韓國 컴퓨터情報學會 論文誌
第8卷 第1號, 2003. 3.

2003-8-1-3-4

이상 침입 탐지를 위한 베이지안 네트워크 기반의 정상행위 프로파일링

차 병 래* 박 경 우** 서 재 현***

Normal Behavior Profiling based on Bayesian Network for Anomaly Intrusion Detection

Byung-rae Cha* Kyung-woo Park** Jae-hyeon Seo***

요 약

프로그램 행위 침입 탐지 기법은 데몬 프로그램이나 루트 권한으로 실행되는 프로그램이 발생시키는 시스템 호출들을 분석하고 프로파일을 구축하여 침입을 효과적으로 탐지한다. 시스템 호출을 이용한 이상 탐지는 단지 그 프로세스가 이상(anomaly)임을 탐지할 뿐 그 프로세스에 의해 영향을 받는 여러 부분에 대해서는 탐지하지 못하는 문제점을 갖는다. 이러한 문제점을 개선하는 방법이 베이지안 확률값 이용하여 여러 프로세스의 시스템 호출간의 관계를 표현하고, 베이지안 네트워크를 이용한 어플리케이션의 행위 프로파일링에 의해 이상 탐지 정보를 제공한다. 본 논문은 여러 침입 탐지 모델들의 문제점들을 극복하면서 이상 침입 탐지를 효율적으로 수행할 수 있는 베이지안 네트워크를 이용한 침입 탐지 방법을 제안한다. 행위의 전후 관계를 이용한 정상 행위를 간결하게 프로파일링하며, 변형되거나 새로운 행위에 대해서도 탐지가 가능하다. 제안한 정상행위 프로파일링 기법을 UNM 데이터를 이용하여 시뮬레이션하였다.

Abstract

Program Behavior Intrusion Detection Technique analyses system calls that called by daemon program or root authority, constructs profiles, and detectes anomaly intrusions effectively. Anomaly detections using system calls are detected only anomaly processes. But this has a problem that doesn't detect affected various part by anomaly processes. To improve this problem, the relation among system calls of processes is represented by

본 논문은 목포대학교 2001년도 학술연구비에 의해 이루어졌음.
,* 목포대학교 컴퓨터공학과 교수

접수일자 : 2003. 1.10
심사완료 : 2003. 3. 7

bayesian probability values. Application behavior profiling by Bayesian Network supports anomaly intrusion informations. This paper overcomes the problems of various intrusion detection models. we propose effective intrusion detection technique using Bayesian Networks. we have profiled concisely normal behaviors using behavior context. And this method be able to detect new intrusions or modificated intrusions. we had simulation by proposed normal behavior profiling technique using UNM data.

Ⅰ. 서론

인증과 접근 제어 등의 보안 기술은 비인가된 사용자로부터 불법적인 정보의 조작과 접근을 방지하기 위한 컴퓨터 보안의 목적을 달성하기 위해서 개발되어 졌다. 인터넷의 활성화, 가용 정보량의 증가와 정보 보호 위협 요인의 증가로 인하여 이러한 인증과 접근제어만으로는 보안 문제를 해결하기에 충분치 못하였고 정보 보호를 위한 2차 방어선으로 침입 탐지 시스템이 개발되어 졌다.

침입 탐지 시스템은 오용 탐지와 이상 탐지로 분류가 된다. 이상 탐지는 정상 시스템의 행위로부터 주목할만한 특이한 행위 패턴을 침입으로 규정하여 침입을 탐지한다. 반면에 오용 탐지는 알려진 침입 방법들을 수집하여 지식 베이스에 유지하고, 동일한 침입 기술을 지식 베이스 검색을 통해 침입을 탐지하는 방법이다. 이상 탐지를 위한 방법들은 연구 초기단계이며 일반적으로 오용탐지 방법이 많이 상업화되어 사용되지만 알려지지 않은 침입 패턴은 탐지할 수 없는 문제점을 해결하지 못하고 있다[1].

이상탐지를 위해 프로그램의 행위를 분석하는 연구들이 수행되어 진다[1]. 프로그램 행위 기반의 침입 탐지 기법은 데몬 프로그램이나 루트 권한으로 실행되는 프로그램이 발생시키는 시스템 호출들을 분석하고 정상행위 프로파일을 구축하여 잠재적인 공격을 효과적으로 탐지하고 있다.

시스템 호출을 이용한 이상탐지 기법은 열거형 방법, 빈도 기반 방법, 데이터마이닝 접근 방법 그리고 유한상태 기계 방법으로 분류할 수 있다. 열거형 순차 방법은 정상행위를 경험적으로 추적하여 알려지지 않은 패턴을 모니터링하여 이상을 탐지한다. 빈도 기반의 방법은 다양한 이벤트의 빈도 분포를 기준으로 하여 침입을 탐지하며 데이터마이닝 접근 방법은 정상행위 데이터로부터 발생하는 공통의 원소로부터 특징을 추출하여 규칙집합으로 기술함으로서 침입탐지가 가능하도록 한다. 또한 유한 상태 기계 방법은 기계 학습 기법으로 프로그램을 추적하여 인식하는 유한 상태 기계를 구축하여 침입을 탐지하는 방법이다[2].

시스템 호출을 이용한 이상 탐지는 단지 그 프로세스가 이상임을 탐지할 뿐이며 그 프로세스에 영향을 주거나 영향을 받는 다른 여러 부분과의 관련성에 에 대해서는 고려하지 못하는 문제점이 있다. 이러한 문제점을 개선하는 방법으로 베이지안 네트워크를 활용할 수 있으며, 여러 프로세스의 시스템 호출간의 관계를 표현하고, 프로그램 레벨에서 프로세스간의 관계를 고려함으로써 프로세스간의 이상이 미치는 범위까지 이상 탐지 정보를 제공할 수 있다[3,7].

본 논문에서는 시스템 호출 기반의 침입 탐지 모델들의 문제점들을 극복하기 위해 베이지안 네트워크를 적용하였으며 프로그램 레벨에서 이상 침입 탐지를 효율적으로 수행할 수 있도록 하는 프로파일링 기법을 제안하였다.

베이지안 네트워크를 이용한 침입 탐지 모델은 베이지안 이론을 기본으로 전후 관계를 확률값으로 추정하며, 발생하는 이벤트의 전후 관계를 그래프화한 베이지안 네트워크로 표현할 수 있다. 본 논문에서는 행위의 전후 관계를 고려하여 정상 행위를 간결하게 프로파일링하여 이상침입을 탐지하며, 변형되거나 새로운 행위에 대해서도 탐지가 가능하도록 제안하였다. 제안된 기법은 은닉 마코프 모델에 비해서 계산 복잡도가 크지 않다는 장점을 갖는다. 본 논문에서 제안한 방법을 UNM 데이터를 이용하여 시뮬레이션 하였다.

Ⅱ. 관련 연구

2.1 시스템 호출을 이용한 이상탐지 기법

침입 탐지를 위해 프로그램 행위를 분석하여 프로파일을 구축하는 기법들은 사용자 행위 침입 탐지 기법의 대안으로 연구되어 왔다. 프로그램 행위 프로파일은 정상적인 프로그램이 수행되면서 발생시키는 시스템 호출들을 수집 및 분석하여 구축할 수 있다. 시스템 호출을 이용한 이상 탐지 기법들은 다음과 같다.

2.1.1 열거형 순차 방법

열거된 순차에 의존하는 방법들은 lookahead pairs, tide(time-delay embedding) 그리고 stide

(sequence time-delay embedding) 등이 있다. 이 방법들은 정상 행위를 경험적으로 추적하여 알려지지 않은 패턴을 모니터링한다. 초기에 이 기법들은 패턴에 대한 통계적 분석이 적용되지 않았다[2].

2.1.2 빈도 기반의 방법들

빈도 기반의 방법들은 다양한 이벤트의 빈도 분포를 모델로 하며, 텍스트 문서를 분류하는데 사용된 n-그램 벡터(n-gram vector)가 여기에 속한다. 시스템 호출 추적의 방법으로 이 기법은 프로그램이 종료되어야 추적 벡터를 계산할 수 있기 때문에 온라인 테스팅에서는 부적절하다. 또한 벡터의 크기를 결정하는데 어려움이 있으며, 동일한 프로그램의 정상 행위와 비정상 행위를 추적을 위한 충분한 정밀도를 제공하지 못한다[2].

2.1.3 데이터마이닝 접근 방법

데이터마이닝 접근법은 많은 수집된 데이터로부터 가장 중요한 특징을 결정하기 위해 설계되었다. 이상 침입 탐지에서는 발생된 정상 행위의 모든 패턴을 단순하게 나열하여 얻기보다는 간결하게 정의할 수 있는 정상 행위 패턴을 발견하는데 있다. 데이터마이닝 접근법으로 RIPPER는 정상 행위 데이터로부터 발생하는 공통의 원소를 작은 규칙 집합으로 특징을 기술하는 능력을 제공한다[2].

2.1.4 유한 상태 기계 방법

기계 학습 접근법으로 프로그램을 추적하여 인식하기 위하여 유한 상태 기계(Finite State Machines)를 구축하여 이상을 탐지한다. 매우 강력한 유한 상태 기계로는 은닉 마코프 모델이 있으며, 이 모델은 이중 추정 통계적 과정으로 기술된다. 여러 모델중에서 가장 이상탐지 능력이 뛰어난 것으로 판명되었으나, 단지 계산 복잡도가 크다는 단점을 갖고 있다[2].

위에 언급한 이상 탐지 모델의 문제점들로는 (1) 통계적 분석의 필요, (2) 실시간 처리의 어려움, (3) 정상 행위의 간결한 정의, (4) 계산 복잡도 문제 등으로 기술할 수 있다.

본 논문에서는 베이지안 기법를 이용함으로서 통계적 분석과 계산 복잡도의 문제점들을 어느 정도 극복할 수 있다. 더불어, 시스템 호출 정보들를 베이지안 네트워크로 구축함으로써 프로세스간의 관련성과 이상이 미치는 범위까지 파악이 가능하도록 하였다.

2.2 N-gram 기법

프로그램 행위 기반 침입 탐지 기법의 전제는 대부분의 공격은 프로그램 결함이나 버그로 인하여 발생할 수 있으며 프로그램의 정상적인 사용과는 그 행위가 다르다는데 있다. 그러므로, 프로그램의 행위가 적합하게 표현될 수 있다면 침입 탐지를 위한 행위 특성으로 활용될 수 있다.

프로그램의 정상행위를 자동적으로 추출하고 정의하기 위한 대표적인 연구는 뉴 멕시코 대학의 Forrest 연구팀에서 개발한 N-gram 기법이다. 이 기법은 면역학의 개념을 침입탐지에 적용한 사례이다[2,4,5].

N-gram 기법은 프로그램에 의해 발생되는 시스템 호출들을 순차(sequence)적으로 고정 길이로 분할하고 정상행위로 간주하여 프로파일을 구축한다. 만약, 임의의 순차적인 시스템 호출이 프로파일에 존재하지 않는다면 이상행위로 간주한다. 세션내의 총 스트링의 개수에 대해 이상 행위로 간주된 스트링의 개수의 비율이 매우 크다면, 그 세션을 비정상으로 판정한다[2,4,5].

N-gram 기법은 단순한 알고리즘과 높은 탐지율을 보이지만, 프로파일 데이터의 크기 및 오버헤드가 매우 크다는 단점을 갖고 있다.

2.3 침입탐지를 위한 베이지안 방법

베이지안 확률값 계산은 사전 확률과 사후 확률, 우도 함수 등을 이용하여 다음과 같은 식(1)로 제시할 수 있다.

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

사전확률 $P(I)$는 침입의 발생 빈도에 의해 좌우되며, 침입에 대한 완전한 정보를 제공하지 못한다. 그러나 사후 확률 $P(I|E)$는 이벤트 E라는 조건에 의한 그림 1의 (a)와 같이 가장 유력한 침입 $P(I|E)$ 부분에서 집중적인 확률 분포를 보일 것이다[6].

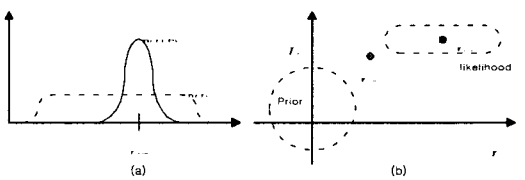

그림 2. 침입에 대한 확률 관계
Fig. 1 Probability relation for Intrusion

사후 확률 $P(I|E)$은 사전 확률 $P(I)$와 우도 함수 $P(E|I)$의 곱에 이벤트의 확률 $P(E)$로 나눔으로써 계산되며, 위의 관계를 그림으로 나타내면 그림 1의 (b)와 같다. 즉, 두 침입 I_i와 I_i에 대해서 사전 정보와 최우도 함수값에 의한 0과 1 사이의 확률값으로 가장 가능한 침입과 이벤트의 정보를 획득할 수 있다[6]. 가장 가능한 침입 I_j 는 확률값 $P(I_j|E)$가 1에 근사하므로 가장 유력한 침입이 되며, 침입 I_i는 확률값 $P(I_i|E)$가 0에 근사하므로 침입과는 무관하게 된다.

Ⅲ. 베이지안 네트워크를 적용한 이상 침입 탐지

시스템 호출의 이상을 탐지하기 위해서 사전정보와 정상 행위 패턴 정보들을 프로파일로 구축하며, 베이지안 네트워크를 적용하여 이상을 탐지하고자 한다.

시스템의 이벤트들을 모니터링 하므로 써 시스템 호출들이 발생하는 사전정보의 확률값을 얻을 수 있으며 그림 2와 같이 사전 정보로써 평균, 표준편차 그리고 순차적인 시스템 호출들을 얻을 수 있다.

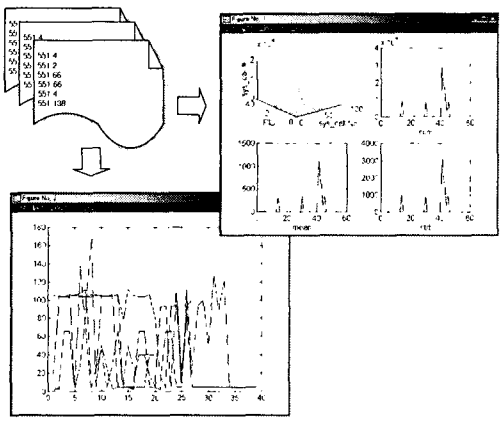

그림 3. 확률, 평균, 편차, 시스템호출 순차
Fig. 2 Probability, Mean, Deviation and System Call Sequence

시스템 호출 과정은 대부분이 무작위의 순서로 이루어지지 않으며, 시스템 호출의 각 상태는 전후의 순서관계가 존재하며 일정한 순서를 갖고 있음을 알 수 있다. 이러한 시스템 호출 과정은 발생하는 순서 관계를 고려하여 상태 전이 그래프인 베이지안 네트워크와 확률값으로 표현 할 수 있다.

본 논문에서는 다음과 같은 제약사항을 전제로 연구를 진행하였다.

제약1) 각 시스템 호출 과정이 전후의 순서 관계에 의한 시간적 순차 과정으로 이루어짐을 가정한다.

제약2) 시스템 호출 과정을 DAG(Direct Arc Graph)를 이용해서 베이지안 네트워크를 표현한다. DAG는 초기상태(◎), 방향성 아크(→), 시스템 호출(이벤트)의 집합(E), 상태(○) 그리고 상태의 확률(P)로 구성된다.

시스템 호출의 연속적인 이벤트($E_1, \cdots, E_{i-1}, E_i$)에 대해서 시스템 상태의 침입 확률($P(I|E_1, \cdots, E_i)$)은 결합 확률 함수를 이용하여 다음과 같이 식(2)로 바꿔 쓸 수 있으며,

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

위의 식(2)로부터 다음을 정의한다.

정의 1) 연속적인 이벤트 $E = (E_1, \cdots, E_{i-1}, E_i)$에 대한 침입 확률값 계산은 $P(I|E_1, \cdots, E_{i-1}, E_i)$으로 그림 3의 (a)와 같이 정의한다. 그리고 각 상태는 전 단계에 독립적임을 가정한다. 전 단계와 독립임을 가정함으로써 계산 복잡도 문제를 해결할 수 있다.

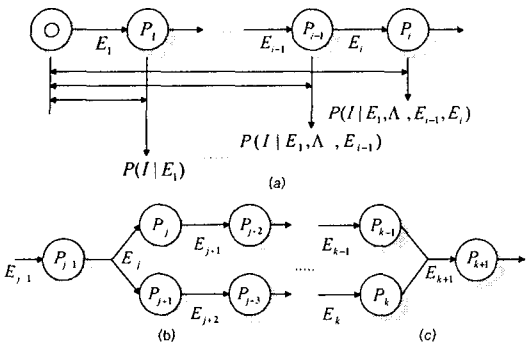

그림 4. 상태전이의 연속, 분기, 병합 처리
Fig. 3 Contiuous, Branch and Merge process of State transition

정의 2) P_{j-1}상태에서 분기시 침입확률값 계산은 $P_j = P(I|E_j, E_{j-1}, \cdots)$와 $P_{j+1} = P(I|E_j, E_{j-1}, \cdots)$이고, P_j와 P_{j+1}의 침입 확률값은 동일하다고 그림 3의 (b)와 같이 정의한다.

정의 3) P_{k-1}과 P_k의 상태에서 병합시 침입 확률값 계산은 $P_{k-1} = P(I|E_{k-1})$와 $P_k = P(I|E_k)$의 결합 확률함수로써, $P_{k+1} = P(I|P_{k-1}, P_k)$으로 그림 3의 (c)와 같이 정의한다.

정의 4) 시스템 호출 과정의 각 상태를 연결, 역, 접두사, 접미사, 길이 그리고 반복으로 표현이 가능하다.

4-1) 연결(concatenation)은 두 개의 상태 P_v와 P_w를 연결하는 것은 P_v의 상태 뒤에 P_w의 상태를 붙이는 연산으로 정의한다.

4-2) 역(reverse)은 어떤 상태의 역순은 주어진 목적들을 거꾸로 나열한 것으로 정의한다. 상태 P_v와 P_w에 대해 $(P_v P_w)^R = P_w^R P_v^R$이 성립한다.

4-3) 접두사(prefix)와 접미사(surfix)는 만약 $P_z = P_v P_w$라면 P_v는 P_w의 접두사가 되고 P_w는 P_t의 접미사라고 정의한다. 어떤 상태에서 접두사나 접미사를 제거함으로써 이루어지는 상태를 '서브 상태'이라 한다.

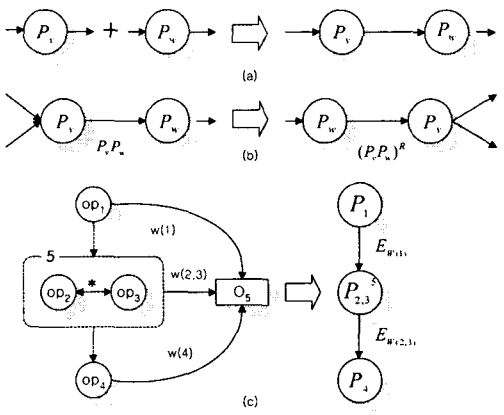

그림 5. 상태 전이의 연결, 역, 반복처리
Fig. 4 Connect, Inverse and Repeat process of State transition

4-4) 상태의 길이는 상태 전이 전과정에 포함된 단위 상태의 개수로 정의한다. $|P_v|$와 같이 절대값을 써서 나타낸다.

4-5) 반복은 P_v가 상태일 때 P_v^n이란 P_v를 n번 연결한 것으로 정의한다.

본 논문에서는 Sendmail 데이터의 다양하고 복잡한 데몬과 프로세스들이 사용되므로, 베이지안 네트워크의 구축 사례로 뉴 멕시코 대학(Univ. of New Mexico)의 Sendmail 데이터를 활용하였다. 먼저 정의 1)을 이용하여 프로세스 아이디(PID : Process IDentification)로 사용하여 시스템 호출 과정을 기본 베이지안 네트워크를 구성하였다. Sendmail의 프로세스 아이디 551의 시스템 호출들은 {4, 2, 66, 66, 4, 138, 66, 5, 23, 45, 4, 27, 66, 5, 4, 2, 66, 66, 5, 4, 2, 66, 66, 5, 5, 85, 5, 5}으로 구성된다. 프로세스 아이디 551의 기본 베이지안 네트워크를 그래프로 표현하면 그림 5와 같다.

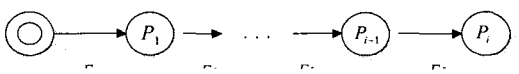

그림 6. 프로세스 아이디 551의 기본 베이지안 네트워크
Fig. 5 Basic Bayesian nework of PID 551

정의2)와 정의 3)을 이용하여 정의 1)에 의해 구성된 기본 베이지안 네트워크를 프로세스 아이디와 무관하게 시스템 호출의 순차를 비교하여 같은 부류끼리 분류한다. 분류된 시스템 호출의 한 부류를 이용하여 하나의 확장된

베이지안 네트워크를 구축한다. 확장된 베이지안 네트워크는 임의의 한 어플리케이션이거나 어플리케이션의 한 부분의 행위에 해당하며, 이것을 프로파일로 사용한다. 다음은 기본 베이지안 네트워크를 확장하는 과정을 보여준다.

그림 6의 (a)는 시스템 호출이 fork인 경우에 프로세스를 복사하여 생성하므로, 유사한 시스템 호출이 발생할 것이다. 그림 6의 (b)는 유사한 두 프로세스간의 매칭을 통한 확장된 베이지안 네트워크를 나타내고 있다. 이와 같은 과정은 정의 2)의 분기로 표현한다.

확장된 베이지안 네트워크를 구축함으로써, 각각의 시스템 호출의 매칭에 의한 이상 탐지에서 탈피하여, 시스템 호출간의 관련성에 의해서 어플리케이션의 행위의 유사도를 이용한 이상 탐지가 가능하게 된다.

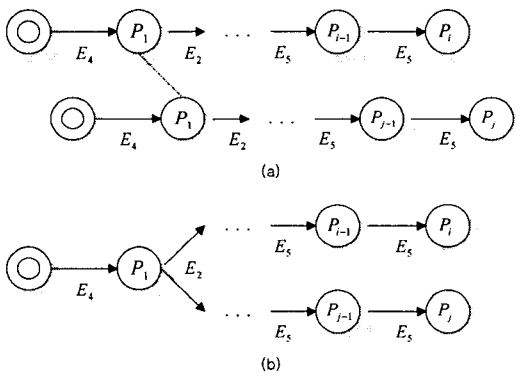

그림 7. 프로세스 아이디 551과 1401 기본 베이지안 네트워크의 확장
Fig. 6 Expanded Basic Bayesian network of PID 551 & 1401

N-gram 방식의 이상 탐지는 정상행위 패턴에 등록되지 않은 패턴이 임계치 이상 발생하면 이상으로 탐지하지만, 제안한 방법은 어플리케이션 행위를 프로파일링하여 이상을 탐지한다. 그러므로, 단순한 매칭에 의한 이상탐지보다는 유연성을 갖으면서 오판율(false positive)이 낮게 나타난다.

Ⅳ. 정상행위 패턴 프로파일링

베이지안 네트워크는 그래픽 형태에서 임의의 시스템 호출간의 인과 종속성(causal dependency)을 표현하고 이웃 노드에 관계되는 확률 집합을 명시한다.

시스템 호출을 이용하여 베이지안 네트워크를 구축하기 위해서는 3단계의 과정이 필요하다.

첫 번째 단계는 프로세스 아이디에 해당하는 여러개의 각각의 기본 베이지안 네트워크를 구축한다.

그림 7의 (a), (b), (c)는 프로세스 아이디가 1409, 1411 그리고 1578의 시스템 호출을 표현한 것이며, (d)는 프로세스 아이디별로 유사도를 표현한 것이다. 시스템 호출을 검사해보면 대부분이 시작과 종료 부분이 같은 순차를 갖는 것을 발견하게 된다. 그러므로, 유사도에 의해 같은 부류로 분류할 수 있다.

두 번째 단계는 유사도에 의해 같은 부류로 분류된 여러 시스템 호출을 이용하여 그림 8과 같이 확장된 베이지안 네트워크를 구축한다. 시스템 호출이 fork인 경우에 정의 2)와 3)에 의해서 기본 베이지안 네트워크를 결합시켜서 확장된 베이지안 네트워크를 구축한다. 그림 7의 (a),(b) 그리고 (c)가 기본 베이지안 네트워크에 해당하고, (d)가 확장된 베이지안 네트워크가 된다.

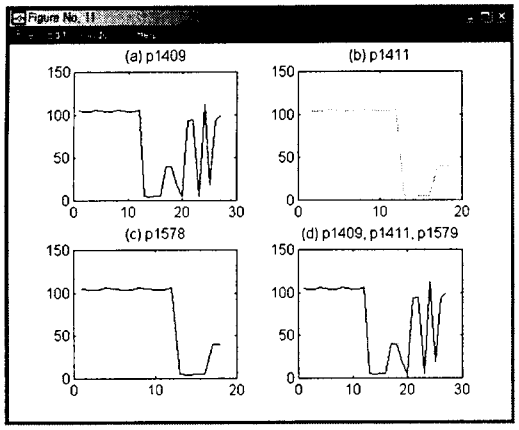

그림 8. 프로세스별 시스템 호출 번호와 유사도
Fig. 7 System call number and Similarity for each process

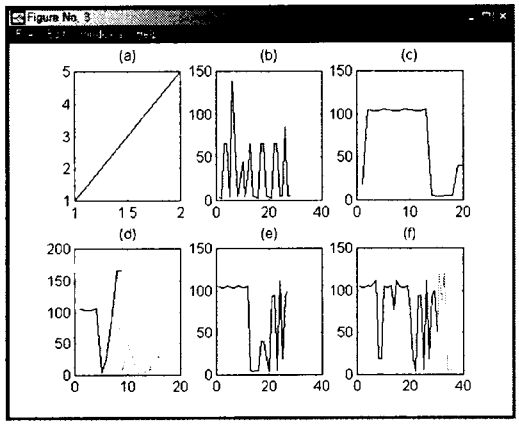

그림 9. 클러스터링에 의한 베이지안 네트워크 구축
Fig. 8 Bayesian network construction by Clustering

세 번째 단계는 확장된 베이지안 네트워크를 최적화된 베이지안 네트워크로 표현하기 위해 재구성한다. 정의 4)를 이용하여 확장된 베이지안 네트워크를 최적화시켜서 재구성한다. 즉, 반복된 부분을 제거하고, 모든 가능한 순차를 베이지안 네트워크로 표현하게 된다. 이 과정에 의해서 시스템 호출에 의한 이상 행위의 변형된 행위나 새로운 행위의 탐지가 가능하게 된다.

프로세스 아이디 551의 경우에는 몇 개의 그룹화된 시스템 호출이 반복되는 것을 발견할 수 있다. 시스템 호출을 단순히 순차적으로 열거하기 보다는 반복부분을 구분하여 제거하는 등의 과정으로 베이지안 네트워크의 최적화를 수행한다. 그림 9은 최적화된 베이지안 네트워크를 나타낸다.

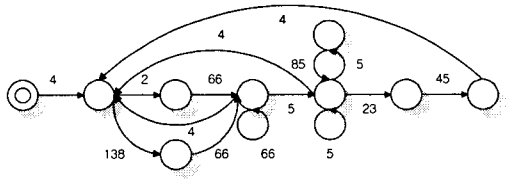

그림 10. 최적화된 베이지안 네트워크
Fig. 9 Optimized Bayesian network

시스템 호출에 의해 생성된 기본 베이지안 네트워크를 분기, 병합 그리고 반복된 부분의 제거 등의 과정을 수행함으로써 프로그램 행위와 유사한 정상 행위 패턴이 생성된다. 이러한 베이지안 네트워크로 표현된 정상 행위 패턴을 정상 행위 프로파일로 구축한다.

시스템 호출에서는 단지 프로세스 아이디에 의한 이상을 갖는 하나의 프로세스밖에는 탐지하지 못한다. 그러나, 정작 하나의 프로그램은 여러 개의 프로세스에 의해 구성되기 때문에 하나의 이상 프로세스에 관련된 프로세스들은 탐지하지 못한다. 즉, 시스템에 이상이 발생했다는 것 밖에는 탐지하지 못한다. 프로그램을 구성하는 임의의 프로세스가 이상 현상을 보이면 결국에는 그 프로세스로 이루어진 프로그램에 이상이 탐지되는 것이다. 한 프로세스 아이디의 이상 탐지에 멈추지 않고 프로세스 아이디의 조합에 의해서 프로그램 레벨의 이상 탐지가 가능하게된다. 또한, 임의의 변형된 시스템 호출에 대해서도 유사도를 측정함으로써 정상과 이상의 탐지가 가능하여 오판율을 낮출 수 있다.

V. 시뮬레이션

제안된 정상행위 패턴 프로파일링을 위해, 뉴 멕시코 대학의 Sendmail 시스템 호출 데이터를 이용하였다.

뉴 멕시코 대학의 Sendmail 데이터는 다양한 데몬과 시스템 호출이 사용되었다. Sendmail 데이터는 많은 양의 정상 행위 데이터가 존재하는 반면에, 비정상 데이터는 상대적으로 협소하게 구성되어 있다. 그러므로, 본 논문에서는 정상 행위 데이터에 대해 베이지안 네트워크를 이용하여 정상 행위 프로파일을 작성하고, 비정상 행위

데이터로는 단지 정상과 이상을 분류하기 위한 베이지안 확률값만 산출하여 사용한다.

시뮬레이션 과정은 먼저, 정상 행위 데이터에 대해 베이지안 네트워크를 이용한 프로파일을 구축한다. 이어서, 비정상 행위 데이터에 대해 베이지안 확률값을 이용하여 이상 침입을 탐지한다.

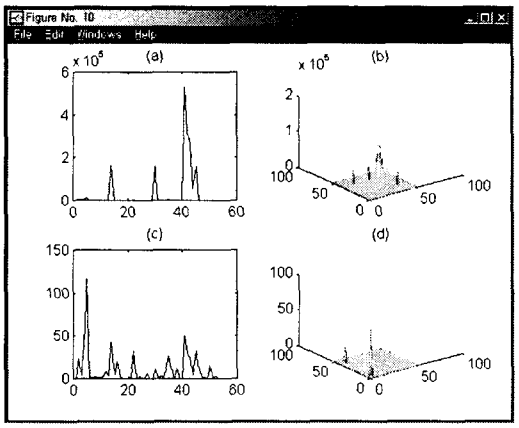

그림 11. 정상 및 이상행위의 분포와 확률값
Fig. 10 Distributions and Probability values of normal & anomaly behavior

그림 10의 (a)와 (b)는 정상 행위의 데이터를 이용한 시스템 호출의 분포와 베이지안 확률값의 분포이며, 그림 10의 (c)와 (d)는 이상 행위 데이터를 이용하여 그래프로 표현한 것이다.

정상 행위의 시스템 호출 분포와 이상 행위 시스템 호출의 분포를 비교할 수 있는데, 정상 행위는 몇몇 시스템 호출에 밀집된 형태를 취하는 반면에, 이상 행위 시스템 호출은 비교적 넓게 분포한 것을 보여준다. 이런 점을 착안하여 시스템 호출의 이상탐지에 베이지안 확률값의 분포에 의한 특징 선택이 가능하게 된다. 그러므로, 모든 시스템 호출을 이용한 이상 탐지보다는 몇몇 시스템 호출로 특징을 선택함으로써 많은 오버헤드 감소와 빠른 이상 탐지가 가능하게 되었다.

그림 11과 그림 12의 시뮬레이션 결과에 의해 정상행위의 베이지안 확률분포는 몇몇 시스템 호출간의 매우 밀접한 전후관계가 존재함을 파악할 수 있다. 이상행위의 베이지안 확률분포 역시, 넓은 분포를 보이지만 서로 연관 시스템 호출간의 전후관계가 존재함을 파악할 수 있다.

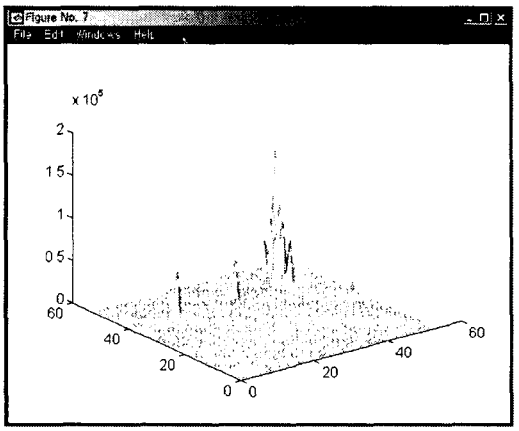

그림 12. 정상행위의 베이지안 확률 분포
Fig. 11 Bayesian probabiity distribution of normal behavior

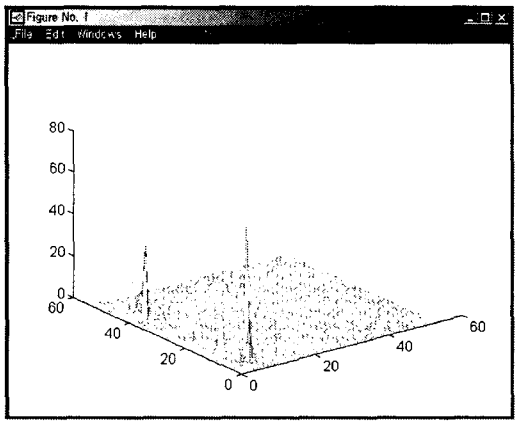

그림 13. 이상 행위의 베이지안 확률 분포
Fig. 12 Bayesian probability distribution of anomaly behavior

몇몇 시스템 호출을 제외한 나머지의 많은 시스템 호출들은 빈번하게 나타나지 않으며, 같은 이벤트를 반복하더라도 확장된 베이지안 네트워크에 의해서 베이지안 확률값의 변화는 매우 미비하게 나타나게 되므로 같은 이벤트가 반복되거나 임의의 시스템 호출에 의한 간섭이 주어지더라도 변형된 침입을 같은 부류로 분류가 가능하였다.

Ⅵ. 결론 및 추후 연구방향

프로그램 행위 침입 탐지 기법은 데몬 프로그램이나 루트 권한으로 실행되는 프로그램이 발생시키는 시스템 호출들을 분석하고 프로파일을 구축하여 침입을 효과적으로 탐지한다.

시스템 호출을 이용한 이상 탐지는 단지 그 프로세스가 이상(anomaly)임을 탐지할 뿐 그 프로세스에 의해 영향을 받는 여러 부분에 대해서는 탐지하지 못하는 문제점을 갖는다. 이러한 문제점을 개선하는 방법이 베이지안 확률값 이용하여 여러 프로세스의 시스템 호출간의 관계를 표현하고, 베이지안 네트워크를 이용한 어플리케이션의 행위 프로파일링에 의해 이상 탐지 정보를 제공한다.

본 논문에서는 시스템의 호출을 이용하는 이상 침입 탐지에 베이지안 네트워크를 적용했으며 베이지안 방법을 이용하여 침입의 각 상태 확률값을 계산하여 정상과 이상을 탐지하였다. 또한 베이지안 네트워크를 확장하여 단순한 시스템 호출의 매칭에 의해 비교하는 이상 탐지에서 탈피하여 어플리케이션의 행위 프로파일링에 의한 이상 탐지가 가능하도록 하였다. N-gram 방식의 이상 탐지는 정상행위 패턴에 등록되지 않은 패턴이 임계치 이상 발생하면 이상으로 탐지하지만, 제안한 방법은 어플리케이션 행위를 베이지안 네트워크로 프로파일링하여 이상을 탐지한다. 그러므로, 단순한 매칭에 의한 이상탐지보다는 유연성을 갖으면서 오판율(false positive)이 낮게 나타난다. 즉, 이상 침입 패턴에 같은 이벤트를 반복하더라도 베이지안 확률값의 변화는 매우 미비하게 나타나게 되고, 다른 시스템 호출의 간섭에 의한 변형된 침입도 확률값의 변화에 미비하게 영향을 미치므로, 같은 부류로 분류가 되었다. 즉, 침입 패턴에 대한 확률값의 요약정보를 생성하여, 프로파일에 정의되지 않은 같은 부류의 침입 패턴을 분류하거나 변형된 침입 패턴을 구분한다.

추후 연구방향으로는 침입 패턴 분류에 대한 프레임워크 연구와 베이지안 확률값에 의한 이상 침입 패턴을 평가하는 기준을 제시하고, 변형된 이상 침입 패턴을 탐지하는 방법을 연구하고자 한다.

참고문헌

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

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

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

[4] S. Forrest, S. Hofmeyr, A. Somayaji ad T. Longstaff, "A sense of self for unix processes", IEEE Symposium on Security and Privacy, p120-p128, 1996.

[5] S. A. Hofmeyr, A. Somayaji and S. Forrest, "Intrusion Detection using Sequences of System Calls", Journal of Computer Security, Vol.6, p151-p180, 1998.

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

[7] K. Jain, R. Sekar, "User-Level Infrastructure for System Call Interposition : A Platform for Intrusion Detection and Confinement", 1999.

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

[9] Paolo Garbolino, Franco Taroni, "Evaluation of scientific evidence using Bayesian networks", Forensic Science International 125, p.149-155, 2002.

[10] E. Biermann, E. Cloete, L.M. Venter, "A

comparison of Intrusion Detection systems", Computers & Security, 20, p676-p683, 2001.

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

[12] Terran Lane, Carla E. Brodley, "Temporal Sequence Learning and Data Reduction for Anomaly Detection", 1999.

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

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

저 자 소 개

차 병 래

2002.9월 : 목포대학교 컴퓨터공학 박사수료

관심분야 : 네트워크 보안, 네트워크, 신경망 등

박 경 우

2001.4월~현재 : 목포대학교 컴퓨터공학과 부교수

관심분야 : 분산시스템, 시스템 소프트웨어 등

서 재 현

2002.10월~현재 : 목포대학교 컴퓨터공학과 부교수

관심분야 : 네트워크 보안, 컴퓨터 네트워크 등