네트워크 서비스별 이상 탐지를 위한 베이지안 네트워크 기법의 정상 행위 프로파일링

Network based Anomaly Intrusion Detection using Bayesian Network Techniques

차 병 래* 박 경 우** 서 재 현***
ByungRae Cha KyoungWoo Park JaeHyun Seo

요 약

최근 급격한 컴퓨팅 환경의 발전과 인터넷의 확산에 따라 손쉽게 필요한 정보를 획득, 이용하는 것이 가능해지게 되었으나, 이에 대한 역기능으로 전산망에 대한 해커(Hacker)들의 불법적인 침입과 위험이 날로 증대되고 있다. 특히 Unix와 TCP/IP로 구성된 인터넷은 정보 보호 측면에서 많은 취약점을 가지고 있어서, 인증과 접근제어 등의 보안기술만으로는 보안 문제를 해결하기에 충분치 못하였고 정보 보호를 위한 2차 방어선으로 침입 탐지 시스템이 개발되었다.

본 논문에서는 베이지안 네트워크를 이용하여 네트워크 행위를 서비스별로 구분하여 프로파일하는 방법을 제안한다. 네트워크 행위를 서비스별로 구분하고, 서비스별 각 세션에서 TCP/IP 플래그를 통한 행위의 전후 관계를 베이지안 네트워크와 확률값으로 정상 행위를 프로파일링을 수행한다. 베이지안 네트워크를 이용한 정상 행위 프로파일에 의해서 변형되거나 프로파일에 존재하지 않는 새로운 행위에 대해서도 탐지가 가능하였다. 본 논문에서는 DARPA 2000년 침입 탐지 데이터 집합을 이용하여 시뮬레이션을 수행하였다.

Abstract

Recently, the rapidly development of computing environments and the spread of Internet make possible to obtain and use of information easily. Immediately, by opposition function the Hacker's unlawful intrusion and threats rise for network environments as time goes on. Specially, the internet consists of Unix and TCP/IP had many vulnerability. the security techniques of authentication and access controls cannot adequate to solve security problem, thus IDS developed with 2nd defence line.

In this paper, intrusion detection method using Bayesian Networks estimated probability values of behavior contexts based on Bayes theory. The contexts of behaviors or events represents Bayesian Networks of graphic types. We profiled concisely normal behaviors using behavior context. And this method be able to detect new intrusions or modificated intrusions. We had simulation using DARPA 2000 Intrusion Data.

Keyword : Network Anomaly Intrusion Detection, Bayesian Network

1. 서 론

최근 컴퓨터와 통신 기술의 급속한 발전은 사회, 정치, 경제, 문화 등 사회 전반에 걸쳐 막대한 영향을 미치고 활용 영역이 넓어지고 있다. 그러나 세계 각지의 컴퓨팅 환경을 마비시킬 수 있는 웜(worm)의 감염과 해킹이 갈수록 증가하고 있는 실태이며, 컴퓨터 시스템에 대한 공격 방법도 매우 다양해지고 있다.

특히 인터넷의 주요 구성요소인 유닉스 운영체제와 TCP/IP가 정보 보호 측면에서 많은 취약점을 가지고 있어 인터넷에 연결된 모든 전산망이 해커에 의한 공격으로부터 노출이 심각한 실정이

* 준 회 원 : 호남대학교 컴퓨터공학과 전임강사
chabr@honam.ac.kr(제 1저자)
** 정 회 원 : 목포대학교 컴퓨터공학과 부교수
kwpark@mokpo.ac.kr(공동저자)
*** 정 회 원 : 목포대학교 정보보호학과 부교수
jhseo@mokpo.ac.kr(공동저자)
[2004/02/23 투고 - 2004/04/02 심사 - 2004/08/12 심사 완료]

다. 그러므로 인증과 접근제어만으로는 이러한 보안 문제를 해결하기에 충분하지 못하므로 정보 보호를 위한 2차 방어선으로 침입 탐지 시스템(IDS : Intrusion Detection System)이 개발되어 졌다.

침입 탐지 모델은 오용(Misuse) 침입 탐지와 이상(Anomaly) 침입 탐지로 분류가 된다. 오용 탐지는 알려진 침입 방법들을 수집하여 지식 베이스에 유지하고, 동일한 침입 유형을 지식 베이스 검색을 통한 비교에 의해 침입을 탐지하는 방법이다. 또한, 이상 탐지는 정상 행위로부터 벗어나는 주목할만한 특이한 행위 패턴을 침입으로 규정하여 침입을 탐지한다. 일반적으로 오용탐지 방법이 많이 상업화되어 사용되지만 새로운 침입 패턴과 변형된 침입 패턴을 탐지할 수 없는 문제점이 있으며, 오용 탐지를 위한 공격 유형을 분석하여 오용 탐지 규칙 등의 인코딩 작업에 시간과 비용이 많이 소요되는 문제점을 해결하지 못하고 있다[2,3,4]. 해결책으로 정상 및 비정상 행위로부터 침입을 탐지하는 이상 침입 탐지 연구가 진행되고 있으나 아직은 연구 단계에 있으며 상업화되지는 못하고 있다.

네트워크 기반의 이상 침입 탐지 시스템은 네트워크 상의 패킷 데이터를 수집하여 이상 침입을 탐지하는 시스템이다. ADAM[5], NIDES, SPADE 등의 네트워크 이상 탐지 시스템들은 감사 데이터로는 패킷의 헤더 정보인 IP 주소, 포트, TCP 상태 등을 이용하여 이상 행위를 탐지한다. Matthew는 네트워크 이상 탐지 시스템에 PHAD와 ALAD의 두 요소로 구성하여, 패킷 헤더 데이터의 이상 탐지와 응용 계층의 이상 탐지를 수행하였다[6].

본 논문에서는 네트워크 기반의 이상 침입 탐지를 위하여 베이지안 기법을 적용하고자 한다. 불확실성을 처리하는 베이지안 이론을 이상 침입 탐지영역에 도입하여 적용함으로써 오용 탐지의 한계성을 극복하여 알려지지 않은 침입 탐지를 하고자 한다.

본 논문의 2장은 관련연구로써 베이지안 이론과 베이지안 네트워크에 대한 내용을 기술한다. 3장은 침입 탐지를 위한 베이지안 기법으로 네트워크 기반의 이상 탐지에 베이지안 기법을 적용을 위한 방법을 제시하였으며, 4장은 베이지안 기법을 적용한 이상 탐지 기법을 DARPA 침입 탐지 평가 데이터[7]를 이용하여 시뮬레이션을 수행한다. 그리고 5장에서는 결론 및 향후 연구방향을 기술한다.

2. 관련 연구

베이지안 이론은 확률, 댐스트-쉐퍼 이론 그리고 퍼지 이론 등과 같이 불확실성을 처리하기 위한 하나의 방법이다. 불확실성이란 의사결정을 하기 위해 필요한 정보가 부족한 상황을 의미하며 불확실성의 원인으로는 정보의 유실에 의한 부분 정보만의 존재, 정보들 간의 충돌, 정보에 대한 신뢰성의 부족, 지식표현 언어의 한계 등을 들 수 있다.

베이지안 이론을 이용한 연구는 다른 학문 분야인 패턴인식, 데이터마이닝, 신경망, OR 그리고 바이오 인포메틱스 등의 영역에서 우수한 성능을 보이고 있다. 그러므로, 이러한 불확실성을 처리하는 베이지안 이론을 이상 침입 탐지영역에 도입하여 적용이 가능하다.

베이지안 이론은 사건 A가 발생한 후 사건 B가 발생할 확률인 조건부 확률 $P(B|A)$의 역 확률 $P(A|B)$를 간단하게 산출할 수 있다. 역 확률 $P(A|B)$는 나중에 발생한 사건 B에 대하여 먼저 발생한 사건 A의 확률을 의미한다. 즉, 증상이 나타나서 그 문제의 원인을 찾으려는 의료 분야, 장비 진단 분야 등 다양한 분야에 적용되고 있다.

즉, 베이지안 이론

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)},$$

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{P(E|H_1)P(H_1)+\cdots+P(E|H_n)P(H_n)}$$

에 의해 결과에 따른 문제의 원인을 찾으려고 한

다. 베이지안 이론을 적용하기 위해서는 먼저 각 사건이 독립적이고, 명확한 사전 확률이 요구된다. 매우 복잡한 문제에는 적합하지 않지만, 잘 정의된 좁은 영역의 문제 해결에는 매우 유용하다.

베이지안 네트워크는 베이지안 이론의 조건부 독립을 그래프 이론에 따른 네트워크 형태로 표현한 것이다. 즉, 실세계의 지식을 확률이 부여된 방향성 비순환 그래프로 표시하며, 인과 네트워크 또는 신뢰 네트워크라 한다.

베이지안 네트워크에 의해서 표현된 지식을 이용하여 추론이 가능하다. 추론의 종류는 인과 추론, 분석 추론 그리고 상호 인과 추론이다. 인과 추론은 사건 A에 의해 사건 B가 발생한다고 할 때, 사건 A의 값을 알면 사건 B의 확률을 계산할 수 있다. 분석 추론은 사건 A에 의해 사건 B가 발생한다고 할 때, 사건 B 값을 알면 사건 A의 확률값을 계산할 수 있다. 상호 인과 추론은 사건 A와 B 모두 사건 C의 원인이 된다고 할 때 사건 C를 알면 사건 A의 확률 변화가 사건 B의 확률에 미치는 영향을 계산할 수 있다.

침입 탐지 분야에서의 베이지안 기법의 연구는 초기 연구단계에 머물러있다[5,8,9,1]. ADAM[5]은 네트워크 이상 탐지에 사전 정보를 이용하여 패킷을 분류하는 Naive Bayes 분류기를 사용하였다. Mehdi Nassehi[8]에 의해 정상 사용자로 가장하는 침입자를 탐지하기 위하여 베이지안 기법의 탐지 방법이 기술 보고서로 발표되었고, Steven L. Scott[9]는 네트워크 침입 탐지 영역에 침입 패턴 인식과 이상 탐지 방법으로써 베이지안 기법을 적용하여 감사 데이터와 침입 모델과의 관계를 확률값으로 표현하였다. 차병래[1]는 프로그램 행위 기반의 이상 침입 탐지 연구에 베이지안 기법을 적용하여 UNM 데이터로 시뮬레이션을 수행하였다.

3. 베이지안 기법의 이상 탐지

베이지안 기법을 이용한 이상 탐지는 통계적 기법의 이상 탐지이다. 그렇기 때문에 베이지안 기법의 이상 탐지를 위해서는 먼저 정상 데이터로부터 사전 확률 정보를 획득하여야 한다. 더불어, 이상 행위를 탐지하기 위해서는 정상 행위 데이터를 이용한 정상 행위에 대한 프로파일링 구축과 행위나 이벤트를 기술하는 행위 패턴 생성 과정이 필요하다.

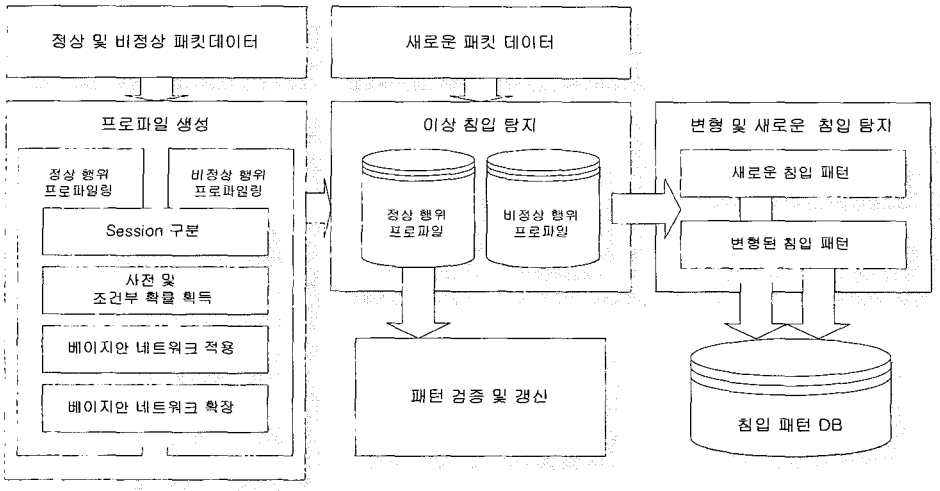
정상 행위가 이루어진 시스템의 네트워크 패킷 데이터를 이용하여 정상 행위 데이터를 수집한다. 그리고 전처리 과정으로 정상 행위 데이터를 세션 단위로 구분한다. 세션에 의해서 행위의 시작과 끝을 구분하고 하나의 행위 단위로 간주하기 위해서이다. 세션의 구분을 위한 척도로는 네트워크 패킷 데이터의 발신지와 수신지의 IP, 포트 번호 그리고 TCP 상태에 의해서 세션을 구분한다.

세션 단위로 구분된 정상 행위 데이터로부터 행위 또는 이벤트에 대한 사전 확률 정보를 획득한다. 그리고 각각의 행위와 이벤트들 간의 관계를 베이지안 확률에 의해서 산출한다. 세션 단위로 구분된 정상 행위 데이터를 클러스터링하여 유사도가 가까운 행위를 베이지안 네트워크를 이용하여 정상 행위를 프로파일링한다. 새로운 행위나 이벤트에 대해 사전 확률과 사후 확률값 계산에 의해서 이상 침입 행위를 탐지한다. 베이지안 기법을 이용한 이상 침입 탐지 구성도는 그림 1과 같다.

이상 침입 탐지를 위한 베이지안 확률값 계산은 사전 확률과 사후 확률, 우도 함수 등을 이용하여 다음과 같은 식(1)으로 제시할 수 있다.

$$P(N|E) = \frac{P(E|N)P(N)}{P(E)} = P(N)\frac{P(E|N)}{P(E)} \quad (1)$$

사전확률 $P(N)$는 정상 행위의 발생 빈도에 의해 좌우되며, 정상 행위에 대한 완전한 정보를 제공하지 못한다. 그러나 사후 확률 $P(N|E)$는 이벤트 E라는 조건에 의한 가장 유력한 정상 행위 이벤트는 $P(N|E)$부분에서 집중적인 확률 분포를 보일 것이다[10].



〈그림 2〉 베이지안 기법의 이상 침입 탐지 구성도

사후 확률 $P(N|E)$은 사전 확률 $P(N)$와 우도 함수 $P(N|E)$의 곱에 이벤트의 확률 $P(E)$로 나눔으로써 계산될 수 있다. 즉, 임의의 정상 행위 N_i와 N_i에 대한 사전 정보와 최우도 함수값에 의해서 가장 가능한 정상 행위와 이벤트에 대한 확률값의 정보를 획득할 수 있다[10]. 가장 가능한 정상 행위 N_j 는 확률값 $P(N_j|E)$가 1에 근접한다면 가장 유력한 정상 행위가 되며, 행위 N_i는 확률값 $P(N_i|E)$가 0에 근접하면 정상 행위와는 무관하게 되어 이상 침입 행위로 간주한다.

본 논문에서는 다음과 같은 제약사항을 전제로 연구를 진행하였다.

제약 1) 모든 행위나 이벤트들은 전후의 시간적 순서 관계에 의한 순차 과정으로 이루어짐을 가정한다.

제약 2) 모든 행위나 이벤트들은 조건부 독립임을 가정 하에 DAG(Direct Acyclic Graph)를 이용해서 베이지안 네트워크를 표현한다.

베이지안 네트워크를 표현하기 위해서 DAG를 사용하는데, DAG는 초기상태(◎), 방향성 아크(→), 시스템 호출(이벤트)의 집합(E), 상태(○) 그리고 상태의 확률(P)로 구성한다.

임의의 행위를 나타내는 연속적인 이벤트 $(E_1,\cdots,E_{i-1}, E_i)$에 대해서 정상 행위로 인식할 정상 행위 확률 $P(N|E_1,\cdots,E_i)$은 결합 확률 함수를 이용하여 다음의 식(2)와 같이 바꿔 쓸 수 있으며,

$$P(N|E_1,\cdots,E_i) = P(N|E_1,\cdots,E_{i-1})\frac{P(E_i|N,E_1,\cdots,E_{i-1})}{P(E_i|E_1,\cdots,E_{i-1})} \quad (2)$$

위의 식(2)으로부터 정의 1)에서부터 정의 4)까지 표 1에 정의한다[1].

〈표 1〉 베이지안 네트워크의 정의

정의 1) 이벤트 연속처리 과정

연속적인 이벤트 $E=(E_1,\cdots,E_{i-1},E_i)$에 대한 정상 행위 확률값 계산은 $P(N|E_1,\cdots,E_{i-1},E_i)$으로 각 상태는 전 단계에 독립임을 정의한다.

정의 2) 이벤트의 분기처리 과정

P_{j-1} 상태에서 분기시 정상 행위 확률값 계산은 $P_j=P(N|E_j,E_{j-1},\cdots)$와 $P_{j+1}=P(N|E_j,E_{j-1},\cdots)$이고, P_j와 P_{j+1}의 정상 행위 확률값은 동일하다고 정의한다.

정의 3) 이벤트의 병합처리 과정

P_{k-1}과 P_k의 상태에서 병합시 정상 행위 확률값 계산은 $P_{k-1}=P(N|E_{k-1})$와 $P_k=P(N|E_k)$의 결합확률 함수로써, $P_{k-1}=P(N|P_{k-1},P_k)$으로 정의한다.

정의 4) 간접 관계

각 상태의 전이 과정 중에서 간접 관계 발견시 각 상태에 벌점에 해당하는 확률값을 적용한다. 벌점 δ는 $(1-\beta)^2$으로 정의하며, β는 베이지안 확률값이다.

정의 1)에서 4)까지를 적용하여 베이지안 네트워크를 이용한 네트워크 기반의 정상 행위 프로파일링을 구축한다.

3.2 네트워크 데이터의 행위 패턴 생성

네트워크 정상 행위의 프로파일을 구축하기 위해서는 하나의 행위를 기술할 수 있는 표현법이 필요하다.

본 논문에서 사용하는 네트워크 행위를 나타내는 표현법은 다음의 표 2와 같다.

〈표 2〉 네트워크 행위 패턴 표현법

네트워크 행위 패턴 표현법	
플래그 값	네트워크 패킷 헤더에 포함된 플래그를 사용
<, >	패턴의 시작과 끝은 각각 <와 >으로 표시하거나, 순차 패턴의 분기와 병합을 표시.
-	패킷과 패킷을 '-'에 의해 구분
X	심볼 X는 모든 플래그에 대응
[ ]	[ ]중괄호는 다양한 플래그를 의미
{ }	{ }중괄호는 제외된 플래그를 의미
( )	괄호()는 반복을 의미
$A,\cdots,Z$	임의의 심볼은 특이 패턴을 정의

네트워크 행위를 표현하기 위하여 DARPA 2000년 NT 데이터 일부를 표현하면 다음의 표 3과 같이 나타낼 수 있다.

〈표 3〉 네트워크 행위의 표현 예제

<S-./ack(2)-P/ack-./ack-P/ack(3)-./ack-P/ack-./ack-P/ack(2)-./ack(2)-P/ack(2)-./ack(2)-P/ack(2)-./ack(2)-P/ack-./ack-F/ack>

표 3과 같이 표현된 네트워크 행위들을 모아서 정상 행위 프로파일 구축에 사용된다.

3.3 네트워크 기반의 이상탐지

네트워크 기반의 이상 침입을 탐지하기 위해서는 먼저 베이지안 네트워크를 적용하여 패킷 데이터의 정상 행위를 프로파일링 한다.

3.3.1 서비스별 정상 행위 프로파일링

네트워크 기반의 침입 탐지에는 네트워크 데이터인 패킷의 헤더 정보를 이용하여 이상이나 오용 침입을 탐지한다. 본 논문에서는 TCP/IP 기반의 서비스에 대한 네트워크의 패킷 헤더 정보를 이용하여 서비스별로 분류하며, 네트워크 서비스별로 정상 행위를 프로파일링하여 이상 침입을 탐지한다.

대부분의 네트워크 침입 탐지는 단지 TCP/IP의 패킷의 이상 유무와 침입시의 패킷의 여러 특징에 의해서 이상 침입을 탐지한다. 본 논문에서는 패킷의 헤더 정보에다가 특정한 서비스에 대해 제약을 적용함으로써 네트워크 이상 침입을 명확히 구분하고자 한다.

FTP 서비스의 정상 행위를 베이지안 네트워크를 이용하여 정상 행위 프로파일링을 수행한다. FTP 서비스의 패킷 데이터을 이용하여 제어 포트와 데이터 포트의 동작을 베이지안 네트워크로 프로파일링하면 그림 2와 같이 표현된다. FTP 서

비스는 제어 포트에서 GET 이나 PUT 명령이 수행되면 데이터 포트가 열리고 데이터가 전송된다. 이런 경우에 정의 2)에 의해서 베이지안 네트워크의 분기 처리 과정이 표현된다. 데이터 전송이 완료되면 데이터 포트가 닫히는데, 이 경우에 정의 3)의 병합 처리 과정이 표현된다.

〈표 4〉 FTP와 FTP-data 세션 분류

Inside Tcpdump Data의 FTP와 FTP-data 세션 : 766					
FTP-data : 599		FTP : 117			
최소 패킷	5 > 3	최소 패킷		1 > 0	
최대 패킷	1508 > 231	최대 패킷		122 > 106	
완료 세션 : 746		미완료 세션 : 20			
745	리셋	리셋		단방향	6
	1	9	1	4	

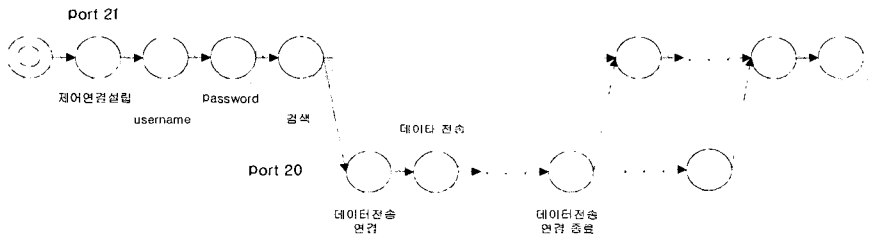
본 논문에서는 MIT 링컨 대학의 2000년 NT 침입 탐지 데이터 집합의 Inside Tcpdump Data를 이용해서 연구를 수행하였다. 침입 탐지 데이터 집합에서 FTP 서비스의 제어 포트와 전송 포트에 해당하는 패킷을 필터링하여 발신지와 수신지 IP 그리고 TCP 통신 절차에 의해서 세션을 구분한다. 발신지와 수신지 IP, 포트 번호 그리고 TCP 통신 절차에 의한 구분된 하나의 세션은 하나의 네트워크 행위가 된다. Inside Tcpdump Data집합은 26,606개의 패킷과 TCP 프로토콜의 절차에 의한 766개의 세션으로 분류되며, 표 4와 같다.

766개 세션의 패킷 정보를 이용하여 각각의 FTP와 FTP-data 서비스에 대한 하나의 행위를 베이지안 네트워크를 이용하여 각각 프로파일링한다. 클러스터링하면 48개의 그룹으로 분류되며, 각 그룹마다 베이지안 네트워크를 이용하여 프로파일링을 확장한다. 그리고 각 그룹의 베이지안 네트워크는 FTP와 FTP-data 서비스의 동작을 나타낸다.

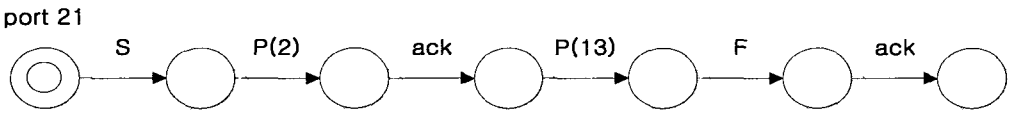
표 3의 네트워크 행위 예제를 이용하여 베이지안 네트워크를 작성하면 FTP 서비스에 대한 하나의 행위가 되며, 그림 3과 같다.

관련된 행위들이 모여서 하나의 프로파일링 생성된다. 그리고 시간 정보에 의해서 근원지 IP 172.16.112.100이고 port 번호가 20 이고, 목적지 IP가 172.16.113.204이고 port 번호가 1052, 1104, 1106인 FTP와 FTP-data 서비스에 대한 하나의 동작을 프로파일링하면 그림 4와 같다.

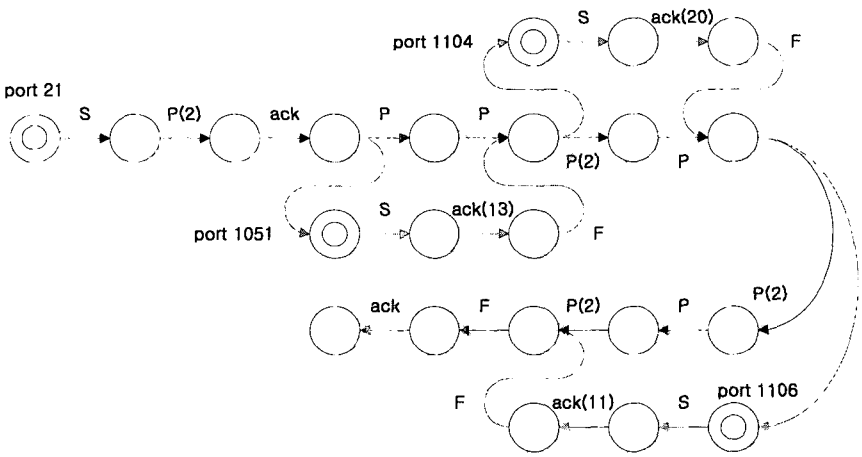
NTIS 공격을 탐지하기 위하여 간단한 네트워크 정상 행위를 프로파일링하기 위하여 RESET 동작이 이루어진 패킷 데이터을 포함하는 세션만을 MSA(Multiple Sequence Alignment) 알고리즘



〈그림 3〉 베이지안 네트워크를 이용한 FTP 서비스 프로파일링

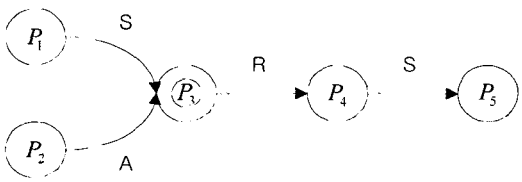


〈그림 4〉 FTP 서비스의 한 행위



〈그림 5〉 확장된 베이지안 네트워크 프로파일링

에 의해 클러스터링한다. 클러스터링된 세션의 패킷 데이터 정보를 이용하여 네트워크의 RESET에 해당하는 정상 행위를 베이지안 네트워크로 프로파일링한다. 클러스터링된 프로파일을 매칭시킴으로써 베이지안 네트워크를 확장시킬 수 있다. 확장된 베이지안 네트워크는 해당하는 네트워크 서비스에 대한 동작을 모델링하게 된다. RESET 동작의 프로파일에서 패킷 데이터가 일치하는 곳이 두 군데가 발견되어 이곳을 기점으로 베이지안 네트워크를 확장하면 그림 5와 같다.



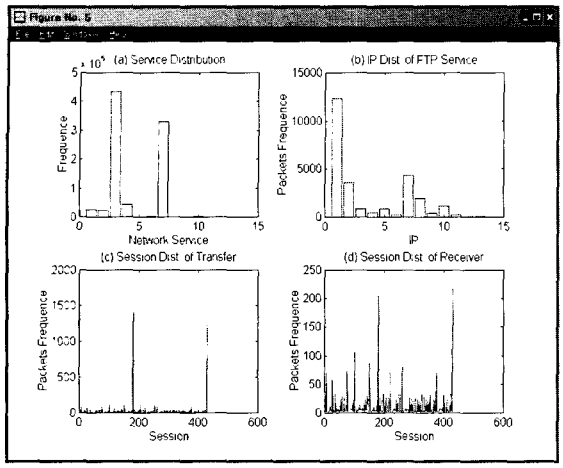
〈그림 6〉 RESET 동작의 프로파일링

RESET 동작의 확장된 베이지안 네트워크를 이용하여 서비스별 네트워크 행위의 이상 침입을 탐지한다. 네트워크 행위의 변형과 새로운 침입의 정의는 확장된 베이지안 네트워크안에 새로운 네트워크 행위가 포함되나 세션에 의한 개별적인 베이지안 네트워크에 존재하지 않으면 네트워크 행위의 변형으로 간주한다. 그리고 네트워크 서비스의 행위를 표현하는 확장된 베이지안 네트워크에 나타나지 않는 행위는 새로운 네트워크 침입으로 규정한다. 이러한 규칙을 이용하여 네트워크 행위의 이상 행위, 변형 그리고 새로운 네트워크 침입 행위를 탐지한다.

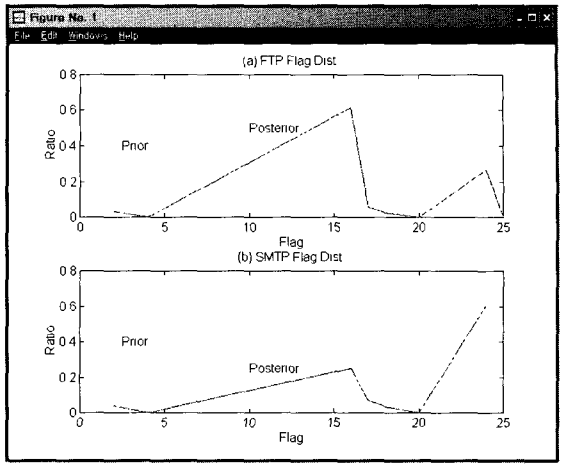
4. 베이지안 기법의 이상탐지 시뮬레이션

베이지안 기법을 적용한 이상행위 탐지를 검증하기 위한 본 시뮬레이션에서는 MIT의 DARPA Intrusion Detection Data 집합의 2000년 NT 네트워크 공격 데이터를 이용하여 네트워크의 FTP 서비스에 대한 공격들을 탐지한다. 시뮬레이션 툴은 Windump, Tcptrace, Perl을 이용하였다. 그리고 그래픽 표현은 Matlab을 이용하였다.

네트워크 기반의 이상 침입 탐지를 위한 시뮬레이션은 FTP와 FTP-Data 서비스에 포함된 CASESEN 공격, SECHOLE 공격 그리고 NTIS 공격을 탐지한다. CASESEN 공격은 NT 시스템의 민감한 객체 디렉토리를 악용하는 U2R 공격이다 , SECHOLE 공격은 공격자가 정상 사용자



(그림 7) 네트워크 이상 탐지를 위한 사전 정보



(그림 8) FTP 서비스와 SMTP 서비스의 플래그별 사전 확률과 사후 확률 분포

로 가장하여 시스템에 SECHOLE 프로그램에 의해서 시스템을 잠그는 공격이다. NTIS 공격은 NT 시스템의 파일 시스템, 접근 허가, 시스템 계정 정보, 그리고 시스템 환경에 대한 정보를 스캔하는 행위이다.

Matthew[11]는 네트워크 기반의 이상 탐지에 네트워크 하위 계층의 33개 특징 정보를 이용하여 이상 탐지를 수행하였다. 본 논문에서는 네트워크의 상위 계층의 정보인 네트워크 서비스의 분류 정보와 세션, 패킷의 플래그를 이용하여 이상 탐지를 수행한다.

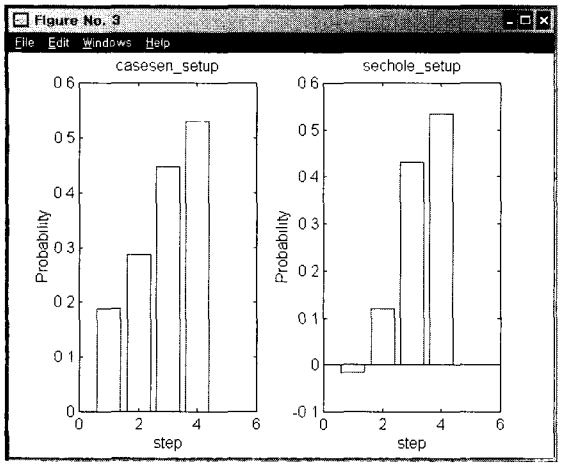
네트워크 기반의 이상 침입을 탐지하기 위하여 서비스별 네트워크 동작의 사전 정보 획득과 베이지안 네트워크를 이용하여 네트워크 서비스를 프로파일링한다.

네트워크의 패킷 데이터를 이용하여 사전 정보와 전처리 과정을 그림 6에 나타낸다. 전처리 과정으로 먼저, 네트워크 서비스별로 분류하고, FTP 서비스에 대해서 IP와 포트 번호, 그리고 TCP 프로토콜의 연결 설정, 데이터 전송, 연결 해제의 통신절차에 의해서 세션을 구분하였고 세션별로 전송과 수신된 패킷 정보 그리고 TCP 플래그에 의한 정보를 획득하였다.

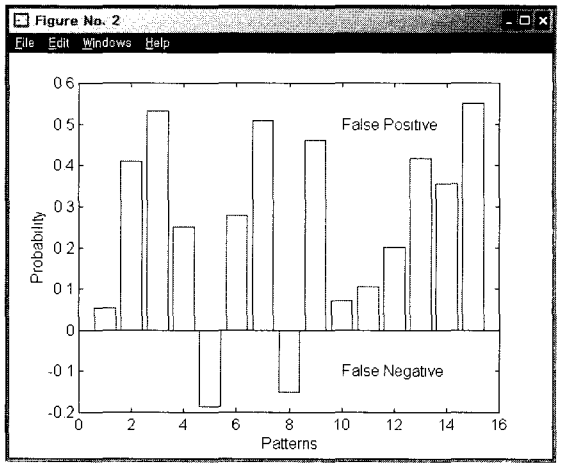
그림 7은 네트워크의 FTP 서비스와 SMTP 서비스의 플래그에 의한 사전 확률과 사후 확률의 분포를 나타낸다. 다른 서비스에 비해 FTP 서비스는 다른 플래그와 비교하여 PUSH와 ack 플래그에 밀집된 분포를 보였다.

그림 4와 같이 구축된 정상 행위 프로파일에 의해서 CASESEN 공격과 SECHOLE 공격을 베이지안 네트워크에 의한 이상 침입은 그림 8과 같이 100% 탐지가 되었다. CASESEN 공격의 경우에는 공격이 진행됨에 따라 단계적으로 이상 탐지 확률값이 증가함을 보였고, SECHOLE 공격의 경우는 첫 세션에서는 전혀 탐지되지 않았으나, 확장된 베이지안 네트워크에 의해서 급격한 이상 탐지 확률값의 변화를 보였다.

CASESEN 공격과 SECHOLE 공격을 베이지안 네트워크에 의한 이상 침입 탐지에 간접 관계를 나타내는 정의 4)를 적용함으로써 그림 9의 (a)와 (c)에 비교해서 (b)와 (d) 같이 확률값이 두드러지게 확대되어 패턴이 명백하게 탐지가 되었다. 베이지안 기법의 정의 4)에 의해서 침입 패턴 분류와 변형된 침입 패턴 탐지에 패턴들간의 명확한 분류를 확대된 확률값으로 제공한다. 그림 9의 (c)와 (d)는 SECHOLE 공격에 대한 4개의 세



(그림 9) CASESEN과 SECHOLE 공격의 탐지율



(그림 10) 베이지안 기법의 NTIS 이상 침입 탐지율

션이 존재하여 하나의 세션을 탐지하지 못하였다. 그러나 정의 4)의 간접 관계에 의해서 SECHOLE 공격 행위의 확률값을 스케일링하는 효과를 가져왔으며, 그림 9의 (c)에 나타난 확률값이 (d)와 같이 확대되어 나타났다. FTP 서비스에 의한 개별적 세션에 의해서는 명확한 공격을 탐지하지 못하지만 간접관계에 의한 탐지 정보를 제공하므로써 공격이 진행됨에 따라 확률값에 의해서 공격을 탐지하였다.

그림 5의 구축된 베이지안 네트워크의 프로파일링에 의해서 ACK 스캔 공격중의 하나인 NTIS 공격에 그림 10과 같이 86%의 탐지율을 보였다. 정상 행위의 RESET 패킷 데이터를 갖는 세션의 빈도가 적어서, 이를 이용한 사전 및 사후 정보를 산출하기 위한 계산량은 작으나, 산출된 베이지안 확률값에 의한 이상 침입 탐지율이 빈약하였다.

FTP 서비스에 대한 정상 행위 세션의 패킷 데이터가 빈약하면, 사전 및 사후 정보의 산출에 계산량은 작았지만, 베이지안 확률값에 의한 이상 탐지가 어려웠다. 결국에는 미탐지가 증가함을 발견하게 되었다. 그러나 한 세션에 대해서만 베이지안 네트워크로 프로파일링하면 미탐지가 존재하지만, 베이지안 네트워크의 확장과 간접관계에 의해 미탐지를 감소시킬 수 있음을 보였다. 역으로, 하나의 세션에 대해 세션을 구성하는 패킷 데이터의 개수가 증가함에 따라 확률값의 계산량이 증가하지만 이상 침입 패턴의 탐지와 분류가 명확하게 이루어졌다.

5. 결 론

인터넷의 활성화에 따라 가용 정보량의 증가와 정보 보호 위협 요인의 증가로 인하여 인증과 접근제어의 보안 기술만으로는 보안 문제 해결에 충분치 못하여 정보 보호를 위한 2차 방어선으로 침입 탐지 시스템이 개발되어 졌다.

본 논문에서는 불확실성 문제를 해결하기 위한 기법 중의 하나인 베이지안 기법을 이상 침입 탐지의 불확실성을 해결하기 위하여 적용하였다. 베이지안 네트워크를 이용하여 네트워크 행위를 서비스별로 구분하여 프로파일하는 방법을 제안하였다. 네트워크 행위를 서비스별로 구분하고, 서비스별 각 세션에서 TCP/IP 플래그를 통한 행위의 전후 관계를 베이지안 네트워크와 확률값으로 정상 행위 프로파일링을 구축하였다. DARPA 2000년 NT 침입 탐지 데이터 집합을 이용하여 이상 침입 탐지를 위한 시뮬레이션을 수행하였다.

DARPA 침입 데이터 집합을 이용한 베이지안 기법의 네트워크 이상 침입 탐지는 FTP 서비스에 대한 CASESEN 공격, SECHOLE 공격 그리고 NTIS 공격을 탐지하였다. FTP 서비스에 대한 정상 행위 세션의 패킷 데이터가 빈약하면, 사전 및 사후 정보에 의해 산출된 계산값이 작아서, 베이지안 확률값에 의한 이상 탐지가 어려웠다. 결국에는 미탐지가 증가함을 발견하게 되었다. 그러나 한 세션에 대해서만 베이지안 네트워크로 프로파일링하면 미탐지가 존재하지만, 베이지안 네트워크의 확장과 간접관계에 의해 미탐지를 감소시킬 수 있음을 보였다. 역으로, 하나의 세션에 대해 세션을 구성하는 패킷 데이터의 개수가 증가함에 따라 확률값의 계산량이 증가하지만 이상 침입 패턴의 탐지와 분류가 명확하게 이루어졌다.

향후, 다양한 데이터에 대한 이상 침입 패턴 분류와 베이지안 확률값에 의한 이상 침입 패턴을 평가하는 기준을 제시하고, 다른 이상 탐지 모델과의 성능 비교 평가가 필요하다. 그리고 변형된 이상 침입 패턴을 효과적으로 탐지하기 위한 침입 패턴 계보 분류에 대한 연구가 요구된다.

참 고 문 헌

[1] 차병래, 박경우, 서재현, "베이지안 네트워크 기반의 변형된 침입패턴 분류 기법", 한국인터넷정보학회 Vol.4, No.2. Apr., p. 69-80, 2003. 4월호

[2] Dorothy E. Denning, "An Intrusion-Detection Model", IEEE Transaction on Software Engineering, Vol. SE-13, No.2, p. 222-232, February 1987.

[3] Steven Noel, D. Wijesekera, Charles Youman, "Modern Intrusion Detection, Data Mining, and Degrees of Attack Guilt", Applications of Data Mining in Computer Security, Daniel Barbara and Sushil Jajodia (eds.), Kluwer Academic Publishers, 2002.

[4] S. Kumar and E. H. Spafford, "A Software Architecture to Support Misuse Intrusion Detection", Proceedings of the 18th National Information Security Conference, p.194-204, 1995.

[5] Daniel Barbara Julia Couto Sushil Jajodia Leonard Popyack Ningning Wu, "ADAM _Detecting Intrusions by Data Mining", Proceedings of the 2001 IEEE Workshop on Information Assurance and Security, 2001.

[6] Matthew V. Mahoney and Philip K. Chan, "Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks", 2002.

[7] http://www.ll.mit.edu/IST/ideval/data/data_index.html

[8] Mehdi Nassehi, "Characterizing Masqueraders for Intrusion Detection", Computer Science/Mathematics, 1998.

[9] Steven L. Scott, "A Bayesian Paradigm for Designing Intrusion Detection Systems To Appear in Computational Statistics and Data Analysis", June 20, 2002.

[10] Christoper M. Bishop, Neural Networks for Pattern Recognition, Oxford Press, p.385-433, 1995.

[11] Matthew V. Mahoney and Philip K. Chan, "PHAD : Packet Header Anomaly Detection for Identifying Hostile Network Traffic", Florida Institute of Technology Technical Report CS-2001-04, 2001.

[12] Christina Warrender, Stephanie Forrest, Barak Pearlmutter, "Detecting Intrusion Using System Calls : Alternative Data Models", 1998.

[13] Paolo Garbolino, Franco Taroni, "Evalu-

ation of scientific evidence using Bayesian Networks", Forensic Science International 125, p.149-155, 2002.

[14] E. Biermann, E. Cloete, L.M. Venter, "A comparison of Intrusion Detection systems", Computers & Security, 20, p. 676 -p683, 2001.

[15] Terran Lane, Carla E. Brodley, "An Application of Machine Learning to Anomaly Detection", February 14, 1997.

[16] Jonatan Gomez, Dipankar Dasgupta, "Evolving Fuzzy Classifiers for Intrusion Detection", IEEE Workshop on Information Assurance, June 2001.

[17] Richard O. Duda, Peter E. Hart, David G. stork, Pattern Classification, 2nd, Wiley, 2001.

[18] Marco Pagni, "Introduction to Patterns, Profiles and Hidden Markov Models", Swiss Institute of Bioinformatics(SIB), August 30, 2002.

[19] Yingjiu Li, Ningning Wu, Sushil Jajodia, X. Sean Wang, "Enhancing Profiles for Anomaly Detection using Time Granularities", Journal of Computer Security, 2000.

[20] Susan M. Bridges, Rayford B. Vaughn, "Fuzzy Data Mining And Genetic Algorithms Applied to Intrusion Detection", 23 rd National Information Systems Security Conference October 16-19, 2000.

[21] Eleazar Eskin, "Anomaly Detection over Noisy Data using Learned Probability Distributions", In Proceedings of the 17 th International Conference on Machine Learning (ICML-2000), 2000.

[22] Guy Helmer, Johnny Wong, Vasant Honavar, Les Miller, "Automated Discovery of Concise Predictive Rules for Intrusion Detection", 2000.

[23] Wenke Lee, Salvatore J. Stolfo, Kui W. Mok, "A Data Mining Framework for Building Intrusion Detection Models", IEEE Symposium on Security and Privacy, 1999.

[24] Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim, "Efficient Algorithms for Mining Outliers from Large Data Sets", Technical report, Bell Laboratories, Murray Hill, 1998.

[25] Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2001.

◐ 저 자 소 개 ◑

차 병 래(Byung Rae Cha)
1995년 호남대학교 수학과 졸업(학사)
1997년 호남대학교 대학원 컴퓨터공학과 졸업(석사)
2004년 목포대학교 대학원 컴퓨터공학과 졸업(박사)
2005년 3월~현재 호남대학교 컴퓨터공학과 전임강사
관심분야 : 정보보호, 컴퓨터 네트워크, 신경망 etc.
E-mail : chabr69@empal.com

박 경 우(Park Kyung Woo)
1986년 전남대학교 계산통계학과 졸업(학사)
1988년 전남대학교 대학원 전산통계학과 졸업(석사)
1994년 전남대학교 대학원 전산통계학과 졸업(박사)
1995.3~현재 목포대학교 정보공학부 부교수
관심분야 : 분산시스템, 컴퓨터 시스템 etc.
E-mail : kwpark@mokpo.ac.kr

서 재 현(Seo Jae Hyeon)
1985년 전남대학교 계산통계학과 졸업(학사)
1988년 중앙대학교 대학원 전자계산학과 졸업(석사)
1996년 전남대학교 대학원 전산통계학과 졸업(박사)
1996. 9 ~ 현재 목포대학교 정보공학부 부교수
관심분야 : 네트워크 보안, 컴퓨터네트워크, 디지털저작권보호기술 등
E-mail : jhseo@mokpo.ac.kr