, grapenet, worldnet, exnet}
DstAddType	{homenet, plumnet, grapenet, worldnet, exnet}
Doff	Real
WindowSize	Real
SrcPortType	{echo, ftpdata, ftp, telnet, smtp, tftp, gopher, http, rlogin, pop3, ·wins,etcport}
DstPortType	{echo, ftpdata, ftp, telnet, smtp, tftp, gopher, http, rlogin, pop3, ·wins,etcport}
Fin	{on, off}
Syn	{on, off}
Rst	{on, off}
Psh	{on, off}
Ack	{on, off}
Urg	{on, off}
BoundType	{inbound, outbound}
Class	{abnormal, normal}

한 개의 자료를 구성하는 attribute 집합은 IP, TCP 패킷의 헤더범위 내에서 그 값들이 추출 되었다. 한 개의 자료를 구성하는 각각의 attribute 및 가능한 값들은 <표 2>에 설명하였다. 특히 Address Type은 실험용 가상네트워크(DARPA)를 고려해서 구분 하였으며, Port Type은 일반적으로 자주 사용되는 echo, ftp data, ftp, telnet, smtp, tftp, gopher, http, rtelnet, pop3, snmp, ipx, https, telnets, pop3sx, wins와 나머지를 etcport로 구분하였다. 패킷의 발생 근원지에 따라 inbound, outbound로 구분하였는데, inbound라 함은 패킷의 발생지가 내부 네트워크 즉 침입탐지 시스템이 설치되어 운용되고 있는 home network를 의미한다.

4. 실 험

용어의 일관된 의미를 위하여 정상을 정상이라 판단하는 것을 true positive, 정상을 비정상이라 판단하면 false positive, 비정상을 비정상이라 하면 true negative, 비정상을 정상이라고 하면 false negative라고 정의하겠다.

4.1 XIBL Training

XIBL의 학습 지식을 위해 1998년 Training data set에서 우선 각 주의 요일마다, 해당 공격들에 해당하는 패킷들을 해당 세션(Session)별로 불리하고, 동일한 패킷 개수 비율로 정상 패킷들을 분리해 낸다. 이렇게 하여 정상과 비정상 비율이 동일한 실험용 전체 data를 구성하고, 이 자료를 30 : 70 비율로 나눈다. 이때, 각각의 분리 자료 내부의 정상 및 비정상의 비율은 50 : 50로 동일하다. 전체 자료 중 30% 자료는 XIBL의 특성 및 성능을 측정하기 위한 용도이고, 나머지 70%는 XIBL의 Training을 위해 사용하였다. Training data는 10개의 폴더로 나누어 10CV(Cross Validation)를 시행하고, 이중 가장 성능이 우수한 경우의 PCD를 학습 지식으로 지정하였다. 이렇게 선택된 PCD와 전체자료를 이용하여 앞장에서 제안된 모델의 가능성 실험을 실시하였다.

이 실험의 목적은 Snort alarm에 대한 XIBL의 행동 분석으로 각 공격 타입별로 진성경보와 오경보 사이 서로 구별될 수 있는 패턴이 존재하는가 인데, 만약 존재한다면 위에서 제안한 결합 모형이 그 논리적 타당성을 가지고 있음을 알 수 있다.

실험에서 사용된 각종 인수는 다음과 같다. XIBL의 패킷별 정상, 비정상 판단을 위한 PCD 내 참조 인스턴스 수는 voting 등의 다른 변수를 제거하기 위하여 1로 두었고, 결합 모형의 역추적 패킷수는 사전 공격 유형별 평균 패킷수에 근거하여 100으로 두었다.

실험 학습에 사용된 총 비정상 패킷은 44726개였고 정상은 45453개였으며 최적 PCD는 약 8%(정상 365 비정상

294)를 포함하며, 집합의 적응도(Fitting Rate)는 99.4%였다.

XIBL만의 정상 및 비정상 탐지 능력을 실험하기 위해, 전체 실험 자료 중 70%로 training을 실시하고 나머지 30%로 성능을 test하였으며, 공격로그에서 추출된 자료는 모두 공격으로 라벨링하였고 - 즉 어떤 공격이 10개의 패킷으로 이루어져 있다면 이 10개의 패킷을 모두 공격이라고 라벨링하였다 - 정상로그에서 추출된 자료는 모두 정상으로 라벨링하였다. 이렇게 구분된 실험 자료를 XIBL에서 Test 하였을 때, 공격이라고 라벨링 된 것을 공격이라고 판단한 것은 전체 공격 자료의 68.93%, 정상을 정상이라고 탐지 한 것은 전체 정상자료의 88.05%였다. 정상을 공격이라고 오인한 경우는 전제 정상의 11.95%, 공격을 정상이라고 오인한 경우는 전체 공격의 31.07%였다.

4.2 진성경보 분석

본 절에서는 Snort에 각종 공격 data를 보냈을 때, snort에서 공격을 정확하게 잡아내는 경우(true alarm)에 대하여 실험하였다. <표 3>은 총 82개의 공격 중 Snort에서 바르게 Alert을 발생시킨 35개의 공격유형에 대한 XIBL의 Alert 결과의 일부를 보여 준다.

〈표 3〉 진성경보에 대한 XIBL의 결과

공격 분류	공 격 명	Packet 수	XIBL 반응		비정상 비율(%)
			정 상	비정상	
Back	2_fri_back	219	195	24	10.9589
	3_wed_back	222	5	217	97.74775
	6_wed_back	4350	3	4347	99.93103
	7_fri_back	4195	5	4190	99.88081
Phf	3_mon_phf	9	8	1	11.11111
	6_mon_phf	7	2	5	71.42857
	7_mon_phf	7	2	5	71.42857
	7_wed_phf	7	6	1	14.28571
Portsweep	2_mon_portsweep	216	0	216	100
	3_mon_portsweep	336	0	336	100
	4_thurs_portsweep	1129	0	1129	100
	5_fri_portsweep	199	0	199	100
	...	...	...	...	...
Satan	3_mon_satan	84	5	80	94.11765
	4_tues_satan	226	215	11	4.867257
	6_mon_satan	12	6	6	50

예를 들어 Back 공격유형 중 2주차 금요일에(2_fri_back) 행한 공격에서는 관측된 패킷의 개수는 219개이며, Snort는 이 공격을 감지하였고 XIBL은 219개 중 195개는 정상으로 나머지 24개는 비정상으로 감지하였다. 따라서 XIL는 전체 패킷 219개 중 비정상으로 감지한 비율은 약 10% 정도이다. 지면의 제약상 전체 실험결과 중 일부만 표시하였다. 이 분석의 목적은 Snort가 인식하는 비정상에 대하여 XIBL의 행동 양식을 관찰하기 위한 것이다. 특히 각각의 공격에서 관측되는 패킷의 크기가 일정 하지 않은 이유는 대부분의 공격 유형이 공격의 끝과 시작이 명확하게 정해져 있지 않고 공격자가 임의의 시간동안에 공격을 행했기 때문이다.

4.3 오경보 분석

본 절에서는 실험 자료 중 정상 자료를 Snort에게 보냈을 때, Snort가 공격이라고 오인한 alarm에(False positive) 대한 XIBL Alert 결과를 실험하였다. Snort에서 Alert이 발생하면 해당 공격 타입을 조사하여 해당 길이에 해당하는 만큼 XIBL의 Alert를 측정하였다.

〈표 4〉 오경보에 대한 XIBL의 결과

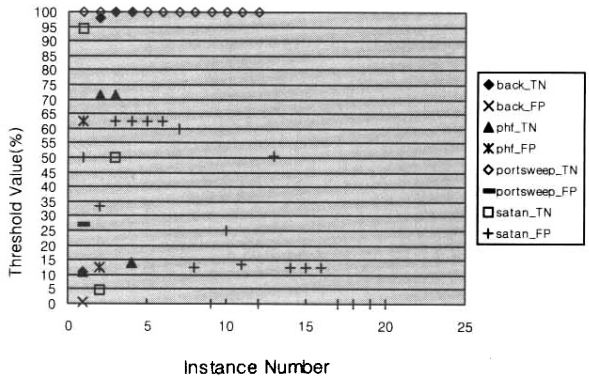
공격분류	정상파일명	XIBL 반응		비정상 비율(%)
		정상	비정상	
Back	7_fri	4881	33	0.671551
Phf	3_wed	3	5	62.5
	3_fri	7	1	12.5
Portsweep	3_fri	16	6	27.27273
Satan	1_mon	4	4	50
	1_thur	10	5	33.33333
	2_fri	3	5	62.5
	3_mon	3	5	62.5
	3_wed	3	5	62.5
	...	...	...	

예를 들어 7주차 금요일에서(7_fri) 추출한 정상 data를 Snort에 보냈을 때, Snort는 Back 공격이라고 잘못된 alarm을 발생 시켰다. 동일한 data를 XIL로 보내고 Snort가 공격이라고 alarm 한 시점을 기준으로, 공격 data에서 관측된 back 유형의 평균적 패킷 4881개에 대한 XIBL의 반응을 조사하여, 비정상 비율이 0.67%임을 관측한 것이다. 각 공격유형의 패킷 평균 길이는 snort가 인식한 경우와 인식하지 못한 경우를 모두 포함하여 계산하였다. 이 분석의 목적은 Snort에서 잘못된 alarm이 발생했을 때, 이것에 대한 XIBL의 행동 양식을 관찰하는 것이다.

4.4 통합 분석

앞 절의 실험 결과들을 통하여 분석해 보면, 적당한 길이의 패킷으로 이루어진 공격의 경우 진성 경보에 대한 XIBL의 반응은 보통 50%를 넘는 반면 오경보의 경우 대부분 50% 미만이 XIBL의 개별 패킷 검사 결과 alarm으로 나타난다. 이는 약 50%를 기준으로 공격을 이루는 패킷 흐름을 XIBL로 분석하면 진성경보와 오경보의 차이를 구분해 낼 수 있다는 것으로, 다소의 예외(예 : 2_fri_back, 3_mon_phf)는 있으나 대부분의 경우 이것을 기준으로 나눌

수 있다. 전체 실험 data 중 진성경보와 오경보가 동시에 발생하는 유형은 1998년 DARPA data set에서는 4가지로 관찰 되었으며 각각의 유형에 대한 반응을 아래 (그림 4)에 표시 하였다. 즉 (그림 4)는 <표 3> <표 4>를 동시에 고려하기 위해 한 공간에 <표 3>, <표 4>의 실험 결과를 같이 표시한 것이다.



(그림 4) 임계값 측정

세로축은 XIBL을 사용한 비정상 정도를 백분율로 표시하였으며, 가로축은 각각의 공격유형별 공격 번호를 표시하였다. 또 공격유형 명칭 뒤쪽에 첨가된 postfix TN은 true negative(Snort가 공격을 공격이라고 한 경우)를 의미하며, FP는 false positive(Snort가 정상을 공격이라고 오인한 경우)를 의미한다. Back 유형은 5회, phf 6회, portsweep 13회, satan 23회의 snort alarm에 대한 XIBL의 비정상 정도를 측정하였다.

(그림 4)에서 보면 Snort가 공격이라고 alarm 했을 때, XIBL의 alert결과를 적용하면, Snort의 alarm이 false positive인지 true negative인지 구분 할 수 있는 패턴들이 존재함을 알 수 있다. 예를 들어 Snort가 satan을 alarm 했을 때, 그 시점을 기준으로 XIBL로 비정상 정도를 측정한다고 했을 때, 관리자는 XIBL의 비정상도가 65% 이상 일 때 만 공격이라고 결정하였다고 하자. 그렇다면 위의 실험 결과에 의하면 satan과 관련하여 Snort가 내는 모든 진성경보는 alarm되고 snort가 내는 모든 오경보는 alarm되지 않아 전체적으로 볼 때, Snort만을 사용하여 alarm하는 경우와 동일한 진성 경보 율을 유지하면서 오경보율은 100% 줄일 수 있음을 알 수 있다. 그러나 이러한 방법으로 모든 Snort alarm에 대한 정확한 판단이 가능하다고 할 수는 없다. 예로 기준 비정상도를 30%으로 지정하는 경우 satan의 경우 일부 오판이 되는 경우도 있다. 또 비정상도를 적용하는 방법에는 각 공격마다 개별적으로 적용하는 방법도 있고, 일괄적으로 한 개의 값을 공통으로 적용하는 방법도 있을 수 있다. DARPA data에는 약 200여 개의 unknown type 공격이 Snort를 통해 탐지되는 바, 유형별 적용 문제는 이론적 논거보다는 구현 상의 문제 성격이 더 강하다. 이러한 결합 방법이 실제환경에서 얼마나 효과적인지를 정략적으로 측정하는 실험은 Darpa 1998년 data와 1999년 data를 이용하여 측정하였는데, 그 구현과정과 확장된 모델의 내용이 많아, 이 논문과는 별도의 실험 논문으로 구성하였다. 다만 본 논문에서는 snort와 XIBL의 결합에 의해 침입탐지의 전체 성능을 향상시킬 수 있는 방법이 존재한다는 것을 증명하는 것과, 이것을 만족시키는 모델을 제시하는 것이 주 목적이다. 여기서 주의해야 할 것은 본 논문에서 사용한 침입탐지 성능의 향상이라 함은 단순히 정확도의 수치적 증가를 의미하지 않고 오경보의 줄임을 의미한다. 즉 본 논문에 제시된 모델의 주 목적은 Misuse system과 학습 기반 anomaly system의 결합으로 Misuse System의 오경보를 줄이는 것이 주된 목적인 것이다. 또한 이런 모델의 적용은 어느 정도 진성경보률의 희생을 요구한다. 다만 관리자의 입장에 따라 그 최적 값을 선택해야 할 것이다.

5. 결 론

네트워크 기반 침입탐지 영역에서 Misuse 방식을 주로 사용하고 학습 기반 기법을 보조적으로 사용하려는 시도는 각 방식의 특성을 인정한 것으로, 한가지 방법만을 단독으로 사용하는 것 보다 효과적이다. 학습 기반 방식이 Misuse 방식과 결합하기 위해서는 네트워크 상태를 패킷 단위로 분석하는 방법이 유효하며, 학습 알고리즘적인 면에서 IBL은 C4.5 계열에 비하여 안정적인 학습 능력을 제공하므로, 침입탐지 영역에서의 학습 알고리즘으로 더 적합하다.

본 논문에서는 Snort에서 공격 alarm이 발생한 시점을 중심으로 특정 크기의 패킷에 대하여 XIBL로 패킷의 행동양식을 역추적 평가하고, 정상 및 비정상 클래스의 비율을 측정하여 Snort의 alarm에 대한 진성경보와 오경보를 구분하는 모형을 제시하였다. 이러한 구조로 Snort의 모든 false positive를 줄일 수는 없지만, Snort를 단독으로 운영하는 것보다는 훨씬 효과적이라고 할 수 있다.

추후 과제는 이러한 모형을 기반으로 실제 시스템에 적용하고, 그 시스템의 성능을 정량적으로 측정해 보는 작업이 필요하다. 또 Snort가 탐지하지 못하는 공격 유형을 탐지 할 수 있는 새로운 방법의 연구도 병행되어야 결합 모형의 효용성이 더욱 증대될 수 있다.

참 고 문 헌

[1] S. Patton, W. Yurcik and D. Doss, "An Achilles' Heel in Signature-Based IDS : Squealing False Positives in SNORT," *Lecture Notes in Computer Science*, 2001.

[2] W. LEE, "A Data Mining Framework for constructing

Features and Models for Intrusion Detection Systems," Ph.D. Dissertation, Columbia University, 1999.

[3] I. Weon, D. Song, C. Lee, Y. Heo and J. Jang, "A Machine Learning approach toward an environment-free network anomaly IDS A primer report," In Proc. of 5th International Conference on Advanced Communication Technology, 2003.

[4] C. Kruegel and T. Toth, "Using decision trees to improve signature-based detection," In 6th Symposium on Recent Advances in Intrusion Detection (RAID), *Lecture Notes in Computer Science*, Springer Verlag, USA, September, 2003.

[5] M. Mahoney and P. Chan, "PHAD : Packet Header Anomaly Detection for Identifying Hostile Network Traffic," Florida Institute for Technology Technical Report CS-2001-04.

[6] R. Lippman et als., "Evaluation intrusion detection systems : The 1998 DARPA Off-line intrusion detection evaluation," Proc. Of DARPA Information Survivability Conference and Exposition, pp.12-26, 2000.

[7] K. Julisch, "Mining alarm clusters to improve alarm handling efficiency," In 17th Annual Computer Security Application Conference (ACSAC), pp.12-21, 2000.

[8] K. Julisch and M. Dacier, "Mining Intrusion Detection Alarms for Actionable Knowledge," In 8th ACM International Conference on Knowledge Discovery and Data Mining, 2002.

[9] S. Manganaris, M. Christensen, D. Zerkle and K. Hermiz, "A Data Mining Analysis of RTID Alarms," In 2nd Workshop on Recent Advances in Intrusion Detection (RAID99), 1999.

[10] SNORT, http://www.snort.org.

[11] D. Aha and D. Kibler, "Noise-tolerant instance-based learning algorithms," Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, pp.794-799, 1989.

[12] McHugh, J., "Testing Intrusion Detection Systems : A critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory," ACM Transactions on Information and System Security, Vol.3, No.4, Nov., 2000.

[13] M. Roech, "SNORT-lightweight Intrusion Detection in Networks," USENIX/LISA Conference, 1999.

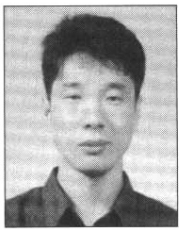
[14] J. R. Quinlan, "Probabilistic Decision Trees, in Machine Learning : An Artificial Intelligence Approach," in *Machine Learning* III, (ed. Yves Kodratoff), Morgan Kaufmann Publishers, Inc., San Mateo, California, pp.140-152, 1990.

[15] C. Stanfill and D. Waltz, "Toward memory-based reasoning," *Communications of the ACM*, 1986.

[16] S. Cost and S. Salzberg, "A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features," Machine Learning 10, pp.57-78, 1993.

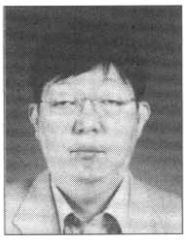
[17] 김도진, "IBL을 사용한 네트워크 기반 침입탐지 시스템과 평가 모델의 연구," 건국대학교 석사학위 청구논문, 2003.

원 일 용



e-mail : clcc@konkuk.ac.kr
1997년 경원대학교 전자계산학과 졸업
1998년 건국대학교 컴퓨터 공학과 석사
2000년~현재 건국대학교 컴퓨터공학과 박사과정
관심분야 : 지능시스템, 네트워크 보안, 복잡성의 과학

송 두 헌



e-mail : mypham@naver.com
1981년 서울대학교 계산통계학과 졸업
1983년 한국과학기술원 전산학과 석사
1994년 캘리포니아대학교 전산학과 박사 수료
1983년~1986년 KIST 연구원
1997년~현재 용인송담대학 컴퓨터 S/W과 교수
관심분야 : 기계학습, 데이터마이닝, CRM, 데이터베이스, 보안, 지능 시스템등

이 창 훈

e-mail : chlee@konkuk.ac.kr
1980년 연세대학교 수학과 졸업
1977년 한국과학기술원 전산학과 석사
1993년 한국과학기술원 전산학과 박사
1996년~2000년 건국대학교 서울캠퍼스 정보통신원 원장
2000년~2002년 건국대학교 정보통신대학원 원장
2001년~2002년 건국대학교 정보통신대학 학장
1980년~현재 건국대학교 컴퓨터공학과 교수
관심분야 : 지능시스템, 운영체제, 보안, 전자상거래