情報保護學會論文誌
第14券 第5號, 2004. 10

행위 프로파일링을 위한 그래픽 기반의 베이지안 프레임워크

차 병 래
목포대학교 컴퓨터 공학과

The Bayesian Framework based on Graphics for the Behavior Profiling

Byung-Rae Cha
Dept. of Computer Engineering, Mokpo Univ.

요 약

인터넷의 급속한 확장과 새로운 공격 형태의 출현으로 인해 공격 기법 패러다임의 변화가 시작되었다. 그러나, 대부분의 침입 탐지 시스템은 오용 탐지 기반의 알려진 공격 유형만을 탐지하며, 새로운 공격에 대해서는 능동적인 대응이 어려운 실정이다. 이에 새로운 공격 유형에 대한 탐지 능력을 높이기 위해 이상 탐지의 여러 기법들을 적용하려는 시도들이 나타나고 있다.

본 논문에서는 그래픽 기반의 베이지안 프레임워크를 이용하여 감사 데이터에 의한 행위 프로파일링 방법을 제안하고 이상 탐지와 분석을 위한 행위 프로파일을 시각화하고자 한다. 호스트/네트워크의 감사 데이터를 이상 탐지를 위한 준 구조적 데이터 형식의 행위 프로파일인 BF-XML로 변환하고, BF-XML을 SVG로 시각화를 시뮬레이션한다.

ABSTRACT

The change of attack techniques paradigm was begun by fast extension of the latest Internet and new attack form appearing. But, Most intrusion detection systems detect only known attack type as IDS is doing based on misuse detection, and active correspondence is difficult in new attack. Therefore, to heighten detection rate for new attack pattern, the experiments to apply various techniques of anomaly detection are appearing.

In this paper, we propose an behavior profiling method using Bayesian framework based on graphics from audit data and visualize behavior profile to detect/analyze anomaly behavior. We achieve simulation to translate host/network audit data into BF-XML which is behavior profile of semi-structured data type for anomaly detection and to visualize BF-XML as SVG.

Keywords : *Anomaly Intrusion Detection, Bayesian Framework, Behavior Profiling*

Ⅰ. 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 마비시키기 위한 악의적인 웜의 감염과 해킹이 갈수록 증가하고 있는 실태이며, 컴퓨터 시스템에 대한 공격 방법이 다양해지고 새로운 형태의 공격 기법들이 나타나고 있다. 특히 인터넷의 주요 구성요소인 유닉스 운영체제와 TCP/IP가 정보 보호 측면에서 많은 취약점을 가지고 있어

접수일 : 2004년 5월 21일 ; 채택일 : 2004년 9월 26일
† 주저자 : chabr69@empal.com

인터넷에 연결된 모든 전산망이 해커에 의한 공격으로부터 노출이 심각한 실정이다. 이러한 인증과 접근 제어만으로는 보안 문제를 해결하기에 충분하지 못하고 정보 보호를 위한 2차 방어선으로 침입 탐지 시스템이 개발되어 졌다.

침입 탐지 모델은 오용 침입 탐지와 이상 침입 탐지로 분류가 된다. 오용 탐지는 알려진 공격만을 탐지하는 방법이다. 새로운 공격과 변형된 공격을 탐지하지 못한다는 오용 탐지의 단점을 극복하기 위해서 이상 침입 탐지가 연구되고 있으며, 정상 행위로부터 벗어나는 주목할만한 특이한 행위 패턴을 침입으로 규정하고 탐지한다. 이상 침입 탐지를 위해서는 우선적으로 행위에 대한 프로파일이 구축되어야 하며, 구축된 프로파일과 새로운 행위를 비교하여 정상과 비정상을 구분한다. 이상 탐지에 앞서, 어떤 방법으로 행위 프로파일링을 구축하는가가 매우 중요하다.

침입 탐지 영역은 데이터 근원지에 의하여 호스트 기반과 네트워크 기반으로 분류된다. 호스트 기반의 침입 탐지 시스템은 호스트로부터 생성, 수집된 감사 데이터를 근거로 이상 침입을 탐지하는 시스템이다. 감사 데이터로는 프로세스를 이용하며, 프로세스에 대한 원래의 소유자와 그룹, 수행되고 있는 프로세스의 현재 사용자와 그룹, CPU 사용량, I/O 할당량, 이 프로세스가 사용하는 파일, 이 프로세스가 호출하는 시스템 호출 등을 수집하여 정상 패턴과 침입 패턴을 구축하여 여러 가지 탐지 방법을 통하여 침입을 탐지하게 된다. 네트워크 기반의 침입 탐지 시스템은 네트워크 상의 패킷 데이터를 수집하여 침입을 탐지하는 시스템이다. 네트워크 패킷의 헤더 정보인 IP 주소, 포트, TCP 상태 등을 이용한다.

침입 탐지 분야에서의 베이지안 기법의 연구는 초기 연구단계에 머물러 있다. Mehdi Nassehi[1]에 의해 정상 사용자로 가장하는 침입자를 탐지하기 위하여 베이지안 기법의 탐지 방법이 기술 보고서로 발표되었고, Steven L. Scott[2]은 네트워크 침입 탐지에 패턴 인식과 이상 탐지 방법으로써 베이지안 기법을 적용하여 확률 값으로 표현하였고, 데이터와 침입 모델과의 관계를 기술하였다.

본 논문에서는 이상 침입 탐지에서 행위를 표현하는 프로파일 생성에 베이지안 프레임워크를 적용하고 시각화를 위한 BF-XML과 SVG 변환을 제안한다. 행위 프로파일 생성에 베이지안 프레임워크를 적용하므로 써 이상 침입 탐지의 불확실성을 해결한다. 그리고 새로운 침입 패턴과 변형된 침입 패턴도 탐지/분류 정보 및 분석하기 위한 그래픽 표현을 제공한다. 본 논문의 2장은 관련연구로써 침입 탐지를 위한 정상 행위 프로파일링 기법에 대해 조사하고, 침입 탐지 모델, 베이지안 이론 그리고 베이지안 네트워크에 대해 기술한다. 3장은 이상 침입 탐지를 위한 정상 행위 프로파일 생성에 적용할 베이지안 프레임워크를 정의한다. 그리고 XML 기반의 베이지안 프레임워크의 DTD를 정의한다. 4장에서는 제안한 베이지안 프레임워크를 호스트 기반과 네트워크 기반의 행위 프로파일에 적용하며, XML 기반의 베이지안 프레임워크를 적용하여 준 구조화된 데이터 형식과 시각화를 위한 SVG 변환, 그리고 MSA 클러스터링 알고리즘에 의한 패턴 매칭을 수행한다. 5장에서 결론 및 향후 연구방향을 기술한다.

Ⅱ. 관련 연구

2.1 행위 프로파일링과 침입 탐지 모델

이상 침입 탐지를 위해서는 제일 먼저 우선되어야 할 과제가 어떻게 정상 행위를 구축할 것인가가 될 것이다. 이상 침입 탐지를 위해서는 프로파일에 이상과 정상을 구분할 행위를 기술하여야 한다. 시스템 또는 사용자의 행위를 기술하는 것을 행위 프로파일링이라 한다. 행위 프로파일링은 객체에 대한 주체 행위의 특징을 기술하거나 주-객체간의 정상 행위 기술, 또는 이상 징후를 제공한다. 행위 프로파일링 방법으로 행렬과 통계적 모델들이 존재한다[3].

행렬 방법은 주기적으로 누적된 측정된 값을 표현하여 이상 탐지에 정보를 제공한다. 행렬방법은 대부분이 3가지 형태로 정의하는데, 이벤트 계수기, 이벤트의 간격 시간 그리고 사용된 자원의 척도로 구성한다.

통계적 모델은 모든 지식을 행위를 관측함으로써 정보를 획득하며, 행위에 대한 기본 분포를 가정하지 않아도 된다. 통계적 모델의 IDS는 운영적 모델, 평균과 표준 편차 모델, 다변량 모델, 마코프 프로세스 모델 그리고 시계열 모델로 구분한다.[3] 운영적 모델은 고정된 제한과 새로운 관측을 비교하여 비정상을 결정하는 모델이다. 평균과 표준 편차 모델은 이벤트의 합, 합 제곱, 평균, 표준편차 등의 정보를 이용하여 임의의 신뢰구간을 벗어나면 비정상으로 간주한다. 장점은 정상 행위의 제한 설정에 대한 사전 지식이 필요하지 않다는 점이다. 또한 신뢰구간은 증가된

지식을 반영하며, 관측된 데이터에 의존한다. 다변량 모델은 행위 행렬간의 상관관계를 기반으로 좋은 식별 능력은 제공한다. 행위 개별적인 측정보다는 관련된 측정의 조합에 의해 얻어진다. 마코프 프로세스 모델은 여러 이벤트 형태를 상태 변수, 상태간의 전이 빈도로 특징을 기술하는 상태전이행렬을 사용한다. 어떤 명령과 명령 순차간의 전이 조사에 매우 유용한 모델이다. 시계열 모델은 시간 측면에서 행위의 경향을 측정한다. 장점은 행위의 중요한 변화를 단계적 탐지할 수 있다는 점이다. 단점은 평균과 표준 편차 모델에 비해 비용증가가 크다는 점이다.

침입 탐지 모델로는 Denning 모델, Shieh 모델 그리고 Kumar 모델 등이 있다. 침입 탐지에 대한 연구는 1980년 John Anderson에 의해 처음 소개되어 1987년에 일반적인 침입탐지 모델이 제시되었다. Dorothy Denning 모델[3]은 시스템의 비정상적인 형태의 사용에 대해서 시스템의 로그 기록을 모니터링 함으로서 침입을 탐지하는 모델로 미리 정의된 통계적인 방법들을 사용하여 시스템의 행위를 계산하는 변수들을 이용하고 있다.

Shiuh-Pyng Shieh 모델[4]은 직접관계에서 시스템 상태와 상태 전이, 주체와 객체사이의 간접관계를 나타내는 규칙으로 정의된다. 시스템 상태는 감사 추적에서 캡처되고, 보호 그래프로 표현된다. 보호 그래프는 주체와 객체 두 가지 타입의 노드를 가지고 있으며, 주체는 프로세스와 사용자들을 표현하는 능동적인 노드로서 주체와 객체 사이의 데이터와 권한의 흐름을 발생하는 것이다. 또한 객체는 수동적인 노드로서 파일이나 디렉토리와 같은 데이터 컨테이너를 나타내고, 데이터나 권한의 흐름과 같이 행위를 초기화한다. 시스템 상태는 방향에 있는 보호 그래프 $G(V,E,C,F)$로 표현한다. 이 그래프는 노드들의 집합 V, 레이블에 있는 간선들의 집합 E, 보호집합들의 집합 C, 합법적인 흐름 행렬 F로 구성된 구조를 가진다. 즉 노드들의 집합 V는 주체와 객체들로 구성되고 주체 S_i와 객체 O_i는 그래픽 보호로 표현되고 접속 오퍼레이션은 노드 V_1과 V_2 사이에서 발생하고 표현 관계 (V_1, V_2)으로 나타낸다. 여기서 표현 관계는 집합 {*r, w, d, cd*}의 원소이다. 표현관계는 *r, w, d, cd*로 정의하는데 알려진 침입 패턴들은 데이터와 권한의 흐름의 네 가지 형태로 특성화시킨 모델이다.

Sandeep Kumar 모델[5]은 Jensen에 의해서 Colored Petri Net에 근거하고 있으며, 침입 행위는 칼라 페트리 네트로 표현하고 네트에서 하나 이상의 초기 상태들과 하나의 최종 상태는 모델에서 매칭을 정의하기 위해서 사용되는 모델이다. 특히 문맥은 토큰의 칼라로 저장되고, 조건은 가이드식(Guard expression)을 이용하고 행위는 상태의 행위를 이용하여 표현된다.

2.2 베이지안 이론과 베이지안 네트워크

이상 침입 탐지의 프로파일링 구축을 연구하기 위해서 베이지안 이론과 베이지안 네트워크에 대해서 기술한다. 베이지안 이론은 확률, 댐스트-쉐퍼 이론 그리고 퍼지 이론 등과 같이 불확실성을 처리하기 위한 하나의 방법이다. 먼저, 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 바이오인포메틱스 등의 영역에서 우수한 성능을 보이고 있다. 그러므로, 이러한 불확실성을 처리하는 베이지안 이론을 이상 침입 탐지의 프로파일링 영역에 도입하여 적용하고자 한다.

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 의미하며, 즉, 증상이 나타나서 그 문제의 원인을 찾으려는 의료 분야, 장비 진단 분야 등의 여러 분야에 적용되고 있다.

베이지안 이론은

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (1)$$

에 의해서 증상에 의한 문제의 원인을 찾으려고 한다. 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 그래프의 네트워크 형태로 표현한 것이다. 즉,

실세계의 지식을 확률이 부여된 방향성 비순환 그래프로 표시한다. 베이지안 네트워크를 인과 네트워크 또는 신뢰 네트워크라고도 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 확률 값을 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.

Ⅲ. 베이지안 프레임워크

이상 침입 탐지를 위해서는 행위를 표현할 수 있는 방법이 필요하다. 침입 탐지 모델은 Shieh 모델, Denning 모델 그리고 Kumar 모델 등이 존재한다. Shieh 모델은 사용자 행위 영역의 모델이며, Denning 모델은 이상 행위 탐지에 대한 전반적인 프로파일에 대한 모델을 제시하였다. Kumar 모델은 프로그램 행위 영역에 대한 침입에 대한 상태 전이 모델이다. 본 논문에서 제안한 베이지안 프레임워크는 세 영역을 모두를 표현할 수 있고, 새로운 툴의 설치 없이 분석을 용이하게 하기 위한 그래픽 정보를 제공할 수 있다는 장점을 갖는다. 본 논문에서는 행위 프로파일의 행위를 표현하기 위한 문자열을 이용한 표기법과 그래픽 기법인 베이지안 네트워크를 이용한 베이지안 프레임워크를 제안한다.

침입을 탐지하기 위해서는 시스템의 로그 데이터나 패킷 데이터로부터 데이터를 가공하여야 한다. 행위를 표현하기 위해서는 세션이라는 개념을 정의한다. 세션이란 여러 이벤트를 모아서 하나의 행위를 나타내는 것을 의미한다. 호스트 기반의 사용자 행위의 경우는 시스템에 로그인한 후부터 로그 아웃까지를 하나의 세션까지의 사용자가 사용한 명령어와 객체들로 집합으로 세션을 규정한다. 프로그램 행위의 경우에는 프로세스 아이디(PID)가 호출하는 시스템 콜(System Call)의 집합으로 세션을 규정한다. 네트워크 기반의 네트워크 행위는 IP 주소, 포트 번호 그리고 TCP 통신 절차에 의해서 패킷 데이터의 집합을 세션으로 규정한다.

3.1 문자열을 이용한 표기법

정상 행위의 프로파일을 구축하기 위해서는 하나의 행위를 기술할 수 있는 표현법이 필요하다. 행위를 표현하기 위해서는 세션을 기본 단위로 사용한다. 세션을 표현하기 위해서 문자열을 이용한 행위 패턴 표기법을 정의한다. 본 논문에서 사용하는 행위를 나타내는 표현법은 다음의 표 1과 같다.

표 1. 행위 패턴 표현법

메타기호	의미
<, >	세션의 시작과 끝은 각각 <와 >으로 표시하거나, 순차 세션의 분기와 병합을 표시.
-	이벤트와 이벤트를 '-' 에 의해 구분
X	심볼 X는 모든 이벤트에 대응
[ ]	[ ]중괄호는 다양한 이벤트를 의미
{ }	중괄호는 제외된 이벤트를 의미
( )	괄호()는 반복을 의미

호스트 기반 프로그램 행위의 정상 행위 프로파일 구축을 위해 뉴 멕시코 대학의 Sendmail 데이터[6]를 이용하여 프로세스 아이디(PID)별 세션을 표 1의 행위 패턴 표현법으로 나타내면 그림 1과 같이 표현된다. 네트워크의 정상 행위 프로파일 구축을 위해 MIT 링컨 대학의 DARPA 2000년 NT 데이터[7]를 이용하여 FTP 서비스의 세션을 행위 표현법으로 나타내면 그림 2와 같이 표현된다.

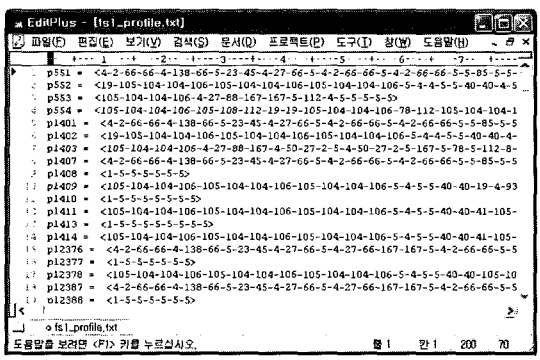

그림 1. 행위 패턴 표현법에 의한 PID별 세션

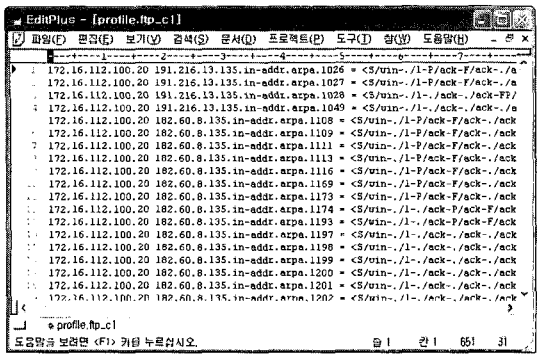

그림 2. 행위 패턴 표현법에 의한 FTP 서비스의 세션

3.2 베이지안 프레임워크

베이지안 프레임워크는 베이지안 네트워크를 기본으로 한다. 베이지안 네트워크는 베이지안 이론의 조건부 독립을 그래프의 네트워크 형태로 표현한 것이다. 행위 프로파일을 세션 정보로 구성되어 있다. 세션은 이벤트의 집합이며, 이벤트들의 시간적 순차적인 관계를 베이지안 네트워크로 표현한다. 즉, 이벤트의 순차 관계를 확률이 부여된 방향성 비순환 그래프(DAG)로 표시한다. 본 논문에서는 다음의 제약사항을 전제로 한다.

(제약 사항) 모든 행위나 이벤트들은 전후의 시간적 순서 관계에 의한 순차 과정으로 이루어짐과 모든 행위나 이벤트들은 조건부 독립임을 가정한다.

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

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 2와 같이 바꿔 쓸 수 있으며,

$$P(ME_1, \cdots, E_i) = P(ME_1, \cdots, E_{i-1}) \frac{P(E_i|N, E_1, \cdots, E_{i-1})}{P(E_i|E_1, \cdots, E_{i-1})} \quad (2)$$

위의 식 2로부터 다음의 정의 1에서부터 정의 3까지를 정의하며, 간접 관계에 대한 정의 4를 정의한다.

[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 $P_j = P(ME_j, E_{j-1}, \cdots)$와 $P_{j+1} = P(ME_j, E_{j-1}, \cdots)$이고, P_j 와 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 $P_{k-1} = P(ME_{k-1})$와 $P_k = P(ME_k)$의 결합확률함수로써, $P_{k+1} = P(MP_{k-1}, P_k)$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)에 의해서 연속되는 이벤트 $E = (E_1, \cdots, E_{i-1}, E_i)$에 대해 베이지안 네트워크로 표현하면 그림 3과 같다. 각각의 이벤트를 수행함에 따라 각각의 상태에 따른 정상 행위 확률값을 갖는다.

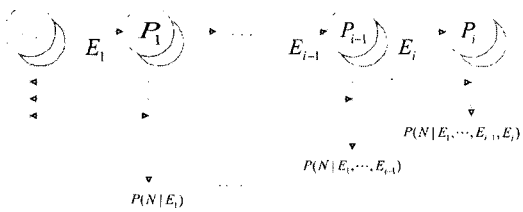

그림 3. 이벤트의 연속처리 과정

정의 2)와 3)은 한 이벤트에 의해서 상태의 분기와 병합을 베이지안 네트워크로 그림 4의 (a)와 (b)로 각각 표현한다. 한 이벤트에 의해 여러 상태로 분기시에는 분기된 모든 상태들은 동일한 확률값을 갖는다. 그러나 두 상태에서 한 상태로 병합시에는 두 상태의 결합 확률값으로 상태의 확률값을 계산한다.

정의 4)는 각 상태의 전이 과정에서 전 상태와의

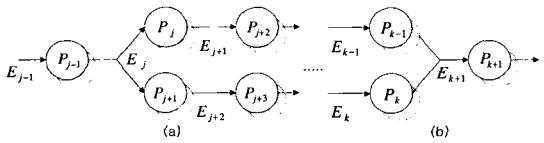

그림 4. 상태전이의 분기와 병합 처리 과정

직접 관계가 아닌 간접 관계가 발생하면 상태에 벌점에 해당하는 임의의 확률값을 부과한다.

위의 정의 1)에서 정의 4)까지를 적용하여 베이지안 네트워크를 이용한 정상 행위를 프로파일링한다. 정상 행위 프로파일링 과정은 호스트 기반과 네트워크 기반으로 구분하여 구축한다. 호스트 기반의 사용자 행위와 네트워크 기반의 정상 행위 프로파일링은 정의1)에서 정의 4)까지 적용하며, 호스트 기반의 프로그램 행위의 정상 행위 프로파일링은 정의 1)에서 정의 3)까지 적용하여 프로파일을 구축한다.

3.3 BF-XML

XML은 웹 상에서 표현하기보다는 데이터를 기술하는 언어이다. XML에 의해 기술된 데이터는 준구조적(semi-structured) 데이터 형식을 취한다. 특히, 웹은 이종 데이터베이스로 구축되어 있으므로, 특별한 데이터베이스 엔진이나, 객체지향 데이터베이스 또는 데이터웨어 하우스 등이 필요하다. 시스템과 데이터에 종속적인 데이터베이스 엔진은 웹 상에서 사용하기 힘들다. 그러나 XML 언어는 일반 웹 언어의 특성, 문법에 맞는 문서(Well-Formed) 형식과 유효한(Valid) 문서 형식을 취하므로 써 준 구조화된 데이터형식으로 자동화할 수 있으며, 특별한 데이터베이스 엔진이 필요하지 않는다.

BF-XML(Bayesian Framework based on XML)은 베이지안 프레임워크를 XML 문서로 정의한 것이다. 먼저, BF-DTD를 표 2와 같이 정의하고, BF-XML에 의한 사용자, 프로그램 행위 그리고 네트워크의 행위 프로파일을 그림 6, 그림 8 그리고 그림 10과 같이 작성할 수 있다.

BF-DTD는 베이지안 네트워크의 요소와 속성을 정의하며, 노드, 아크 그리고 확률 값과 제목 등으로 구성된다.

BF-DTD는 DATA와 STRUCTURE의 두 부분으로 구성된다. DATA 부분은 NODE와 OBJECT로 구성된다. NODE는 능동의 객체이며, 상태를 나타내며, 세션의 이벤트 정보를 이용하여 표시

표 2. BF-XML을 위한 BF-DTD의 정의

```
<!-- DTD for Bayesian Framework -->
<!ELEMENT BF-XML ( DATA | STRUCTURE )+>
  <!ATTLIST BF-XML ID CDATA #REQUIRED
       NAME CDATA #IMPLIED
       TITLE CDATA #IMPLIED>
<!-- Node Data declaration section -->
<!ELEMENT DATA (NODE|OBJECT)*>
  <!ELEMENT NODE EMPTY>
    <!ATTLIST NODE
     NAME NMTOKEN #REQUIRED
          TYPE (S | S1 | G) "G"
          TITLE CDATA #IMPLIED
          PROBABILITY CDATA #IMPLIED
     XPOS CDATA #IMPLIED
     YPOS CDATA #IMPLIED>
  <!ELEMENT OBJECT EMPTY>
    <!ATTLIST OBJ
    NAME NMTOKEN #REQUIRED
    TITLE CDATA #IMPLIED
    XPOS CDATA #IMPLIED
    YPOS CDATA #IMPLIED>
<!-- topological dependency structure
information -->
<!ELEMENT STRUCTURE (ARC)*>
  <!-- specify dependency arc -->
  <!ELEMENT ARC EMPTY>
    <!ATTLIST ARC
     TYPE (D | I) "D"
     PARENT NMTOKEN #REQUIRED
     CHILD NMTOKEN #REQUIRED
         EVENT_NAME CDATA #IMPLIED>
```

한다. OBJECT는 수동의 객체이며, NODE에 의해 접근되는 객체를 나타낸다. STRUCTURE 부분은 ARC로 구성된다. ARC는 NODE들의 관계를 나타낸다. 관계있는 NODE간에 ARC를 이용해서 연결하게 되며, NODE로 표시되는 이벤트들의 전후 관계를 나타낸다. 또한 ARC는 직접 관계와 간접 관계를 구분하여 표기한다. NODE와 ARC를 이용해서 사용자 행위, 프로그램 행위 그리고 네트워크 행위를 프로파일링하게 된다.

베이지안 확률 $P(B|A)$를 BF-XML로 나타내면, NODE는 이벤트 A와 이벤트 B로 구성되며, ARC는 이벤트 A와 B의 순서 관계를 표시한다. NODE 원소는 NODE A와 NODE B로 구성되고, 각각의 확률값 $P(A)$와 $P(B)$의 상태를 갖는다. ARC 원소는 PARENT 항목에는 A, CHILD 항목에는 B 그리고 TYPE 항목은 D가 입력되어 이벤트 순서가

A, 다음의 이벤트가 B이며, 직접관계를 나타낸다.

베이지안 프레임워크는 전 상태에 독립임을 가정하고 있다. 각 NODE는 상태의 확률값을 갖게 되며, ARC에 의해서 이벤트의 순서 관계를 표현한다. 이벤트의 순서 관계를 MSA 클러스터링 알고리즘에 의해서 행위 프로파일을 분류 및 패턴 매칭을 수행하게 된다. 상태의 확률값에 의해서 이벤트의 빈도와 분포 정보를 얻을 수 있다.

Ⅳ. XML 기반의 베이지안 프레임워크에 의한 행위 프로파일링

이상 침입을 탐지하기 위해서는 행위의 영역에 따라 호스트 기반과 네트워크 기반으로 분류된다. 호스트 기반의 이상 행위는 또한 사용자 행위와 프로그램 행위로 구분된다. 호스트와 네트워크 기반의 감사 데이터로부터 BF-XML에 의해 정상과 비정상 행위를 기술하고, SVG(Scalable Vector Graphics)[8]로 변환한다. SVG 변환에 의해서 정상과 비정상 행위에 대한 시각화와 분석이 용이하도록 그래픽 정보를 제공한다.

4.1 BF-XML을 이용한 행위 프로파일링

BF-XML을 이용하여 Shieh 모델의 침입 행위를 프로파일링하면 그림 5와 같다. 실선과 점선은 정의 4에 의해서 각각 직접 관계와 간접 관계를 나타낸다. 또한 I, w, cb, *은 간접(Indirect), 쓰기, 제어, 시간의 전후를 관계를 나타낸다. BF-XML을 이용한 Shieh 모델의 침입 행위를 프로파일링하면 그림 6과 같다.

프로그램 행위 기반을 위한 뉴 멕시코 대학의 Sendmail 데이터와 네트워크 기반의 MIT 링컨 대

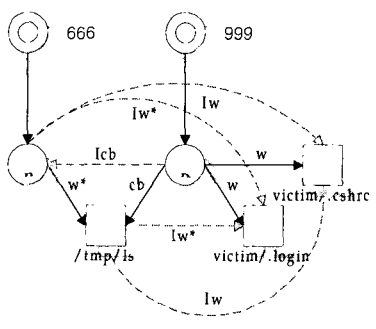

그림 5. Shieh 모델의 프로파일

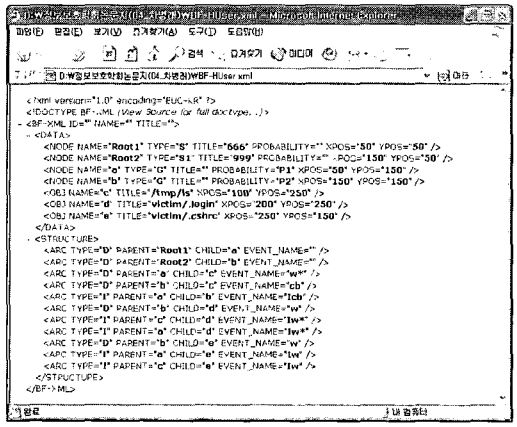

그림 6. BF-XML에 의해 기술된 Shieh 모델의 예제

학의 DARPA 2000년 NT 데이터 이용하여 정상 행위와 비정상 행위 프로파일을 베이지안 네트워크와 BF-XML로 기술한다. 그림 1과 그림 2의 세션을 구성하는 이벤트 정보를 이용하여 BF-XML에 근접한 프로토타이핑을 생성한다.

그림 7은 정상 행위 프로세스인 PID 551의 행위 프로파일링을 베이지안 프레임워크로 나타낸 것이다. 그림 8은 베이지안 프레임워크를 BF-XML을 이용

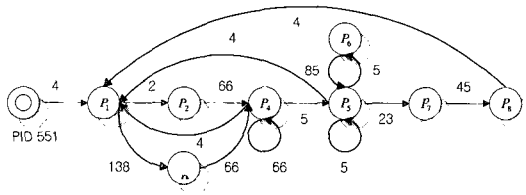

그림 7. BF에 의한 PID 551의 정상행위 프로파일

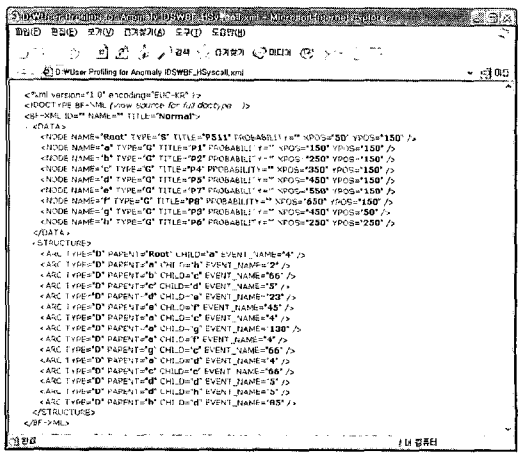

그림 8. BF-XML에 의해 기술된 PID 551

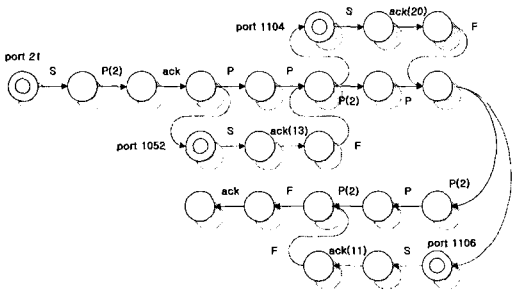

그림 9. BF에 의한 casesen 공격의 네트워크 프로파일

하여 기술한 프로파일을 나타낸다.

그림 9는 FTP 네트워크 서비스에서 공격 행위인 CASESEN 공격의 베이지안 프레임워크에 의한 프로파일링을 나타낸 것이며, 그림 10은 CASESEN 공격을 BF-XML로 기술한 것이다. CASESEN 공격은 NT 시스템의 민감한 객체 디렉토리를 악용하는 U2R(User to Root)공격이다.

4.2 BF-XML의 SVG 변환

BF-XML의 시각화를 위해서는 SVG 파일로 변환한다. BF-XML을 이용한 베이지안 프레임워크로 행위를 프로파일링한다. 베이지안 프레임워크를 BF-XML로 기술하면 XSL에 의해서 SVG 파일로 변환

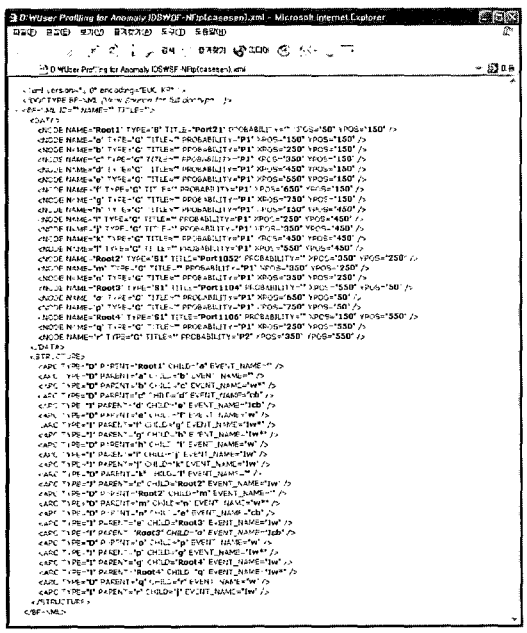

그림 10. BF-XML에 의해 기술된 CASESEN 공격

한다. SVG 파일은 베이지안 네트워크에 의해 기술된 행위 프로파일을 그래픽 형태로 나타내 주므로, XML의 단점인 시각적인 효과를 제공한다.

BF-XML을 이용하여 사용자 행위, 프로그램 행위 그리고 네트워크 행위를 기술하였다. BF-XML로 기술된 행위 프로파일은 프로그램에 의해서 자동으로 SVG로 변환된다. 행위 프로파일의 그래픽 정보인 SVG로 표현하면 그림 11, 그림 12 그리고 그림 13과 같이 나타낸다.

그림 11은 Shieh 모델의 예제의 BF-XML을 SVG로 변환하여 브라우저에 나타낸 것이다. 그림

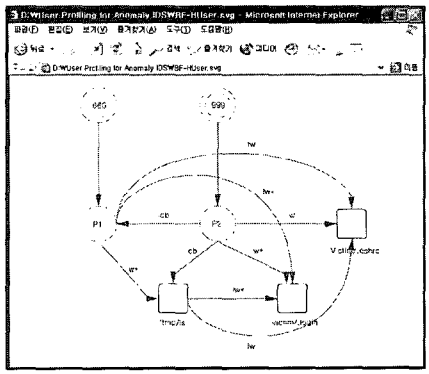

그림 11. Shieh 모델의 SVG 변환

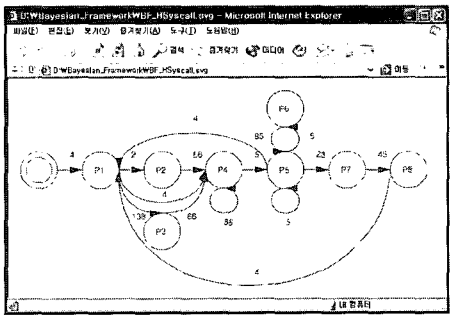

그림 12. P551의 SVG 변환

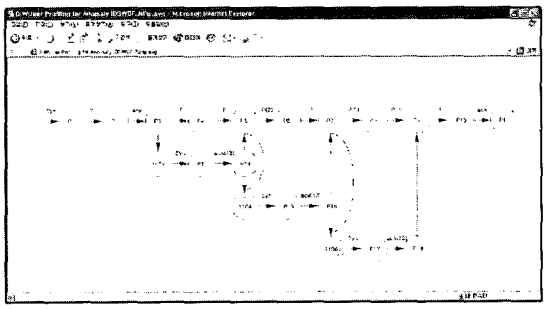

그림 13. CASESEN 공격의 SVG 변환

12는 뉴 멕시코 대학 Sendmail 데이터의 정상 행위 프로파일인 프로세스 아이디 551에 대한 BF-XML을 SVG로 변환한 것이며, 그림 13은 DARPA 2000년 NT 데이터의 FTP 서비스에 포함되어 있는 침입 행위 프로파일인 CASESEN 공격에 대한 BF-XML을 SVG로 변환하여 나타낸 것이다.

4.3 행위 프로파일의 패턴 매칭

BF-XML에 의해서 호스트 기반의 Sendmail 데이터를 MSA 클러스터링 알고리즘으로 분류하면 그림 14와 같으며, 패턴의 일부분만을 나타낸 것이다.

MSA 클러스터링 알고리즘은 DNA서열이나 단백질의 특징이나 구조, 화학적 반응에 결합된 특별한 영역을 탐지하는 방법 중의 하나이다. 그러한 영역에 의해 표현된 정보를 서열 정렬을 이용하여, 새로운 서열 또는 유사한 서열을 데이터베이스에서 검색하는 방법이다[9]. 바이오 인포메틱스에서 사용된 서열 탐색기법들을 이용하여 행위 프로파일의 클러스터링과 패턴 매칭에 적용하였다.

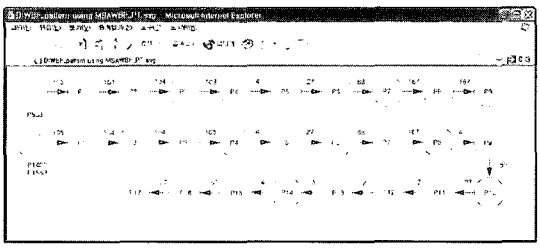

그림 14. 클러스터링에 의해 분류된 세션 패턴

행위 프로파일의 패턴들을 시각화를 위해서는 BF-XML에 의해서 기술된 프로파일 패턴을 필을 이용해서 SVG로 변환하여 표시한다. 또한 BF-XML로 프로파일링된 패턴들을 MSA 클러스터링 알고리즘에 의해서 세션의 클러스터링과 패턴 매칭을 수행하였다. 시스템 호출의 순차로 구성된 세션을 MSA 클러스터링 알고리즘에 의해서 일치된 부분은 노란색으로 표시하며, 불일치된 부분은 흰색으로 표시하였다.

V. 결론 및 향후 연구

이상 침입 탐지를 위해서는 행위에 대한 프로파일이 구축되어야 한다. 구축된 행위 프로파일과 새로운 행위를 비교하여 정상과 비정상을 구분한다. 이상 탐지에 앞서, 어떤 방법으로 행위 프로파일링을 구축하는가가 매우 중요하다.

본 논문에서는 호스트 기반과 네트워크 기반의 이상 탐지에서 행위를 기술하고 분석하기 위한 그래픽 기반의 베이지안 프레임워크를 제안한다. 베이지안 프레임워크는 베이지안 네트워크를 이용한 그래픽 형태로 구축되며, XML을 이용하여 준 구조화된 데이터 형식으로 기술된다. 그리고 시각화를 위해서 SVG로 변환하여 그래픽 형태로 구현한다. 시뮬레이션에서는 호스트 기반의 사용자 행위, 프로그램 행위 그리고 네트워크 행위를 베이지안 프레임워크로 프로파일을 구축하였고, BF-XML과 SVG 변환을 이용하여 그래픽으로 구현하였다. 더불어, MSA 클러스터링 알고리즘에 의해서 세션 패턴을 클러스터링하고 패턴 매칭을 수행하였다.

향후 연구로는 BF-XML로 구축된 정상 행위 프로파일링을 이용한 효율적인 클러스터링 알고리즘 개발이 필요하다. 그리고 변형되거나 새로운 행위에 대해 분류기법과 분석기법에 대한 연구와 오용 탐지에 침입 탐지 기법을 제공하기 위한 자동화된 탐지 패턴 기술 방법에 대해서 연구가 필요하다.

참 고 문 헌

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

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

[3] Dorothy E. Denning, "An Intrusion-Detection Model", IEEE Transaction on Software Engineering, Vol. SE-13, No.2, p222-232, February 1987.

[4] Shiuh-Pyng Shieh and Virgil D. Gligor, "On a Pattern-Oriented Model for Intrusion Detection", IEEE Transaction on knowledge and Data Engineering, Vol. 9, No. 4, July/August, 1997.

[5] Sandeep Kumar and Eugene H. Spafford, "An Application of Pattern Matching in Intrusion Detection", Techni-

cal Report CSD-TR-94-013, June 17, 1994.
[6] http://cs.unm.edu/~immsec/data/synth-sm.html
[7] http://www.ll.mit.edu/IST/ideval/data/ data_index .html
[8] 나방형, 심규찬, 이종연, "XML 그래픽 입문", 21세기 출판사, 2001.
[9] Marco Pagni, "Introduction to Patterns, Profiles and Hidden Markov Models", Swiss Institute of Bioinformatics(SIB), August 30, 2002.
[10] Steven A. Hofmeyr, Stephanie Forrest and Anil Somayaji, "Intrusion Detection using Sequences of System Calls", August 18, 1998.
[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] 유은진, 전문석, 이철희, "페트리네트를 이용한 침입탐지 전자지불 프로토콜의 설계와 검증", 정보보호학회논문지, 2000.
[13] 이종성, 정찬호, 채수환, "특권 프로세서의 시스템 호출 추적을 사용하는 침입탐지시스템의 설계 : 면역 시스템 접근", 정보보호학회논문지, 2000.
[14] 류회재 , 예홍진, "특집 : 네트워크 정보보호 : 침입탐지율 향상을 위한 네트웍 서비스별 클러스터링(clustering)", 정보보호학회지, 2003.

〈著者紹介〉

차 병 래 (Byung-Rae Cha) 정회원
1995년 2월 : 호남대학교 수학과 졸업
1997년 2월 : 호남대학교 컴퓨터공학과 석사
2004년 2월 : 목포대학교 컴퓨터공학과 박사
〈관심분야〉 정보보호, 컴퓨터 네트워크, 신경망