네트워크 기반 비정상 행위에 대한 다계층 침입 탐지 시스템 설계 및 구현

이정현*,김현정*,원일용*,곽주현*,김성학**,이창훈*
*건국대학교 컴퓨터공학과,**유한대학교 컴퓨터공학과
*e-mail:corona,clcc,chlee@konkuk.ac.kr
**e-mail:saint@Yuhan.ac.kr

Design and Implementation of Multi Layer IDS for Network Based Anomaly Behaviors

Jung-Hyun Lee*,Hyun-Jung Kim*,Il-Yong Won*,Ju-Hyun Park*,Sung-Hak Kim**,Chang-Hoon Lee*
*Dept of Computer Science, Kon-Kuk University
**Dept of Computer Science, Yu-Han University

요 약

인터넷 사용자가 급격히 증가함에 따라 정보통신 산업의 발전이 되었지만, 이에 따른 역기능 또한 크게 증가하고 있는 추세이며, 이를 차단하고, 탐지하는 기술이 해킹기술을 앞지르지 못하고 있으며 수동적 입장에서 해킹 사례를 분석하거나, 접근 차체를 차단하는 방법을 택하고 있지만 새로운 해킹 시도에 노출되고, 피해가 계속되고 있다. 따라서 우리가 제안하는 Anomaly IDS는 능동적 입장에서 해킹 기법에 대해 대처하고, 새로운 형태의 해킹기술을 탐지함으로써, 보호하려는 시스템에 대한 능동적 보안수단을 제공한다. 본 논문에서는 기존 Anomaly IDS에서의 문제점을 보완하는 다계층적 측면에서 Audit Data를 통계적으로 학습하여 패턴을 생성하고 탐지하는 시스템을 설계 및 구현하였다.

1. 서론

네트웍을 통한 비정상 행위 탐지를 탐지하는 것은 정상행위에 대한 학습을 통해서 비정상 행위가 있을 경우 이를 탐지하고 경고하는 과정이 Anomaly IDS의 핵심적 구조라고 말할 수 있다. 하지만 정상행위의 로그 자료가 너무 많고 이것을 처리하는데 너무 많은 자원이 필요하다. 그러므로 일반적인 방법으로는 정상 행위 로그 자료를 분석하고 처리하는 것이 불가능하다. 우리는 이러한 대용량의 로그 자료에서 중요한 특징을 선택하고 선택된 특징에 대해 통계처리를 하며, 실시간 탐지를 할 수 있는 패턴으로 만든다. 그러나 이렇게 통계 처리된 패턴은 어떤 비정상적인 행위가 있을 경우 단순히 현재의 상황이 정상 상황에 비해 얼마 비율로 이상하다라는 식으로 경고될 수밖에 없는 단점을 가지고 있다.

통계 처리의 단점을 보완하기 위하여 다계층 침입 탐지(Multi Layer IDS)를 제안한다. 다계층 침입 탐지(Multi Layer IDS)는 정상 행위 로그를 처리할 때 여러 관점에서 그 데이터를 봄으로써 데이터의 특징을 세밀히 표현하는 것이 가능하게 되었다.

본 논문에서 제시한 것은 Network에서의 비정상 행위를 탐지하기 위한 방법으로 한정되며, Ethernet으로 이루어진 네트웍에 내에 있는 시스템으로 들어가는 비정상적 패킷을 탐지하는 시스템이다. 네트웍에서 발생하는 침입을 탐지하는 방법을 2장에서 알아보고, 3장에서 제안한 Multi Layer IDS 분석 및 설계에 대해서 알아본다. 4장에서는 제안한 시스템의 구현에 대해서 알아보고, 마지막으로 5장에서 결론을 맞는다

2. 침입 탐지 방법

침입 탐지는 내부 사용자나 외부의 침입자에 의한 컴퓨터 시스템의 권한 없는 사용함으로써, 시스템에 피해를 주는 행위를 탐지하는 방법이다.

침입 탐지 시스템의 침입 탐지 접근 방법에는 오용 탐지(Misuse detection)와 비정상 행위 탐지(Anomaly detection)방법 있다. 또한 침입 탐지 시스템의 위치에 따라 네트웍 기반 탐지(Network based detection)와 호스트 기반 탐지(Host based detection) 시스템으로 분류 할 수 있다.

오용 탐지는 정해진 공격 패턴을 정의하거나 침입 패턴 DB를 구축함으로써 현재의 네트웍에 있는 필드의 구조, Data Field의 내용이 정의된 공격 패턴과의 유사성을 계산하고 일치하면 침입으로 간주하고,

유사할 경우 침입의 일부라고 여겨지는 행위가 일어나고 있다고 예측할 수 있다.

이런 침입 패턴에 기초한 IDS는 규칙의 형태로 하드코딩 하거나 스크립트 언어로 정의하는 경우가 일반적이고, 다양한 방법으로 침입 행위를 나타낼 수 있다는 장점이 있다. 그러나 여기에서 정의되지 않은 방법으로 침입을 할 경우 탐지할 수 없다는 단점과 취약점이 발견되면 계속 업데이트를 해야 하는 번거러움이 있고, 하드코딩이나 스크립트를 보안 전문가에 의해서 작업이 가능하기 때문에 유지 보수가 어렵다는 단점이 있다.

또한 오용 탐지에는 상태 전이 분석 기반의 침입 모델이 있다. Petri-Net 표현 기법을 이용하는 방법인데 단순 패턴 매칭 기법보다는 침입 행위의 단계에 따라 상태를 전이함으로써 처리시간을 단축할 수 있다는 장점이 있다. 하지만 상태의 정의에서 벗어난 것에 대한 처리의 어려움 있다는 단점이 있다.

비정상 행위 탐지 기법은 정상 행위에 대한 패턴을 학습하여 이에 벗어날 경우 침입으로 간주하는 방법이다. 이 기법은 정상 행위에 대한 많은 양의 Audit Data를 가지고 정상 행위 패턴을 만드는 방법에 따라서 분류할 수 있다.

Data Mining기법은 Audit Data의 연관성을 찾거나 필요 없는 데이터를 줄이는데 이용되고 있다. 하지만 실제 네트웍상에서 발생하는 각 패킷들에 대한 연관성에 대한 연구는 아직 미흡한 편이며, 신경망, 예상 패턴 생성 방법으로 기존의 상태 전이 다이어그램을 사용하는 방법에 Markov 모델링 기법을 사용한 것으로 특정 이벤트들의 발생 후 다음에 특정 이벤트가 발생할 확률을 정의하는 방법 등 이 있다. 마지막으로 통계적 방법을 이용한 방법으로 네트웍을 통해서 들어오는 패킷을 통계적 방법으로 처리하여 학습하여 패턴을 만든다.

3. Multi Layer IDS 분석 및 설계

이장에서는 우리가 제안하는 Multi Layer IDS의 핵심 아이디어를 제시하고, 시스템 설계를 설명한다.

3.1 분석

네트워크의 상황정보를 4가지 관점으로 분류하여 분석함으로써 얻는 장점으로서 첫째, 다양한 관점에서의 Audit Data를 추출함으로서 네트워크의 전반적인 변화부터 세부적인 변화까지 감지 할 수 있다는 것이다. 둘째, 유연하게 새로운 종류의 침입에 대처하는 Anomaly 의 장점을 살리면서 세부적인 감지가 힘들다는 단점을 보완 할 수 있으며, 마지막으로 계층에 따른 다른 보안정책 적용이 가능하다는 것이다. 하지만 Audit Data를 다각적으로 분석하기 때문에 처리량이 많다는 단점이 있지만 이것은 분산처리를 통해서 시스템 부하량을 줄일 수 있다.

Layer	내 용
Network	네트워크 전체에 발생하는 상황에 대한 데이터 -전체 네트워크상의 패킷량 -종류별 패킷량 및 비율 -source ip 의 숫자, 평균 패킷량, 패킷량 편차 -host 당 패킷량의 편차
Source IP	패킷의 Source IP별로 분류한 데이터 -종류별 패킷량 및 비율 -session 연결의 수 -접근 호스트의 수
Destination IP	보호대상인 Target 호스트를 중심으로 분류한 데이터 -source-destination port 쌍의 tcp flag -종류별 패킷량 및 비율 -접근 IP의 패킷량
Session	하나의 세션(연결)단위로 분류된 데이터 -각 시간당 패킷량 -패킷 평균 크기및 편차 -패킷들 사이의 시간적인 간격 편차 -각세션 연결 시간

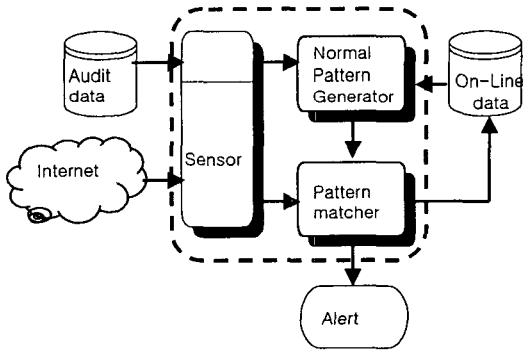
[표1] 각 Layer의 개념적 특징 필드

개념적 다계층을 둠으로써 계산의 양을 줄이고 통계적 처리가 가능토록 전처리 부분에서 부하 없이 Pattern Generator나 Pattern Matcher로 각 필드 값이 들어 갈 수 있도록 한다.

특징 선택 즉, 필요한 packet에서 추출한 이벤트 필드값과 여러 확률에 의해서 생성된 이벤트 필드를 통계적 방법에 의해서 처리하는데 있고, Anomaly IDS에서 통계적 처리방법은 네트워크에 있어서의 데이터 분포를 선택된 데이터를 이용해 표현할 있고, 처리방법이 간단하므로 같은 시간에 더 많은 종류의 요소를 비교분석이 가능하다라는 장점이 있다. 그러나 정규분포형태의 데이터를 가정으로 하고, 각각의 필드가 독립적이므로 여러 필드가 복합적으로 고려되어야 하는 패턴의 감지가 불가능하다라는 단점을 가지고 있다.

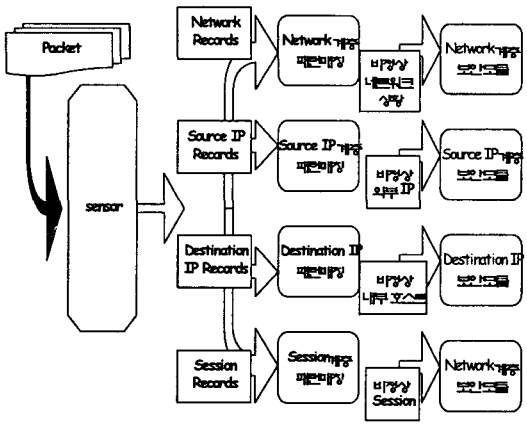
3.2 설계

제안하는 시스템의 구성은 아래 [그림1]과 같은 모듈로 구성되어 있다. Sensor 부분은 Audit data(원시 패킷)를 입력으로 하여 이벤트를 만들고 이 이벤트들의 Normal Pattern Generator에게 보내 패턴을 생성하게 된다. 또 침입탐지 모드일 때 이벤트는 Pattern matcher로 보내져서 패턴 인식 알고리즘에 의해 정상과 비정상을 판단하게 된다.



[그림 1] 시스템의 개념적 구성도

아래 [그림2]는 [그림1]의 점선 안에 있는 모듈을 상세히 나타낸 그림이다.



[그림 2] 핵심부분의 구성도

자료 처리 흐름은 패턴 생성과 탐지의 2가지 모드로 이루어진다. 원시 패킷을 P, 이벤트를 E,패턴은 PT, Alert는 A라고 하면 시스템의 자료 처리 과정은 다음과 같이 표현하면,

1. 패턴 생성 모드(샘플링 모드)

$\{(P_1,P_2,P_3),(P_4,P_5,P_6), \ldots ,(P_{n-2},P_{n-1},P_n)\}$

$\{E_1 , E_2 , \ldots , E_n\}$

PT_n

2. 탐지 모드(탐지 모드)

$\{(P_1,P_2,P_3),(P_4,P_5,P_6), \ldots ,(P_{n-2},P_{n-1},P_n)\}$

$(E_1,E_2,E_3), \ldots ,(E_{n-2},E_{n-1},E_n)$

$A_1 , \ldots , A_n$

이상패턴 판단 알고리즘-통계적 방법중심으로 양적인 변화와 비율적 변화를 사용-은 두 가지 특징을 갖는다.

첫째, 양적인 변화에 중점을 둠으로서 모집단의 평균과 표준 편차를 이용하여 표본 집단의 평균이 모집단의 분포와 이상이 있는가를 판단하며, 확률변수 X 가 $N(m, \partial^2)$를 따른다고 할 때 크기가 n인 표본 집단 의 평균이 $\overline{X}$이라고 하면 H0: 모평균은 모집단의 분포에 비해 이상이 있다.

$$Z= (\overline{X} - m) / (\partial / \sqrt{n})$$

에서 0.05의 유의 수준으로 $|Z| < 1.96$ 이면 가설을 기각(자료의 커다란 변동이 없다. 즉 normal 하다) 하고 그 외에는 가설을 채택-이상이 있다, 즉 abnormal 하다- 한다.

둘째, 방법으로 비율에 중점을 둠으로서 모집단의 분산과 표본 집단의 분산을 이용하여 분산의 차이를 비교하여 표본 집단의 분포를 판단하며, 확률 변수 X의 분산을 $\partial 1$이라고 하고 크기 N인 표본 집단의 분산을 $\partial 2$라고 하면 $|\partial 1 - \partial 2 | > \alpha$ 이면 abnormal 그 외에는 normal 하다고 판단한다.

4. 구현 및 실험

구현 및 실험 환경은 Linux (와우 리눅스 7.0)를 사용했으며, 구현 언어는 ANSI C를 사용했고, 패킷 캡춰 모듈은 libpcap 0.4를 사용하였다.

본 시스템의 구조적 특징은 다음과 같다. 이벤트 프로그래머들은 시스템에서 정의한 4개의 메시지만을 이용하여 자신의 이벤트 모듈을 작성하고, 자신이 정의한 이벤트를 이벤트 맵에 등록하여 사용자 정의 이벤트를 시스템에 설치 할 수 있도록 하는 확장 구조를 가지고 있다. 본 시스템에서 이벤트 프로그래머에게 지원하는 시스템 이벤트는 "INI_", "END_ ", "PACKET_ ", "EVENT_", "PATTERN_ " 이 있다.

시스템은 시작되면 초기화 파일을 로딩하여 환경을 설정 한 후 , 네트웍의 패킷을 모니터링 하기 시작한다. 메인 루프는 4개의 중요한 이벤트가 발생되면 각각의 해당 핸들러 함수를 호출하게 된다. 4개의 메인 이벤트에서의 처리 알고리즘은 아래와 같다.

원시패킷 발생

if 이벤트발생 단위 시간 이상 then
  이벤트 발생 시킴
else
  패킷에서 필요한 정보로 이벤트 가공
end if

이벤트 발생

if 탐지 모드 then
  if Alert발생 단위 시간 이상 then
  Alert발생시킴.
  else
  Alert를 위한 자료 가공
  end if
else if 샘플링 모드 then
  if 패턴 발생 단위 시간 이상 then
  패턴 발생시킴.
  else
  패턴을 발생시키기 위한 가공
  end if
end if

패턴 발생
프로파일을 만듦.

Alert 발생
Monitor 및 Alert파일에 Alert함

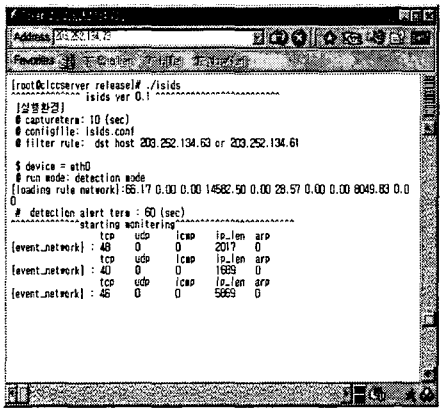
[그림 3]처리 알고리즘

본 시스템에서 구현한 이벤트는 "Network", "Source IP", "Destination IP", "Session"의 4 가지이며, 패턴의 로그파일 및 각종 이벤트 파일들은 가독성을 위해 텍스트모드의 라인 단위 입출력을 사용 하였다.

사용자는 먼저 보호하고자 하는 네트워크를 지정한 후 일정 시간동안 본 시스템에서 정한 이벤트를 생성하여 네트워크의 정상적인 상태를 모델링하여 패턴을 생성한 후 프로파일을 만들게 된다. 정상적인 패턴이 만들어지고 나면 시스템은 탐지 모드로

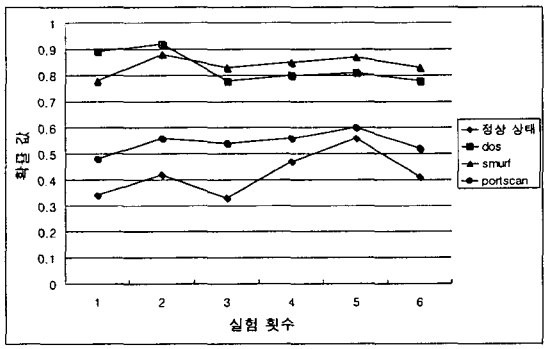
변경되어 네트워크를 감시하게 된다. 먼저 일정 주기동안 원시 패킷을 가공한 후 이벤트을 발생시키고 이 이벤트를 다시 일정 주기 동안 가공하여 Alert패턴을 생성한후 프로파일에 있는 정상적 패턴과 비교하여 Alert패턴의 이상 여부의 정도를 판단하게 된다.

그림()은 본 시스템을 실행했을 때의 스크린 스냅샵으로 실행 모드는 탐지 모드이며 , network 계층의 정상적 패턴을 로드한 상태이며, 이벤트 캡춰 시간은 10초 간격이며 이벤트 체크 간격은 60초로 설정되어 있는 예를 보여준다.



[그림 4] 실행 모습

시스템의 성능 관찰을 위하여 본 연구실에 있는 LAN 상의 특정 파일 서버의 7일간의 네트워크 데이터를 설계된 방법으로 모델링하여 통계적 패턴를 만들고 DOS, smurf, portscan으로 서버를 공격했다. 아래 그래프는 network계층의 이벤트를 관찰한 결과를 그래프로 나타낸 것으로 특정 공격들에 대하여 이상 감지는 비교적 양호하게 나타남을 알 수 있다.



[그림 5] 실험 결과

5. 결론 및 향후 연구 방향

본 논문에서는 Anomaly IDS의 문제점인 속도 향상 및 오버헤드 최소화를 위하여 다계층 필터 및 패턴분석기를 설계 및 구현하였다.

그러나 상위계층(네트워크 > 소스IP = 목적IP > 세션) 과 하위계층간의 관련성을 이용한 복합적 분석에 대한 연구가 부족하다. 따라서 각 계층간, 필드간의 연관관계의 반영을 위해 신경망, 연관 규칙 등을 도입하여 각 층들의 수평적 수직적 관련성에 대한 지속적인 연구가 필요하다.

또 각종 네트워크 기반 침입에 대한 로그 기록들의 데이터마이닝 기법을 통한 패턴 분석으로 각 층의 논리적 필드를 최적화 하는 연구가 필요하다.

참고문헌

[1] Stephan Northcutt, Judy Novak,Donald mcLachlan "Network Intrusion Detection An Analyst' s Handbook" , second Ed, New Riders September,2000

[2] 김주영, 강창구,이극, 이소우 "네트워크 패킷 분석을 통한 침입탐지 기법 개발" ,1999

[3] 성승제, 강창구, 소우영 "네트워크 기반 실시간 침입탐지시스템을 위한 감사자료 수집모듈 설계 및 구현" ,1999

[4] Jean-Philippe Pouzeol , Mireille Ducasse "Handling Generic Intrusion Signatures" ,2000

[5] Steven T. Eckmann , Giovanni Vigna , Richard A. Kemmerer "STATL: An Attack Language for State-based Intrusion Detection ",1999

[6] Giovanni vigna , Richard A,Kemmerer "NetSTAT: A Network-based Intrusion Detection Approach" ,1999

[7] A Data Mining Framework for Building Intrusion Detection Models,199x

[8] R. Agrawal, T.Imielinski, and A. Swami. Mining association rules between sets fo items in large database.1993

[9] 김형종, "지식 기반 시뮬레션 환경을 사용한 분산침입탐지 시스템의 계층적 모델링" ,2000