학습을 통한 탐지 모델 생성 시스템

Detection Model Generation System using Learning

김선영
충북대학교 컴퓨터공학과

오창석
충북대학교 전기전자컴퓨터공학과

Sun-Young Kim
Dept. of Computer Engineering, Chungbuk National University
Chang-Suk Oh
School of Electrical and Computer Engineering, Chungbuk National University

중심어 : 데이터 마이닝, 침입 탐지 시스템, 탐지 모델

요 약

본 논문에서는 탐지 모델을 자동 생성하여 인력, 시간에서의 효율성과 오탐율을 향상시키는 학습을 통한 탐지 모델 생성 시스템을 제안한다. 제안된 탐지 모델 생성 시스템은 agent 시스템과 manager 시스템으로 구성되고 agent 시스템은 탐지 모델 데이터베이스를 기반으로 센서의 역할을 수행하고 manager 시스템에서는 탐지 모델 생성과 모델 분산의 역할을 수행한다. 모델 생성은 유전적 알고리즘에 의해 기존의 정형화된 포맷의 탐지 모델을 학습시켜 모델을 생성하고 새로운 탐지 모델로 적용할 수 있다.

실험 결과에 따라 제안된 데이터 마이닝 기반의 탐지 모델 생성 시스템은 기존의 침입 탐지 시스템보다 효율적으로 침입을 탐지하였다. 구현된 시스템으로 인하여 새로운 유형의 침입 시 탐지 모델 생성과, False-Positive율의 감소를 가져와 기존 침입 탐지 시스템의 성능을 개선하여 탐지 모델 생성 시스템을 제안한다.

Abstract

In this paper, We propose detection model generation system using learning to generate automatically detection model. It is improved manpower, efficiency in time. Proposed detection model generation system is consisted of agent system and manager system. Model generation can do existing standardization by genetic algorithm because do model generation and apply by new detection model.

According to experiment results, detection model generation using learning proposed sees more efficiently than existing intrusion detection system. When intrusion of new type occur by implemented system and decrease of the False-Positive rate, improve performance of existing intrusion detection system.

I. 서론

오늘날 인터넷 기술의 급격한 성장과 인터넷을 통한 조직과 개인의 사회적 활동이 증가함에 따라 인터넷 기술의 발전은 데이터 전송속도의 고속화, 인터넷 쇼핑, 이메일, 웹하드등 부가적인 서비스의 발전을 가져와 업무 효율을 향상시키고, 생활의 질을 높여줄 뿐만 아니라 국가 경쟁력 향상에도 긍정적인 효과를 거두고 있다. 그러나 이러한 인터넷의 범용적인 사용과 발전은 항상 긍정적인 측면만 보여주는 것이 아니라, 정보의 유출, 파괴, 서비스 방해, 위조, 변조 등의 컴퓨터 침해 사고가 점차 증가하고 있다. 따라서 컴퓨터 침해 사고의 증가로 인한 정보 시스템과 통신망의 보안이 필요한 시점이다. 이러한 문제점을 해결하기 위한 보안 정책으로는 방화벽을 설치한다거나 침입 탐지 시스템을 설치하는 거라 할 수 있다. 침입 탐지 시스템은 침입 시도와 침입을 정확하게 탐지 할 수 있어야 한다[1].

2000년대에 들어오면서 침입 탐지 시스템에 대한 연구가 점차 증가되고 있으며 많은 시제품들이 상용화되고 있다. 침입 탐지 시스템은 호스트 기반의 침입 탐지 시스템과 네트워크 기반의 침입 탐지 시스템[2],[3]으로 나누어지지만 최근 대부분의 보안 관련 회사들은 네트워크 기반의 침입 탐지 시스템을 사용하는 추세이다. 이러한 침입 탐지 시스템은 False-Positive와 False-Negative가 빈번한 문제점을 가지고 있다. 그러나 이보다 더 큰 문제는 침입을 탐지 할 때 기존의 탐지 모델을 기반으로 비교하여 탐지하기 때문에

새로운 유형의 침입이 발생할 경우에는 속수무책이라는 것이다. 그러므로 새로운 유형의 침입에 대해서는 수동으로 침입을 분석하고 그에 해당하는 탐지모델[4]을 만들어 적용하는데 인력과 시간에서 많은 문제점을 수반하고 있는 현실이다. 따라서 이러한 문제를 해결하기 위해서는 새로운 유형의 침입이 발생할 경우 자동으로 탐지 모델을 생성하는 탐지 모델 생성 시스템 개발이 요구되고 있다. 본 연구에서는 기존의 침입 탐지 시스템의 문제로 제기된 새로운 유형의 공격에 대한 신속한 대처를 위해 효율적으로 탐지 모델을 생성하는 학습을 통한 탐지 모델 생성 시스템을 제안한다. 학습을 통한 탐지 모델 생성 시스템[5]은 기존에 존재하는 탐지 모델을 정형화된 패턴으로 데이터베이스에 저장하고 새로운 유형의 침입이나 알려지지 않은 침입이 발생할 경우 침입 데이터를 분석한 후 분석된 자료를 토대로 데이터베이스에 저장된 모델을 가지고 유전자 알고리즘을 이용하여 학습시켜 탐지를 위한 새로운 모델로 적용하는 시스템이다. 실험을 위해서는 기존의 탐지 모델을 데이터베이스에 맞는 포맷으로 정형화시킨 후 탐지 모델에 있지 않은 공격으로 agent 시스템을 공격한 후 manager 시스템에서 유전자 알고리즘을 적용하여 생성된 모델에 대한 False-Positive율의 변화를 계산한다.

II. 침입 탐지 시스템

1. 침입 탐지 시스템의 개요

침입 탐지 시스템은 침입 즉 자원의 무결성, 기밀성, 가용성등을 파괴하기 위한 일련의 시도들을 탐지하기 위한 보안 시스템이다. 침입 탐지 시스템은 그림 1에서 도시한 바와 같이 데이터 수집 단계, 데이터의 가공 및 축약 단계, 분석 및 침입 탐지 단계, 보고 및 대응 단계로 구성된다.

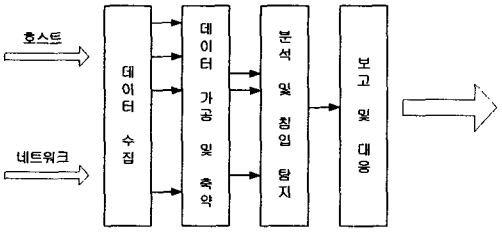

그림 1. 침입 탐지 시스템의 구성

침입 탐지 기법은 일반적으로 단순 정보에 의한 침입 탐지 기법과 수집 정보에 의한 침입 탐지 기법으로 나뉜다. 단순 정보에 의한 침입 탐지 기반은 데이터베이스에 저장된 침입 탐지 규칙 집합에 제시된 기준과 탐지된 개별 패킷 헤더를 단순 비교하여 시스템이 침입을 판정한다. 그러나 수집 정보에 의한 침입 탐지 기법은 일정 기간 수집된 패킷의 헤더를 분석하여 침입 여부를 판정한다. 이는 다시 행위의 결과에 따라 크게 두 가지로 양분할 수 있는데, 비정상적인 침입 탐지 와 오용 탐지로 나눌 수 있다. 비정상적인 침입 탐지란 컴퓨터 자원의 비정상적인 행위에 근거하여 탐지하는 것을 말한다. 예를 들면, 한 사용자가 컴퓨터 사용 시간이 오전 9시부터 오후 5시까지인데, 근무시간 이후에 컴퓨터를 사용하는 경우 올바른 로그인 네임과 패스워드를 사용한 정당한 사용일지라도 침입으로 간주하는 경우를 들 수 있다. 또한 침입 탐지 시스템은 자료 수집원의 대상에 따라 네트워크 기반 침입 탐지 시스템과 호스트 기반 침입 탐지 시스템으로 크게 나눌 수 있다. 그림 2는 네트워크 기반의 침입 탐지 시스템의 배치도를 도시하였다. 크래커가 인터넷을 경유해서 라우터를 통해 침입할 때 침입 탐지 시스템에서 먼저 패킷을 수집한 후 분석하여 침입 여부를 판단하는 구조로 되어 있다.

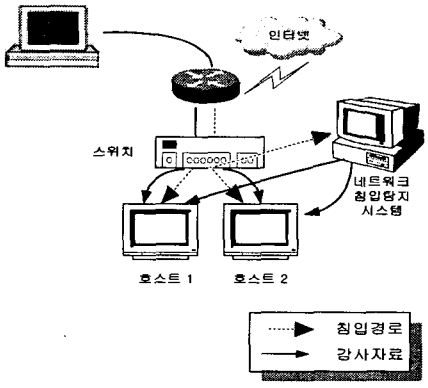

그림 2. 네트워크 기반 침입 탐지 시스템

반면, 호스트 기반 침입 탐지 시스템은 모니터링 하려는 시스템에 설치하여야 하고 단일 호스트로부터 생성되고 수집된 감사 자료를 분석해서 침입을 탐지하게 된다. 호스트 기반 침입 탐지 시스템은 시스템의 로그나 파일 변경 여부를 분석하여 침입 여부를 판정한다. 그림 3은 호스트 기반 침입 탐지 시스템의 구조를 도시하였다.

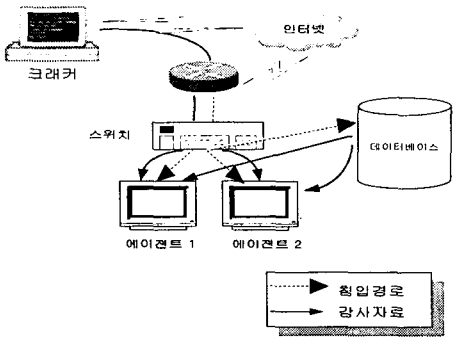

그림 3. 호스트 기반 침입 탐지 시스템

침입 탐지 방법에 따른 분류는 크게 비정상 탐지 기법과 오용 탐지 기법으로 양분할 수 있다. 비정상 탐지 기법은 통계적인 방법을 적용하여 경험적인 자료를 토대로 탐지하는 방법이다. 오용 탐지 기법은 소프트웨어의 취약점이나, 시스템의 취약점을 이용한 공격에 대비하여 이러한 것을 토대로 생성된 탐지 룰에 의하여 탐지하는 방법이다. 이외에도 많은 새로운 방식을 적용한 침입 탐지 시스템이 현재 상용화되고 있으나 침입 탐지 시스템의 근본적인 한계를 개선하는데 중점을 두고 연구 개발하고 있다.

2. 침입 탐지 시스템의 한계

현재 침입 탐지 시스템은 보안 강화를 위한 목적으로 방화벽이나 기타 보안 응용 프로그램 중 가장 선호하고 또한 연구 개발 중에 있다. 그러나 이렇게 선호하고 연구하고 있는 중에도 불구하고 침입 탐지 시스템은 많은 문제점들이 있다. 다음은 현재 침입 탐지 시스템의 문제점으로 지적되고 있는 상황이다[6],[7].

● 궁극적인 대응력이 부족하다.

현재까지는 침입이 발생하게 되면 시스템에서 경고나 운영자에게 메일을 보내는 등 궁극적인 크래커를 차단하는 것이 아니라 크래커를 검출하는 시스템이기 때문에 역추적을 한다거나, 연결 종료 등의 대응력은 부족하다.

● 오탐율이 높다.

최근 들어 다양한 상용 침입 탐지 시스템이 출시되고 개발되고 있으나 공통적인 것은 모든 침입 탐지 시스템이 완벽하게 침입을 판정하는 것이 아니라 정상 연결을 침입으로 판정하는 경우, 침입을 정상 연결로 판정하는 오탐율이 높은 실정이다.

● 속도의 한계를 가지고 있다.

네트워크 기반의 침입 탐지 시스템의 경우는 네트워크상에 많은 트래픽이 있을 경우, 혹은 기가비트의 빠른 네트워크에서는 수집, 대응, 리포팅에 있어 단점을 가지고 있다.

이처럼 침입 탐지 시스템은 많은 문제점을 가지고 있으나 그 중에서 가장 큰 문제는 새로운 유형의 침입이나 알려지지 않은 침입이 발생할 시 탐지 할 수 있는 기본이 되는 탐지 모델이 없다는 것이다. 만약 침입이 이루어지고 분석하여 그러한 침입을 검출하기 위한 탐지 모델을 만들었다 하더라도 이것은 침입이 이루어진 후 이다. 또한 이러한 침입을 위한 탐지 모델을 생성하는 데에도 많은 인력과 비용, 시간이 소요된다. 본 연구에서는 이러한 침입 탐지 시스템의 문제점중의 하나인 새로운 유형의 침입을 고려하여 학습을 통한 탐지 모델 생성시스템을 구현하였다.

III. 학습을 통한 탐지 모델 생성 시스템

본 절에서는 학습을 통한 탐지 모델 생성[5] 시스템의 구성 요소에 대한 설계를 제안한다. 탐지 모델 생성 시스템은 크게 침입을 탐지하고 탐지된 침입 관련 데이터를 전송하는 agent 시스템과 전송된 침입 관련 데이터를 분석하고 학습시켜 새로운 모델을 생성하는 manager 시스템으로 구성된다.

1. 시스템 구조

학습을 통한 모델 생성 시스템은 모델 학습을 기반으로 하여 각 네트워크에 있는 agent로부터 침입 관련 데이터만을 manager 시스템으로 전송하게 된다. manager 시스템은 데이터베이스에 저장되어 있던 기존 모델과 새로운 침입 데이터를 토대로 모델 생성을 위해 학습을 하게 된다. 학습을 통한 새로운 모델이 생성되게 되면 데이터베이스에 저장되고 agent는 데이터베이스와 미러링을 통하여 탐지 모델을 빠른 시간에 업그레이드할 수 있게 된다. 그림 4는 모델 생성 시스템 구성도이다.

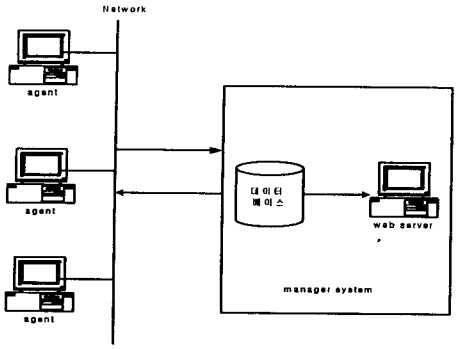

그림 4. 모델 생성 시스템 구성도

1.1. 침입 탐지 agent 시스템

침입 탐지 agent 시스템에서는 센서에서 패킷을 캡쳐하기 위해 libpcap함수를 이용하여 네트워크에 흐르는 패킷을 수집 및 정형화시킨 후 탐지기로 정형화된 데이터를 전송한다. 또한 사후 분석을 위해 데이터베이스로 전송하는 역할을 수행하게 된다. 센서에서 모니터링하고 수집된 데이터는 탐지기에서 감사 데이터를 분석하고 탐지 모델을 사용하여 침입을 탐지한다. 탐지기는 계속적으로 manager 시스템에 있는 모델 데이터베이스에서 탐지 모델을 업그레이드하게 된다. 그림 5는 위에서 말한 agent 시스템 구조를 도시한다.

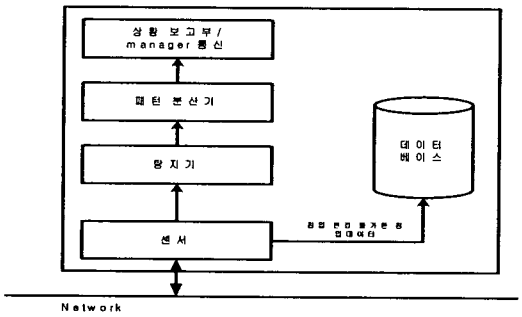

그림 5. agent 시스템 구조

● 센서

센서는 감사 스트림으로부터 정보를 수집하고 데이터를 정형화시킨 후 탐지를 위해 탐지기로 정형화된 데이터를 전송한다. 또한 저장을 위해 데이터베이스로 전송, 모니터링을 한다. 본 연구에서는 네트워크 활동을 모니터링 하기 위한 센서로서 네트워크 기반의 오픈 소스인 snort를 사용하여 네트워크 센서를 이용하였다. 그림 6은 센서에서 패킷을 수집하는 과정을 도시하였다[8].

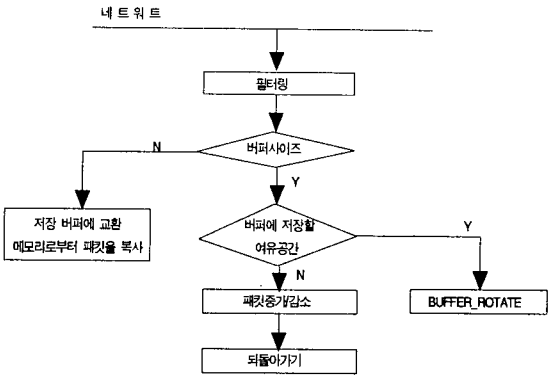

그림 6. 패킷을 수집하는 과정

● 탐지기

탐지기는 센서로부터 수집된 감사 스트림 데이터를 분석하고 탐지 모델을 사용하여 침입을 탐지하고 각각의 기록에 대한 모델 평가를 수행한다. 이것은 탐지 모델을 사용하여 센서에서 수집된 데이터가 침입인지 혹은 정상 연결인지를 판정하여 사후 데이터 보존을 위한 판단을 하게된다.

● 패턴 분산기

패턴 분산기에서는 침입 탐지에 사용되는 성능 메트릭 변수나 시스템 관리 변수를 분석식을 통하여 침입의 종류를 분석하게 된다. 분석 기준은 기존의 분석 자료와 비교하거나 분석식에 의해 계산하게 된다. 분석식에 의해 계산된 값은 허용 임계값을 기준으로 침입의 종류를 분류하게 된다. 패턴 분산기는 모든 정보를 분석하여 프로세스 공격 패턴, 휘발성 공격 패턴, 비휘발성 공격 패턴, 파일 시스템 공격 패턴, 기타 패턴으로 분류되어진다.

● 상황 보고부/manager 통신

센서에 의해 수집된 데이터를 기반으로 탐지기는 탐지 모델에 의하여 침입 유무를 판단하게 된다. 침입으로 판정된 데이터는 agent의 상황 보고부에 의하여 manager 시스템의 데이터베이스로 침입 관련 데이터를 전송하게 된다. 전송을 할 때는 manager 통신부에 해당하는 통신 응용프로그램 rsynk 데몬을 이용하여 통신한다.

2. manager 시스템 설계

탐지 모델 생성 시스템은 침입이 발생하였을 때 agent의 상황 보고부/manager 통신부로부터 침입 관련 데이터를 전송받고 모델 생성기에서는 데이터베이스에 있는 모델들과

비교 후 기존에 있는 침입 모델과 일치하게 되면 알람만 울리고 기존의 침입 모델에 없을 경우에는 모델 생성기에서 학습시켜 새로운 탐지모델을 위한 학습을 수행하게된다. 그림 7은 Manager 시스템 구조를 나타낸다. agent의 상황 보고부에서 받은 감사데이터를 가지고 데이터 정형화기는 정형화된 포맷으로 정형화한 후 데이터베이스에 저장하고 데이터베이스에서는 탐지 모델 생성기에서 전송받은 탐지 모델을 탐지 모델 분산기를 통하여 agent의 탐지기에 배치한다.

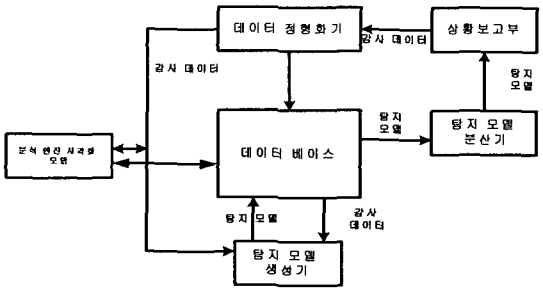

그림 7. Manager 시스템 구조

● 데이터 정형화기

agent 시스템의 상황 보고부로부터 전송 받은 침입 관련 감사자료는 manager 시스템의 정형화기에서 재차 검증을 받게된다. 전송 받은 침입 관련 감사자료가 정확한 침입인지, 기존에 있던 유형의 침입인지, 새로운 유형의 침입인지를 데이터베이스에 저장된 탐지 모델과 비교 분석한 후 데이터베이스에 맞는 포맷으로 정형화하여 데이터베이스로 저장하거나 재차 감사 자료를 분석하게 된다. 분석된 감사자료가 새로운 유형의 침입일 경우는 정형화된 포맷으로 변형한 후 모델 생성기로 전송하게되고 기존에 있던 침입 유형일 경우는 분석 엔진 시각화 모델을 통하여 실시간 볼 수 있도록 전송한다.

● 탐지 모델 분산기

agent의 상황 보고부에 의해 전송된 침입 관련 감사데이터가 탐지 모델에 있지 않고 새로운 침입 관련 감사 데이터일 경우 모델 생성기는 감사 데이터와 기존 탐지 모델을 기반으로 새로운 모델을 학습한 후 모델을 생성한다. 이렇게 생성된 모델들은 데이터베이스에 일단 저장된 후 모델 분산기에 의해 네트워크에서 모니터링하고 있는 agent에 의해 새로운 모델을 신속히 업그레이드할 수 있도록 모델을 배치하는 역할을 수행한다.

● 탐지 모델 생성기

탐지 모델 생성기는 데이터베이스에 저장되어 있는 정형화된 기존의 모델들과 새로운 침입이나 알려지지 않은 침입 시 기존 데이터와의 학습을 통하여 모델을 생성하는 역할을 한다. 모델 생성기에서 학습을 통해 생성된 모델은 신속히 분산기에서 침입을 위해 사용될 수 있도록 배치된다. 모델 생성기에서 사용하는 학술 알고리즘은 유전적 알고리즘을 사용한다. 본 연구에서는 유전적 알고리즘을 이용하여 모델 생성 알고리즘을 위한 최적 조건을 사전에 산출하였고, 산출된 조건으로 모델을 생성하도록 설계하였다. 먼저 컴퓨터 시뮬레이션 범위 안에서 많은 개체의 모집단이 생성되고, 각 개체는 가능한 이론적 모델들을 나타낸다. 각 개체는 원인과 영향 방법에서 개체에 대한 기본적인 설명으로 동작한다. 초기 모집단은 염색체의 무작위 추출에 의해 생성되고 다음 생성의 개체는 임의로 추출됨에 의해 변형을 수행한다. 이러한 유전적 알고리즘은 각 개체가 가능한 행동 모델을 나타내도록 하였다. 그림 8은 유전적 알고리즘에 관한 것이다.

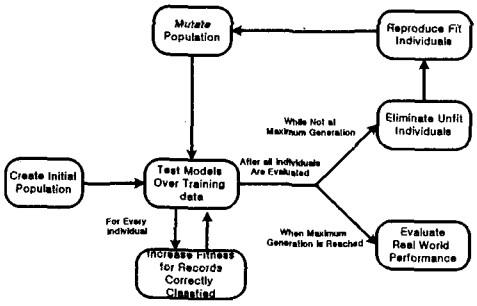

그림 8. 유전적 알고리즘 이론

개체에 대한 성능은 적당 함수에 의해 측정된다. 적당 함수는 알고리즘에 의해 규정된 문제의 결과들과 개체의 염색체들의 결과들을 비교함으로써 개체의 성능을 평가한다. 적당 함수는 일반적으로 처음에 정의한 범위를 가지고 부동 소수점으로 나타내며, 알고리즘에 잘 표현되고 적당하지 않은 것을 제거하는데 사용된다. 함수 F를 위한 적당한 값은 가능한 -1과 이상적인 적당값 1 사이 즉, [-1,1] 사이 안에 있어야 한다. 높은 탐지율과 낮은 False Positive 율은 개체를 위한 적당 함수에서 높은 점수를 나타낸다. 위와 같이 유전적 알고리즘에 의해 생성된 모델은 침입 탐지에서 문제가 되고있는 새로운 유형의 침입에 대한 문제를 해결하고 데이터 분석의 기본이 된다[9].

● 데이터베이스

manager 시스템에서의 데이터베이스는 기존의 탐지모델을 정형화된 패턴으로 정형화한 모델들을 저장한 후 agent 시스템에서 탐지를 위해 역으로 탐지 모델을 생성하고 탐지를 수행할 수 있게 한다. 또한 모델 생성기에서 생성된 탐지 모델을 정형화된 포맷으로 정형화한 것을 새로운 모델로 등록하여 신속한 탐지를 할 수 있다. 그림 9는 기존의 ICMP 침입 탐지 모델을 PhpMyAdmin 이라는 데이터베이스 관리기를 이용하여 정형화된 포맷으로 만든 것을 나타낸다. 테이블의 필드는 type, protocol, extnet, port_ext, homenet, port_home, msg로 나누었다.

```
mysql> DESCRIBE error;
+-----------+--------------+------+-----+---------+-------+
| Field     | Type         | Null | Key | Default | Extra |
+-----------+--------------+------+-----+---------+-------+
| type      | varchar(10)  |      |     |         |       |
| protocol  | varchar(10)  |      |     |         |       |
| extnet    | varchar(20)  |      |     |         |       |
| port_ext  | varchar(4)   |      |     |         |       |
| homenet   | varchar(20)  |      |     |         |       |
| port_home | varchar(4)   |      |     |         |       |
| msg       | varchar(255) |      |     |         |       |
+-----------+--------------+------+-----+---------+-------+
7 rows in set (0.00 sec)
```

그림 9. manager 시스템의 데이터베이스

IV. 실험 및 고찰

본 장에서는 탐지 모델 생성 시스템의 성능을 평가하기 위해 알려지지 않은 침입으로 agent 시스템을 침입하여 성능을 평가하고 일정기간 수집된 데이터를 기반으로 데이터 마이닝 기반의 탐지 모델 생성 시스템과 기존 침입 탐지 시스템의 성능을 비교 분석한다.

1. 실험 환경

우선 실험을 위해 구현할 프로그램은 manager 시스템과 agent 시스템이다. 개발 도구로 Sun MicroSystems사의 java 1.3 SDK와 시스템제어를 위해서 php와 C를 이용하였다. 학습을 통한 탐지 모델 생성 시스템은 생성된 탐지 모델을 이용하여 False-Positive 율을 낮추고 침입 탐지의 정확성을 가져온다. 구현 모델은 PentiumIII 500 시스템 3대에 open source 네트워크 침입 탐지 시스템을 설치하고 또 다른 한 대의 시스템에는 manager 시스템을 설치하여 실험하였다.

본 연구에서 제시한 학습을 통한 탐지 모델 생성 시스템이 효율적이고 정확하게 동작하는지 살펴보기 위해 시간에 대한 모델 수, False-Positive 율에 대한 침입 탐지율, 생성된 모델에 대한 침입 탐지율의 경우를 실험하였다. 그림 10은 실험을 위한 테스트베드 구성이다. 그림에서 manager 시스템은 위에서 서술하였듯이 agent 시스템으로부터 전송받은 침입 관련 데이터와 기존의 모델들을 데이터베이스에 저장한 것과 비교하여 침입을 판정하고, 기존의 모델에 없는 새로운 유형의 침입이 발생할 경우에는 모델을 생성하여 데이터베이스에 저장한다.

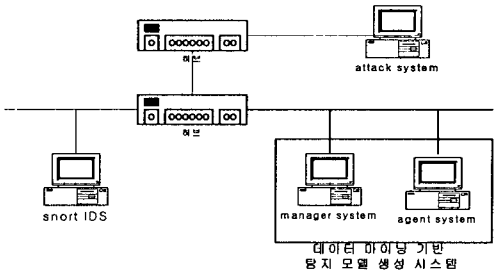

그림 10. 테스트 베드 구성

데이터 전송 및 모델 생성 과정은 다음과 같다.

● 외부 시스템이 네트워크를 통한 agent 시스템에 접근
● agent 시스템은 침입 유/무를 판단하여 침입일 경우 자체 데이터베이스에 저장한 후 manager 시스템으로 침입 데이터 전송
● manager 시스템은 agent 시스템으로부터 전송 받은 침입관련 데이터를 정형화된 탐지 모델로 정형화시킨 데이터베이스의 모델과 비교하여 재차 검증
● 기존의 침입 패턴과 일치할 경우 알람 및 보고 후 침입 데이터 삭제
● 새로운 유형의 침입일 경우 학습을 통한 모델 생성 후 데이터베이스에 모델 추가
● 새로운 모델은 탐지 모델 분산기에 의해 차후 탐지를 위해 모델 분산

학습을 통한 탐지 모델 생성 시스템의 초기 화면은 그림 11과 같다.

```
[root@testnara /]# ./GSODM
Initializing Output Plugins!
initializing agent system.....
initializing manager system.....
starting GSODM.....
[root@testnara /]#
```

그림 11. 초기 화면

학습을 통한 탐지 모델 생성 시스템을 실행하고 정형화에 의해 수행된 데이터베이스 결과는 탐지 모델을 기반으로 동작하기 때문에 탐지 모델이 없을 시에는 탐지를 할 수 없다. 이것은 모든 침입 탐지 시스템의 공통된 문제이다. 학습을 통한 탐지 모델을 실행한 후 210.115.x.x번을 중심으로 포트 스캔을 했을 때 로그 파일의 변화를 실험하였다. 실험한 결과는 그림 12와 같다.

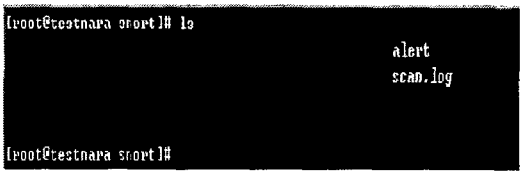

그림 12. 포트 스캔 탐지 모델 추가

실험을 위한 성능 평가 공격은 표 1과 같다.

표 1. 성능 평가를 위한 공격

공 격	프로그램	특 징
remote attack	Tftp버그이용, 패스워드크랙이용, sniff	증폭기처럼 허술한 방어 사이트와 구성 시스템을 사용한 무차별 공격
DDoS 공격	trinoo	tcp 1534,2765,27444,31335 포트를 이용
	TFN2000	UDP, ICMP, TCP등이 복합적으로 사용, 랜덤 포트 사용
	Stacheldraht	tcp 16660, 65000, IMCP ECHO, ICMP RELAY 사용

표 1과 같은 공격 후 나타나는 학습을 통한 탐지 모델 생성 시스템의 성능을 ROC 특성곡선으로 나타내었다. 여기서 ROC 특성곡선은 모델 수에 따른 오탐율의 감소와 모델 생성수를 기준으로 성능을 분석하였다. 위의 실험을 통해서 각각의 ROC 특성곡선을 구할 수 있다. ROC 특성곡선을 측정하기 위한 탐지율과 오탐율을 구하는 부분은 아래와 같다.

● 탐지율 : 침입 탐지 시스템의 탐지 능력을 측정하는 부분

탐지율 = 침입세션의 올바른 분석/ 침입 세션수

실시간 탐지율 = 침입세션 도중 침입 탐지 / 탐지된 침입 세션수

● 오탐율 : 침입 탐지 시스템의 오탐율을 측정하는 부분

False-Positive Rate

= 오탐한 세션수 / 정상행위 세션 수

False-Negative Rate

= 탐지하지 못한 침입 세션수 / 침입 세션수

위와 같은 식에 의해 탐지모델을 증가하여 최종 600개를 기준으로 한 측정 결과이다. 실험 결과 학습을 통한 탐지 모델 생성 시스템은 일반적인 침입 탐지 시스템인 snort 침입 탐지 시스템과 비교하여 0.1%정도 높은 탐지를 하였고 False-Positive율은 0.02%정도 낮게 향상된 것을 표 2에서 볼 수 있다. 표 2.는 탐지 모델 변화에 의한 측정결과이다.

표 2. 탐지 모델 600개를 기준으로 한 측정 결과

항목 / IDS	총 탐지된 패킷	정상 패킷	침입 패킷	탐지율	False-Positive Rate
snort IDS	1027	249	778	0.76	0.32
GSODM	920	125	795	0.86	0.30

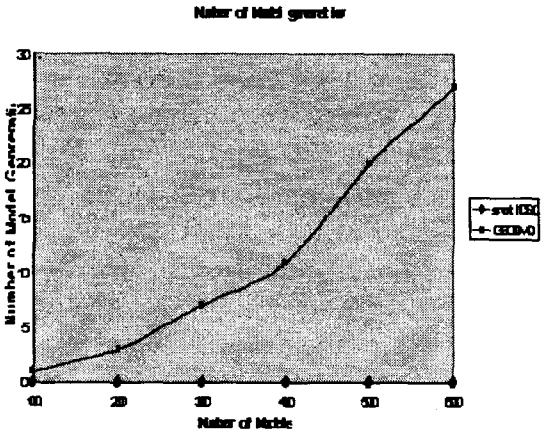

그림 13. 모델 생성 수 비교 결과

그림 13는 모델을 100단위로 증가했을 때 학습을 통한 탐지 모델 생성 시스템과 snort 침입 탐지 시스템을 False-Positive Rate에 대하여 도시한 그림이다. snort 침입 탐지 시스템은 탐지 모델이 증가함에 따라 탐지 모델을 기반으로 하여 False-Positive Rate이 일정 수준 감소하는 것에 비해서 학습을 통한 탐지 모델 생성 시스템은 탐지 모델이 증가함에 따라 새로운 탐지 모델이 생성됨에 의해 False-Positive Rate이 상당한 값으로 줄어드는 것을 나타낸

다. 침입 탐지 시스템에서 가장 중요한 것이 오탐율이 적어야 좋은 것이므로 학습을 통한 탐지 모델 생성 시스템은 오탐율 감소에서 기존 침입 탐지 시스템보다 0.02%의 False-Positive율을 감소시킬 수 있는 것을 보여준다. 그림 14은 모델 증가에 대한 오탐율 변화를 ROC 특성곡선으로 나타낸 것이다.

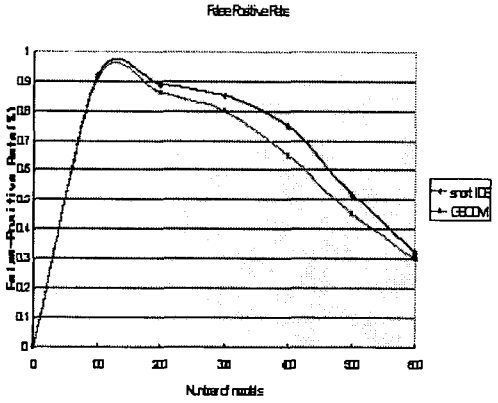

그림 14. 모델 증가에 의한 False-Positive Rate 변화

IV. 결 론

본 연구는 기존 침입 탐지 시스템의 단점을 보완하여 오탐율의 감소를 가져왔고 유전자 알고리즘에 의한 침입 탐지 모델을 생성하는 시스템에 관한 것이다.

특히 탐지 모델 생성에 많은 시간, 인력, 비용이 들게되는 침입 탐지 시스템의 비효율성과 오탐율이 많다는 문제점을 해결하기 위해 agent 시스템과 manager 시스템으로 구성된 학습을 통한 탐지 모델 생성 시스템을 구현하였으며 모델 생성을 위해 기존 모델들을 정형화된 포맷으로 데이터베이스에 저장하여 탐지 모델 생성을 용이하게 하여 학습을 통하여 모델을 생성하게 하여 기존 침입 탐지 시스템의 문제점을 개선하였다.

향후 연구 방향으로는 학습을 통해 생성된 모델이 정확한 탐지를 하기 위해 좀더 효율적인 학습 알고리즘의 보완 연구가 진행되어야 할 것으로 사료된다.

참 고 문 헌

[1] "국내외 침입탐지시스템 제품현황 및 향후 동향", 정보보호 21c, p.98, 2000년 9월

[2] D.E. Denning., "An intrusion detection model," IEEE Transactions on Software Engineering, SE-13:222-232, 1987.

[3] 한국정보보호센터, "호스트기반 침입탐지시스템 개발에 관한 연구", 1998. 12.

[4] S. Kumar, "Classification and Detection of Computer Intrusions," Department of Computer Sciences, Purdue University, PhD Dissertation, Coast TR 95-08, 1995.

[5] A.KGhosh, A.S. Schwartzbard, "A Study in Using Neural Networks for Anomaly and Misuse Detection," Proceedings of the 8th USENIX Security Symposium, Washington, D.C., USA, Auguest pp.23-26, 1999.

[6] 한국정보보호센터, "정보 시스템 현황 및 대응", 1997.

[7] Lee W, stolfo S, Mok K, "Mining Audit Data to Build Intrusion Detection Models," Columbia University, 1998.

[8] 오창석, 데이터 통신, 영한 출판사, 1999.

[9] Adhitya ChitturOssining, "Model Generation for an Intrusion Detection System Using Genetic Algorithms," Ossining High School, November 27, 2001.

김 선 영(Sun-Young Kim) 정회원

2001년 2월 : 한밭대학교 전자공학과 (공학사)
2003년 2월 : 충북대학교컴퓨터공학과 (공학석사)
2003년 3월~ 현재 : 충북대학교 컴퓨터공학과(박사과정)
<관심분야> : 네트워크 보안, Embeded System, 정보 보호

오 창 석(Chang-Suk Oh) 종신회원

1978년 2월 : 연세대학교 전자공학과 (공학사)
1980년 2월 : 연세대학교 전자공학과 (공학석사)
1988년 8월 : 연세대학교 전자공학과 (공학박사)
1985년 ~ 현재 : 충북대학교 전기전자컴퓨터공학부 교수
1982년 ~ 1984년 : 한국전자통신연구원 연구원
1990년 ~ 1991년 : 미국 Stanford대학교 객원교수
<관심분야> : 컴퓨터 네트워크, 뉴로 컴퓨터, 정보 보호