침입탐지 시스템을 위한 은닉 마르코프 모델의 적용

(Application of Hidden Markov Model to Intrusion Detection System)

최 종 호 † 조 성 배 ††
(Jongho Choy) (Sung-Bae Cho)

요 약 정보통신 구조의 확산과 함께 전산시스템에 대한 침입과 피해가 증가되고 있으며 침입탐지 시스템에 대한 관심과 연구가 늘어나고 있다. 본 논문에서는 은닉 마르코프 모델(HMM)을 이용하여 사용자의 정상행위에서 생성된 이벤트ID 정보를 모델링한 후 사용자의 비정상행위를 탐지하는 침입탐지 시스템을 제안한다. 전처리를 거친 이벤트ID열은 전방향-역방향 절차와 Baum-Welch 재추정식을 이용해서 정상행위로 구축된다. 판정은 전방향 절차를 이용해서 판정하려는 열이 정상행위로부터 생성되었을 확률을 계산하며, 이 값을 임계값과 비교함으로써 수행된다. 실험을 통해 침입탐지를 위한 최적의 HMM 매개변수를 결정하고, 사용자 구분이 없는 단일모델링, 사용자별 모델링, 사용자 그룹별 모델링 방식을 비교하여 정상행위 모델링 성능을 평가하였다. 실험결과 제안한 시스템이 발생한 침입을 적절히 탐지함을 확인할 수 있었지만, 신뢰도 높은 침입탐지 시스템의 구축을 위해서는 보다 정교한 모델의 클러스터링이 필요함을 알 수 있었다.

Abstract As the infrastructure for computer communication grows the damage of computer system due to intrusions significantly increases, which has raised interests in the research on intrusion detection systems. This paper proposes an intrusion detection system which models BSM event id sequences generated by users' normal behaviors using hidden Markov model (HMM) and detects anomalous behaviors. Preprocessed event id sequence is modeled as normal behavior using forward-backward procedure and Baum-Welch reestimation formula. Whether user's current behavior is normal or not is determined by comparing the probability at which the sequence is generated from the normal model using forward procedure with pre-determined threshold. Experiments have been made to determine the optimal HMM parameters. HMM models for normal behaviors have been tested and evaluated in three categories: single model without user discrimination, separate model for each user, and models for user groups. Experimental results indicate that the proposed model is promising, and more sophisticated model clustering is required to raise the reliability of the intrusion detection system.

1. 서 론

최근 급속한 정보통신 기반구조의 확산에 힘입어 컴퓨터를 이용한 정보처리에 새로운 장이 열리고 있으며, 그와 더불어 정보보안에 대한 문제가 심각하게 대두되고 있다. 실제로 미국 퍼듀대학의 보고에 의하면 지난 5년동안 정보에 대한 공격 위협이 250% 증가하여 그 피해액이 1천억 불에 달한다고 하며[1], 국내의 경우에도 정보보호센터의 조사에 의하면 네트워크를 통한 침입의 회수가 90년대 들어 급격히 늘어나는 추세이다[2]. 특히 최근 금융망이나 국방망, 전력망 등에 침입하는 사례가 늘고 있어, 불법적인 침입을 사전에 탐지하여 국가의 중요 정보통신 기반구조에 가해지는 피해를 차단할 필요가 있다.

침입탐지 시스템은 불법적인 사용이나 오용, 남용 등에 의한 침입을 알아내는 것으로[3][4], 단일 컴퓨터는

• 본 연구는 정보통신산업기술단의 산업기술개발사업(탐지오인율이 낮은 실시간 Anomaly IDS 개발)의 지원에 의해 이루어졌습니다.
† 비 회 원 : 연세대학교 컴퓨터과학과
hosoft@candy.yonsei.ac.kr
†† 종신회원 : 연세대학교 컴퓨터과학과 교수
sbcho@csai.yonsei.ac.kr
논문접수 : 1999년 9월 17일
심사완료 : 2001년 4월 30일

물론이고 네트워크로 연결된 여러 컴퓨터를 감독할 수 있다. 이러한 시스템은 기본적으로 감사기록, 시스템 테이블, 네트워크 부하기록 등의 자료로부터 사용자의 행위에 대한 정보를 분석하는 작업을 한다. 이제까지 연구·개발된 침입탐지 방법은 공격행위의 정보를 이용하는지 정상행위의 정보를 이용하는지에 따라서 오용탐지와 비정상행위 탐지로 나눌 수 있다.

오용탐지 기법은 알려진 공격에 대한 정보를 이용하여 사용자나 시스템 또는 프로그램의 현재 행동이 공격패턴과 일치하는지를 검사한다. 공격패턴 정보를 가지고 있으므로 정상행위를 공격행위로 간주하는 오류(false-positive error)가 낮고, 공격패턴만 검색하면 되므로 경제적인 장점이 있는 반면, 공격에 대한 정보를 계속 수집하는데 어려움이 있고 알려지지 않은 새로운 공격은 탐지할 수 없다는 단점을 가지고 있다. 이에 비해 비정상행위탐지 기법은 모델링된 정상행위에서 벗어나는 행동은 공격행위로 간주하기 때문에 공격행위를 정상행위로 간주하는 오류(false-negative error)가 적다. 그러나 정상행위 모델링을 위해서 다량의 데이타를 분석해야 하므로 구현비용이 높고, 학습되지 않은 정상행위는 비정상행위로 간주되므로 정상행위가 공격행위로 간주되는 오류가 많다.

본 논문에서는 비정상행위탐지 방식을 사용한 침입탐지 시스템을 구현한다. 비정상행위탐지를 위해서는 참조되는 정상행위모델과 모델링 기법, 그리고 현재 행위가 정상에서 벗어났는지 여부를 알아내기 위한 추론기법이 있어야 한다. 정상모델구축을 위해 모델링하는 대상은 사용자와 프로그램으로 나눌 수 있다. 사용자 모델링[5]은 해당 사용자의 정상적인 사용패턴을 모델링하며 프로그램 모델링[6][7]은 프로그램이 정상적으로 사용되기 위해서 발생되어야 하는 이벤트 및 정보의 변화를 모델링한다. 프로그램을 모델링하는 경우에는 프로그램 문맥에 따른 정상행위 집합을 기술할 수 있다는 장점이 있는 반면, 다른 사용자에 의한 불법적인 계정획득에 의한 침입은 발견할 수 없다. 사용자를 모델링하는 경우에는 다른 사용자에 의한 불법적인 사용을 잡아낼 수 있지만 사용자수에 비례해서 모델링 비용이 증가한다.

본 시스템은 정상행위 모델링과 비정상행위 판정을 위한 기법으로 HMM을 사용한다. 침입탐지시스템은 일반적으로 사용자의 의도가 무엇인지, 어떤 기저에 의해서 정상행위가 이루어지는지 알지 못한다. 침입탐지시스템이 관찰할 수 있는 것은 오직 사용자 또는 프로그램이 생성해내는 이벤트열이며, 이에 의존해서 정상행위 모델을 구축해야 하는 어려움을 가지고 있다. HMM은 이러한 제약조건에서 효과적인 모델링도구로 사용될 수 있다. HMM은 기저가 되는 생성모델를 가정하지 않으며 단지 관찰열에만 의존해서 확률적으로 대상을 모델링하여 특정 관찰열이 구축된 모델로부터 생성되었을 확률 및 최적의 상태전이를 구할 수 있다. 이러한 HMM의 특성은 시스템에서 생성되는 관찰결과로부터 정상행위를 구축하고 현재의 행동이 정상행위인지를 평가할 수 있는 자연스러운 도구를 제공한다. 또한, HMM의 모델링과 평가를 위한 절차들은 잘 정립된 수학적 배경을 가지고 있어서, 음성인식을 포함한 여러 분야에서 소스가 알려지지 않은 대상을 모델링하는데 널리 유용성을 인정받고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 침입탐지시스템의 관련연구들을 살펴보며, 3장에서는 본 논문에서 제안한 시스템의 전체구조, 정상행위 모델링 및 판정기법에 대해 설명한다. 4장에서는 실험을 통해 제안한 시스템의 유용성을 살펴보았으며, 5장에서 결론 및 향후 연구에 대해 언급한다.

2 관련연구

2.1 침입탐지기법[8]

침입탐지시스템은 방화벽을 통과한 외부침입자 뿐만 아니라 내부 사용자의 불법적인 사용을 탐지하는데 그 목적이 있다. 이러한 침입탐지시스템은 크게 표 1과 같이 여러 가지 기준에 의해서 분류될 수 있다[8]. 본 절에서는 이중 현재 사용되고 있는 침입탐지기법들에 대한 설명과 이들을 이용해서 구현된 시스템들에 대해서 개괄한다.

표 1 침입탐지시스템의 분류

분류기준	분류항목	설 명
탐지기법	행동기반	정상행위에 관한 정보를 이용
	지식기반	공격에 관한 정보 이용
탐지시 대응	능동대응	로그아웃, 서비스 폐쇄 등의 순향적인 조치 취함
	수동대응	단순히 경보만을 생성
감사원 위치	호스트기반	감사자료, 시스템 로그
	네트웍기반	네트웍 패킷 등
사용빈도	연속 모니터링	실시간 연속감시 방식을 사용
	주기적 분석	주기적 수행방식

침입탐지시스템의 탐지기법은 공격에 관한 축적된 지식을 사용하여 특정 공격기법을 사용하고 있다는 증거를 찾는 오용탐지 또는 지식기반기법과 감시중인 시스템의 정상행위에 관한 참조모델을 생성한 후 정상행위에서 벗어나는 경우를 찾는 비정상행위탐지 또는 행동기반 기법으로 나뉠 수 있다. 오용탐지는 침입자의 행위에 대한 시나리오를 결정하여 저장하는데, 시스템에 접속하여 사용하는 행위에서 이러한 시나리오 즉 알려진 침입패턴과 동일한 행위를 침입이라고 판정한다. 따라서, 규칙 기반 침입 확인 방법이 가장 많이 사용되고 상태 변이 방법도 종종 사용된다. 오용탐지에는 전문가시스템, 시그너쳐 분석, 페트리넷, 상태전이분석 등의 기법이 사용되고 있다.

전문가시스템은 주로 오용탐지기법에서 사용되고 있는데 공격에 관한 규칙집합을 가지고 있어 감사 이벤트가 전문가시스템 내에서 의미를 가지는 사실로 변환이 된 후 추론엔진에 의해 규칙과 사실을 기반으로 침입을 판단한다. 이러한 접근방식은 지식공학적인 측면과 처리속도면에서 한계를 가지고 있다. 지식공학 측면의 약점은 공격에 관한 지식을 추출하기가 쉽지 않으며 감사자료를 입력으로 사용해서 생성규칙으로 변환하는 작업이 더욱 어렵다는 점이다. 때때로 필요한 정보를 감사자료에서 얻을 수 없는 경우도 있다. 또한 하나의 취약점을 이용하는 방법은 여러 가지가 있을 수 있으므로 여러 규칙을 생성해야 하는 문제점을 가지고 있다.

시그너쳐 분석은 전문가시스템과 동일한 방식으로 지식을 획득하지만 지식을 사용하는 방식이 다르다. 공격에 대한 의미적 기술은 감사 자료에서 곧바로 검색이 가능한 형태의 정보로 변경된다. 예를 들어, 공격 시나리오는 공격시 생성되는 감사이벤트 시퀀스로 변경되거나 시스템에 의해 생성된 감사자료에서 탐색할 수 있는 데이타 패턴으로 변경된다. 이 때 공격에 관한 기술이 저수준에서 이루어진다. 시그너쳐 분석기법은 아주 효율적인 구현이 가능하므로 상업적인 침입탐지 제품에 응용되고 있다. 이 방식의 주약점은 다른 지식기반 접근방법과 마찬가지로 새로 발견된 취약점에 대해 자주 갱신을 해주어야 한다는 것이다.

침입에 관한 시그너쳐를 표현하기 위해서 IDIOT 시스템[3]은 칼라페트리넷(CPN)을 사용한다. CPN은 일반성, 개념적 단순성, 그래프 표현성 등의 장점을 가지고 있다. 시스템 관리자는 공격의 시그너쳐를 작성하고 IDIOT 시스템에 통합할 수 있다. CPN은 일반적으로 아주 복잡한 시그너쳐도 쉽게 작성할 수 있다. 그러나 복잡한 시그너쳐를 감사자료와 비교하는 작업은 상당히 많은 계산비용을 요구한다.

상태전이분석은 Porras와 Kemmerer[9]에 의해 제안된 기법으로 개념상으로는 모델기반 추론과 동일하다. 이 기법은 공격을 목표와 상태 전이의 집합으로 기술하며 상태전이 다이어그램으로 표현한다.

비정상행위 탐지기법은 시스템이나 사용자의 정상행위로부터 벗어나는 행위를 관찰함으로써 침입을 탐지할 수 있다는 가정에 기반한다. 정상 또는 유효한 행동은 다양한 방법으로 수집된 참조 정보로부터 추출된다. 사용자나 프로세스의 활동은 모델링된 정상행위와 비교되고, 정상행위로부터의 이탈이 관찰되면 경보가 발생된다. 즉, 이전에 학습된 행동에 대응되지 않는 행위는 침입행위로 간주된다.

비정상행위 탐지기법 중 가장 널리 사용되는 도구는 통계적 기법이다[10][11][12]. 통계적 기법은 사용자나 시스템 행동을 시간에 따라 샘플링된 여러 가지 변수들에 의해 측정하여 평균과 표준편차로 모델링한다. 탐지시에 변수의 표준편차에 기반해서 관정데이타가 임계값을 넘어섰는지를 검사한다.

전문가시스템은 정상행위를 규칙의 집합으로 표현한다[13]. 탐지시에는 사용자의 활동을 구축된 규칙들과 비교하는 방식으로 비정상행위 여부를 가린다. 신경망은 학습과 일반화 능력을 이용하여 주로 시스템내의 사용자나 데몬 등의 행위를 학습하는데 사용된다[6]. 통계적 기법에 비해 신경망을 사용하는 경우의 이점은 변수들 간의 비선형적 관계를 표현하는 간단한 방법을 가지는 것과 신경망을 자동적으로 학습하는데 있다.

사용자의 의도 파악[14]은 사용자가 시스템에서 수행하는 고수준 작업들의 집합을 사용해서 사용자의 정상행위를 모델링하는 것이다. 각 작업들은 행동들로 세분화되고, 이 행동들이 시스템에서 관찰된 감사이벤트들과 관계된다. 분석기는 각 사용자가 수행할 수 있는 작업들의 집합을 유지하며 작업 패턴에 맞지 않는 행동이 발생하면 경보를 발생시킨다.

컴퓨터 면역학[15]은 자연계의 면역시스템을 침입탐지에 응용한다. 자기(self)와 비자기(nonself)는 각각 정상행위와 비정상행위에 해당한다. 실제 시스템에서 이 모델은 프로세스에서 만들어진 시스템 호출의 짧은 시퀀스들로 구성된다. 코드내의 결함을 사용하는 공격은 일반적으로 수행되지 않는 경로들을 거친다. 따라서 서비스의 적절한 행위를 나타내는 참조 감사들의 집합을 수집한 후 알려진 모든 시스템 호출의 정상시퀀스를 포함하고 있는 참조 테이블을 추출한다. 침입 탐지는 사용자나 프로세스의 행위 시퀀스가 테이블에 있는지 검사

하여 수행된다.

기존 침입탐지시스템을 위의 기준에 따라 분류하면 표 2와 같다[8].

2.2 순서정보를 이용한 정상행위 모델링

컴퓨터 내의 프로그램 수행이나 사용자 활동은 시간에 따라 순서적으로 발생한다는 점이 자연스럽게 모델링에서 순서적인 정보를 이용하려는 시도들로 이어졌다.

Forrest는 유닉스 프로세스의 시작에서부터 끝까지 발생하는 시스템호출의 순서화된 리스트라고 정의된 트레이스를 통해서 프로그램의 정상행위는 트레이스내의 국부적인 패턴들에 의해서 결정될 수 있으며, 결정된 패턴들로부터 벗어난 정도를 가지고 수행중인 프로세스의 보안위반여부를 결정할 수 있다는 것을 보였다[16]. 정상행위 모델링은 프로그램의 특성정상패턴 데이타베이스를 통해서 이루어진다. 정상적으로 수행된 프로그램의 트레이스를 길이 $k+1$의 윈도우를 사용해 이동하면서 현재 시스템호출을 기준으로 이후 k까지의 각 위치에 나타날 수 있는 시스템호출의 집합을 수집한다. 정상패턴 데이타베이스가 구축된 후에는 검사할 현재 트레이스의 각 시스템호출열이 데이타베이스에 있는지 여부를 검사한다. 이 기법은 시스템호출열이라는 정상행위와 침입행위를 구별할 수 있는 단순한 관찰대상을 발견하였고, 분석 및 모델을 위한 간단하면서 계산적으로 효율적인 기법을 사용했다는 중요한 특징을 가지고 있다. 그러나 합법적인 단위 행동으로 이루어진 불법적인 작업을 수행하는 것을 막을 수 없다는 단점을 가지고 있다.

[6]은 동등정합(equality matching) 기법을 사용하는데, 이벤트들을 고정크기 윈도우로 나눈 후 구축되어있는 정상열들과 비교하면서 비정상행위 계수를 통해 추적한다. 계수기는 각 고정윈도우에서 시스템호출을 계수하는 것에서부터 비정상 윈도우를 계수하는 것까지 다양한 수준에서 사용되며, 각 수준에서는 어느 선에서 상위수준으로 비정상행위를 전파할지를 결정하는 임계값들이 적용된다. 충분한 수의 시스템 호출 윈도우가 비정상적이면 침입탐지 플래그를 설정한다.

Elman 신경망을 이용한 비정상행위 탐지[6]는 열간 상태정보를 유지할 수 있는 신경망 기법으로 신경망의 학습과 일반화 능력을 유지하면서 순서적 특성을 반영하려고 시도한 것이다. Elman 신경망은 일반적인 입력, 출력, 은닉 노드에 추가적으로 문맥 노드를 가지고 입력간의 상태정보를 유지한다.

Lane과 Brodley[3][5]는 유닉스 명령행상에서의 사용자 입력을 주요한 이벤트 정보로 사용하여 사용자 프로파일을 구축했다. 이때 연속적인 이벤트들을 처리하기 위해 고정길이의 크기로 분할하는 방법을, 이벤트들의 연관성 소멸을 방지하기 위해 이벤트를 중복시키는 방법을 사용한다. 유사도 평가기준은 사용자의 프로파일에 저장된 패턴과 입력된 패턴이 일치하는지의 여부를 사용한다.

입력된 패턴과 사용자 프로파일과의 유사도를 측정하기 위해 다음 수식과 같은 간단한 매칭 알고리즘을 사용한다. 이러한 방법은 사용자가 입력한 패턴의 순서가 저장된 사용자 프로파일의 패턴순서와 일치하는지를 검사한다. $Sim(X, Y)$는 시퀀스간의 유사도를 나타낸다. 수식에서 보듯이 일치하던 순서가 중간에 다를 경우 큰 폭의 차이를 나타내도록 한다.

$$w(X,Y,i)=\begin{cases}0 & \text{if } i<0 \text{ or } x_i \neq y_i \\ 1+w(X,Y,i-1) & \text{if } x_i = y_i\end{cases}$$

$$Sim(X,Y)=\sum_{i=0}^{l} w(X,Y,i)$$

3. HMM을 사용한 침입탐지

3.1 침입탐지 시스템 개요

본 논문에서 개발한 침입탐지시스템은 그림 1과 같이 데이타 필터링과 데이타 축약을 담당하는 전처리 모듈과 정상행위 모델링과 추론 및 판정을 담당하는 비정상행위 판정모듈로 구성된다. 모델학습을 통해 정상행위를 프로파일 데이타베이스로 구축하여 판정시 사용한다.

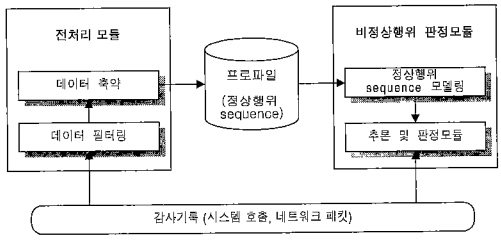

그림 1 침입탐지 시스템 개요

3.2 전처리

이벤트열 정보원으로는 Sun사의 BSM (Basic Security Module)[17]을 통해 획득한 감사자료 중 이벤트ID를 사용하였다. 이벤트ID는 주로 시스템호출로 구성된 커널수준 이벤트ID와 응용프로그램 수행시 발생하는 응용프로그램수준 이벤트ID로 구성되어 있다. 모든 이벤트ID가 다 사용되지 않으므로 통계적으로 빈도가 높은 49개의 이벤트에 대해 0부터 48번까지의 번호를 부여하였고 그 밖의 이벤트는 49번을 부여하여 총 50개의 축약된 이벤트ID를 사용하였다. 표 3은 본 논문에서

표 2 침입탐지시스템 일람
ES : 전문가시스템, SA : 시그너처 분석, PN : 페트리넷, STA : 상태전이분석,
Stat : 통계적기법, NN : 신경망, UII: 사용자 의도, HB : 호스트기반, NB:네트웍기반

기 관	IDS명	기간	지식기반 IDS				행동기반 IDS				HB	NB
			ES	SA	PN	STA	Stat	ES	NN	UII		
Univ. Namur	ASAX	1990-1997	X								X	
AT&T	ComputerWatch	1987-1990						X			X	
USAF	HayStack	1987-1990					X				X	
	DIDS	1989-1995	X				X				X	X
CS Telecom	Hyperview	1990-1995	X				X		X		X	
SRI	IDES	1983-1992	x				X				X	
	NIDES	1992-1995	X				X				X	
	Emerald	1996-		X			X					X
Purdue Univ.	IDIOT	1992-1997			X						X	
UCDavis	NSM	1989-1995		X			X					X
	GrIDS	1995-						X				X
LANL	W&S	1987-1990						X			X	
	Nadir	1990-	X					X			X	
WheelGroup	NetRanger	1995-		X								X
ISS	RealSecure	199-		X								X
Securenet Cons	SecureNet	1992-1996	X				X		X	X	X	
Haystack labs	Stalker	1995-		X				X			X	
	WebStalker	199-									X	
UCSB	STAT	1991-1992				X					X	
	USTAT	1992-1993				X					X	
Stanford Univ.	Swatch	1992-1993	X								X	
MCNC and NCSU	JiNao	1995-	X				X					X

사용한 이벤트ID를 보여준다. 순서적으로 생성되는 이벤트ID는 일정 크기의 윈도우를 옆으로 이동시켜가면서 윈도우 크기 만한 열로 추출하였다.

3.3 침입탐지를 위한 HMM

HMM은 실제적인 생성모델을 알 수 없고 단지 생성된 관찰열에 의해서만 확률적으로 관찰할 수 있는 이중으로 확률적인 절차로서[18], 순서정보를 모델링하기에 유용한 도구이다. 이 모델은 상태라고 불리는 N개의 노드와 상태간의 전이를 표현하는 가지(edge)로 구성된 그래프로 볼 수 있다. 각 상태노드에는 초기상태 분포와 해당 상태에서 M개의 관찰가능한 심볼 중 특정 심볼을 관찰할 확률분포가 저장되어 있으며, 각 가지에는 한 상태에서 다른 상태로 전이할 상태전이 확률분포가 저장되어 있다. 그림 2는 상태수가 3인 우향(left-to-right) HMM을 보여주고 있다.

표 3 축약된 이벤트ID

번호	BSM이벤트	번호	BSM이벤트	번호	BSM이벤트
0	exit	17	mkdir	34	munmap
1	fork	18	rmdir	35	setegid
2	creat	19	setrlimit	36	seteuid
3	link	20	pathconf	37	putmsg
4	unlink	21	open_r	38	getmsg
5	chdir	22	open_w	39	auditon_setcond
6	chmod	23	open_wc	40	statvfs
7	chown	24	open_rw	41	sysinfo
8	kill	25	open_rwc	42	fork1
9	symlink	26	close	43	sockconnect
10	readlink	27	getaudit	44	login
11	execve	28	setaudit	45	logout
12	vfork	29	ioctl	46	telnet
13	setgroups	30	setuid	47	rlogin
14	setpgrp	31	utime	48	su
15	fcntl	32	nice	49	기타
16	rename	33	setgid		

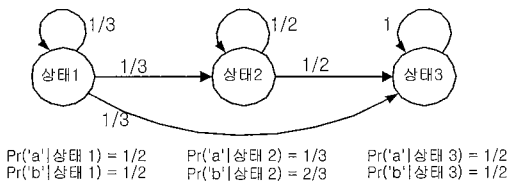

그림 2 우향모델 HMM의 예

$O=O_1, O_2, \ldots, O_T$라는 입력열이 주어지면 HMM은 비록 외부에서 그 상태전이 과정을 직접적으로 알 수는 없어도 자체의 확률 매개변수를 이용하여 이를 마르코프 과정의 확률함수로 모델링할 수 있다. 또한, 일단 모델링 과정을 통해 모델이 구축되면 임의의 입력열이 모델로부터 생성되었을 확률을 계산할 수 있다. HMM은 다음과 같은 매개변수 $\lambda=(A,B,\pi)$로 표현된다.

- 상태전이 확률분포 $A=\{a_{ij}\}$(상태 S_i에서 상태 S_j로 이동할 확률)
- 관찰심볼 확률분포 $B=\{b_j(k)\}$(상태 S_j에서 심볼 v_k를 관찰할 확률)
- 초기 상태 분포 $\pi=\{\pi_i\}$ (초기 상태가 S_i가 될 확률)

본 논문은 시스템에서 발생되는 각 이벤트를 HMM의 관찰심볼로 사용했으며, 시스템의 이벤트 발생기저는 일반적으로 알 수 없으므로 HMM의 상태는 상태수만을 정의하고 각 상태에 대해서 의미를 부여하지는 않았다. 시스템에서 발생되는 이벤트열은 고정길이 T로 샘플링되어 HMM의 입력열이 되며, 정상행위 모델링과 정상행위 평가는 HMM의 관찰열 확률 계산과 모델학습 문제와 대응된다. HMM을 이용한 침입탐지는 그림 3의 과정을 통해서 수행된다.

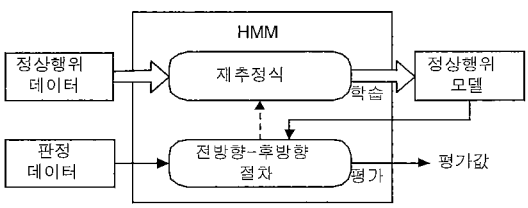

그림 3 HMM을 이용한 정상행위 모델링과 정상행위 평가과정

• 정상행위 모델링

정상행위 모델링은 전처리 단계에서 생성된 정상행위 열을 기반으로 HMM의 매개변수를 결정하는 과정이다. HMM의 매개변수 결정은 주어진 관찰열 O가 해당 모델 λ로부터 나왔을 확률인 $\Pr(O|\lambda)$값이 최대가 되도록 $\lambda=(A,B,\pi)$를 조정한다. 이를 계산하는 해석적인 방법은 알려져있지 않고 반복적으로 λ를 결정하는 방법으로 Baum-Welch의 재추정식이 있다[18].

• 정상행위 평가

정상행위 평가는 이미 구축되어 있는 정상행위별 HMM에 사용자행위열을 입력으로 넣어 각 정상행위에서 현재 행위가 생성되었을 확률을 계산한다. 예를 들어 그림 2와 같이 구축된 모델에 입력열로 '*aba*'가 관찰되었고, 초기 상태는 항상 상태 1에서 시작된다고 하자. 입력열이 상태 1, 상태 2, 상태 3을 거쳐서 발생했을 확률은 초기상태를 상태 1에서 시작하여 상태 1에서 심볼 '*a*'를 관찰한 후, 상태 1에서 상태 2로 전이한 후 상태 2에서 심볼 '*b*'를 관찰하고, 상태 2에서 상태 3으로 전이한 후 상태 3에서 심볼 '*a*'를 관찰할 확률을 곱하면

$$\begin{aligned}\Pr(O &= 'aba', Q=S_1S_2S_3|\lambda)\\ &=\pi_1\cdot b_1('a')\cdot a_{12}\cdot b_2('b')\cdot a_{23}\cdot b_3('a')\\ &=1\cdot 1/2\cdot 1/3\cdot 1/2\cdot 1/2\cdot 1\\ &=1/24\end{aligned}$$

과 같이 계산된다. '*aba*'가 어떤 상태를 거쳐서 발생했는지 모르므로 모든 가능한 상태전이에 대해서 확률을 구한 후 더해주면 해당 모델에서 관찰열 '*aba*'가 발생한 확률을 구할 수 있다. 실제에서는 더 효율적인 전방향-역방향 절차(forward-backward procedure)를 사용할 수 있다[18].

• 비정상행위 판정

정상행위 평가에 의해 수치화된 평가값은 현재 행위가 기준이 되는 모델, 즉 정상행위로부터 생성되었는지를 나타내는 확률값으로 다른 평가값과의 산술적인 직접비교가 가능하다. 따라서 정상행위로 볼 수 있는 가장 낮은 임계값을 결정하고, 현재 행위의 평가값을 임계값과 산술비교하여 더 낮으면 현재 행위를 비정상행위로 판정한다.

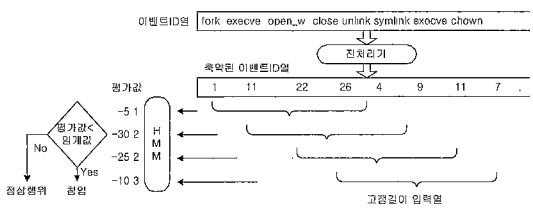

그림 4 비정상행위 판정과정

그림 4는 이벤트ID열이 전처리기에 의해 축약되고 고정길이로 샘플링되어 HMM에 의해 정상행위 평가값이 계산된 후 정상행위 임계값과 비교되어 비정상행위인지 판정하는 과정을 보여준다.

4. 실험결과

실험 데이타로는 3명의 사용자가 1주일간 발생시킨 데이타를 사용하였다. 주사용 프로그램은 문서편집기와 컴파일러, 그리고 사용자가 작성한 프로그램이었다. 학습 데이타와 테스트 데이타는 사용자별로 각각 10,000개, 즉 60,000개를 사용했으며, 테스트 데이타에는 각 사용자별로 일반사용자가 루트 권한을 획득하는 u2r(user2root)형의 공격을 17차례 넣었다. 이 침입은 심볼릭링크를 사용하여 특정 프로그램 수행시에 발생하는 임시파일생성과 경쟁조건에 관련된 취약성을 이용하여 루트 권한을 획득한다. HMM의 모델은 순서정보를 잘 표현하는 것으로 알려진 우향 모델을 사용했다.

각 실험결과는 ROC (Receiver Operating Characteristic) 곡선[6]을 사용하여 비교하였다. 본 논문에서는 임계값의 조정에 따른 탐지율과 false-positive 오류율의 변화를 ROC 곡선을 사용하여 보여준다. 바람직한 침입탐지시스템은 낮은 false-positive 오류에서 높은 침입탐지율을 보여주어야 하므로 곡선이 왼쪽 위로 있을수록 좋은 성능을 나타낸다. 정상행위 평가값은 확률값에 자연로그를 취한 실수값으로, $[0, -\infty]$의 범위를 갖는다.

4.1 단일모델링

첫 번째 실험은 전체 사용자 데이타를 단일 모델로 구축하여 수행하였다. 그림 5는 침입행위가 발생한 경우의 열 평가값의 변화를 보여준다.

그림 5의 세로축은 사용자의 이벤트열 평가값으로 HMM으로 구축된 정상행위모델로부터 현재의 이벤트열이 발생되었을 확률값이다. 컴퓨터 자료표현범위의 제한으로 인하여 실제 확률값을 사용하는 경우 확률값이 매우 낮은 경우 언더플로우가 발생하는 경우가 있으므로 실제 HMM구현시에는 확률값에 로그를 취하기도 한다[18]. 본 실험에서도 확률값에 자연로그를 취한 값을 사용하였다. 수평축은 관심을 가지고 있는 시간범위상에서 이벤트의 순서를 나타낸다. 즉 수평축의 1은 해당 시간범위상의 첫 번째 이벤트이며 31은 31번째 이벤트를 나타낸다. 이때 해당시점에서의 수평축의 값은 현재 이벤트까지를 포함하는 이벤트열의 평가값을 나타낸다. 그림에서 보듯이 침입행위가 시작된 시간 11에서부터 침입행위가 종료된 시간 32사이에 있는 열들의 평가값이 급격히 낮아져 HMM이 효과적으로 침입을 탐지하고 있는 것을 보여주고 있다.

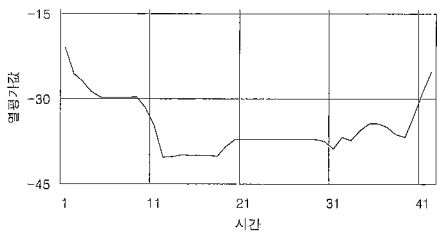

그림 5 침입행위 발생시 평가값의 변화

이 실험에서는 침입탐지에 최적의 성능을 보일 수 있는 HMM의 매개변수를 결정하기 위한 실험도 병행하였다. 상태수를 5, 10, 15, 30으로 변경해가면서 ROC곡선을 얻은 결과 그림 6와 7에서 볼 수 있듯이 상태수 10에서 가장 낮은 false-positive 오류를 보여 가장 좋은 결과를 얻을 수 있었다.

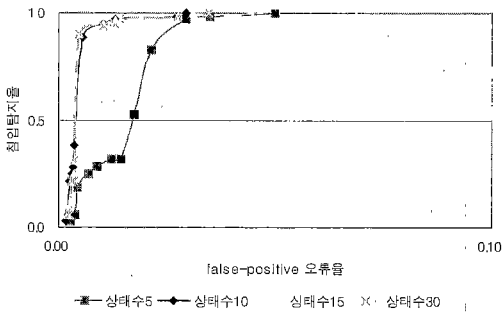

그림 6 상태수에 따른 ROC곡선(단일모델)

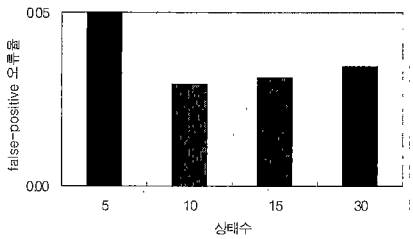

그림 7 상태수에 따른 false-positive오류율(단일모델)

그림 8과 9는 열 길이를 10, 20, 30, 40으로 변경해가면서 시스템의 성능을 평가한 것을 보여주는데, 열 길이 30에서 가장 좋은 성능을 내는 것을 볼 수 있다.

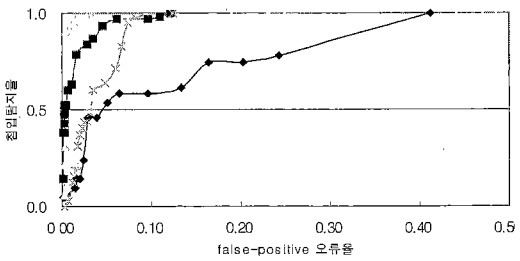

그림 8 열길이에 따른 ROC곡선(단일모델)

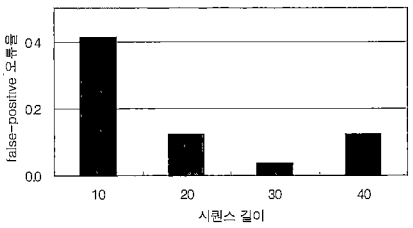

그림 9 열길이에 따른 false-positive오류율(단일모델)

4.2 사용자별 모델링

첫번째 실험에서 침입탐지 문제에 HMM을 적용하는 것이 타당함을 검토하고 최적의 HMM 매개변수를 결정한 후, 사용자별로 별도의 모델을 만들어서 두번째 실험을 수행하였다.

실제 대형 네트웍에서는 일반적으로 내부 네트웍에서의 접근은 신뢰하고 있다. 많은 경우 사용자들은 회의장이나 협력 연구기관이나 외부 사이트에서 ISP나 원격 네트웍을 통해 telnet, ftp, POP 등의 안전하지 않은 응용프로그램으로 네트웍을 접속하는 경우가 있으며 시스템 관리자는 원격 사이트에서 합법적인 원격사용자의 로긴 정보가 유출된 경우를 가장 큰 위협으로 삼고있다 [19]. 이같은 침입이 발생된 경우 침입자는 유출된 일반 사용자 계정을 교두보로 삼아서 시스템의 취약성을 이용해 루트 권한을 획득하고 있는 실정이다. 따라서 현재 사용자가 기존의 사용패턴과 다른 행동을 보이는지를 검사할 수 있는 기능은 위와 같은 위협을 발견하는데 중요한 역할을 한다.

표 4는 각 모델별 탐지율과 false-positive 오류율을 보여준다. 모든 테스트 데이타를 단일 모델로 모델링한 경우보다 각 사용자별로 모델링한 경우에 전반적으로 더 낮은 false-positive 오류율을 보여준다.

표 4 사용자별 모델링에 따른 비정상행위 탐지결과

	단일 모델링	사용자별 모델링		
		사용자 1	사용자 2	사용자 3
정상행위평균 (표준편차)	-14.74 (10.42)	-15.54 (5.88)	-4.72 (8.42)	-14.25 (12.34)
임 계 값	-32.45	-26.21	-31.35	-36.92
탐 지 율	100%	100%	100%	100%
false-positive 오류율	2.95%	4.31%	1.73%	1.96%

그림 10의 ROC 곡선에서도 단일 모델링의 경우보다 세분화된 모델링의 경우가 전반적으로 좋은 성능을 보여줌을 확인할 수 있다.

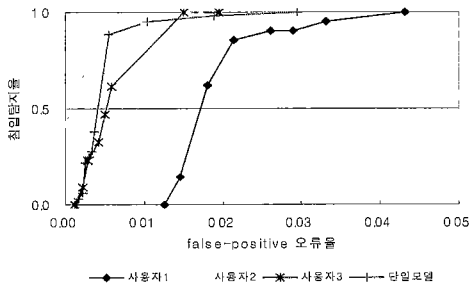

그림 10 ROC 곡선(모델비교)

정상행위를 모델링한 사용자와 판정사용자를 다르게 한 경우 표 5와 같이 열 평가값이 낮게 나왔다. 그림 11, 12, 13에서는 사용자가 달라진 경우 false-positive 오류율이 증가함을 확인할 수 있다. 이는 본 기법이 사용자의 정상행위를 모델링하는데 효과적임을 보여주며 다른 사용자의 불법적인 계정획득에 의한 침입행위 탐지에도 효과적으로 사용될 수 있음을 보여준다.

표 5 열평가값이 -50이하인 열수

		평가 사용자		
		사용자1	사용자2	사용자3
모델링 사용자	사용자1	0	366	55
	사용자2	566	315	336
	사용자3	475	346	0

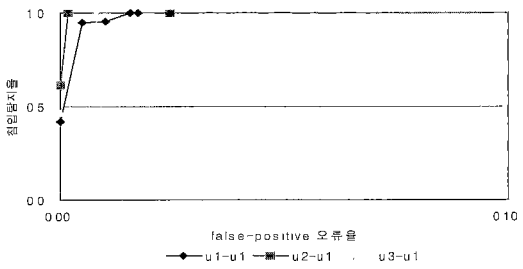

그림 11 정상모델변화에 따른 ROC곡선(사용자1)

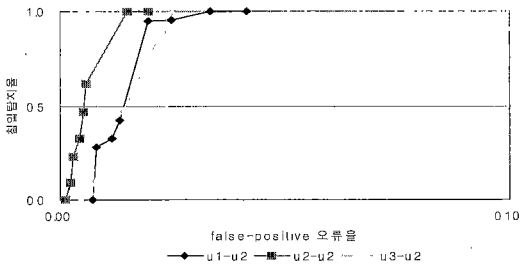

그림 12 정상모델변화에 따른 ROC곡선(사용자2)

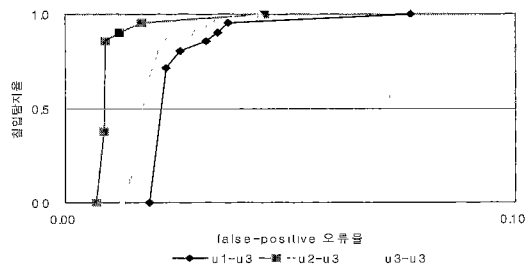

그림 13 정상모델변화에 따른 ROC곡선(사용자3)

4.3 그룹별 모델링

마지막 실험으로 비슷한 행동을 보이는 사용자들을 그룹으로 묶어서 그룹별로 모델링하여 보았다. 이는 모델링할 사용자의 증가에 따른 정상행위 모델링 비용을 감소시키려는 것과 유사한 행동을 보일 것으로 기대되는 사용자들의 행위데이타를 통합함으로써 해당 사용자 그룹의 다양한 정상행위를 수집하려는 것이다. 그룹 선정시에는 사용자1과 3이 동일과제를 수행하고 있다는 것과 표 3에서 보듯이 두 사용자 사이의 행위차가 상대적으로 작다는 점을 고려하였다.

표 6와 7의 결과를 보면 사용자 1과 3을 그룹으로 묶은 경우 사용자 1의 오류율은 개선되었지만 사용자 3의 오류율은 나빠졌다. 실험대상 사용자집단의 크기가 유사성을 찾을 만큼 충분히 크지 않아서 상대적으로 유사한 두 사용자의 행위 유사성이 기대만큼 크지 않았던 것으로 보인다.

표 6 그룹별 모델링에 따른 비정상행위 탐지결과

	사용자별 모델링			그룹 모델링 (사용자1,3)
	사용자 1	사용자 2	사용자 3	
정상행위 평균(편차)	-15.54 (5.88)	-4.72 (8.42)	-14.25 (12.34)	-16.20 (8.70)
임 계 값	-26.21	-31.35	-36.92	-27.83
탐 지 율	100%	100%	100%	100%
false-positive 오류율	4.31%	1.73%	1.96%	4.09%

표 7 모델링방법에 따른 비정상행위 탐지결과

	단일모델링	사용자별 모델링	그룹 모델링
탐 지 율	100%	100%	100%
false-positive 오류율	2.95%	2.67%	3.30%

5. 결론 및 향후연구

본 논문에서는 HMM을 사용하여 사용자의 정상행위를 모델링한 후 순서적으로 생성되는 이벤트를 분석하여 비정상행위를 판정하는 침입탐지 시스템을 제안하였다. 본 시스템은 침입발생시 적절하게 이를 탐지하였으며, 모델링된 사용자 외의 다른 사용자에 의한 사용도 탐지하였다. false-positive 오류가 발생한 부분은 학습시 정상행위로 모델링되지 않은 것이었다. 포괄적으로 정상행위 데이타를 제공할 수 있다면 HMM이 효과적인 침입탐지기법으로 사용될 수 있음을 알 수 있었다.

모델간 비교를 통해서는 하나의 모델로 모델링한 것보다 사용자별로 모델링한 경우가 오류가 더 적음을 확인하였다. 사용자 그룹별로 모델링한 경우에는 성능의 개선을 보여주지 못하였지만 사용자집단의 크기가 충분히 크고 사용자별, 행위별 유사성 판정 방식을 개선한다면 처리성능 및 모델링 능력을 향상시킬 수 있을 것으로 보인다.

현재는 시스템의 판정능력을 보이기 위하여 ROC 곡선을 이용하여 임계값이 변함에 따라 시스템의 오류율이 어떻게 변하는지를 보여주고 있으나 실제 침입탐지에 이용되기 위해서는 임계값을 상황에 맞도록 시스템

내에서 적절하게 설정할 수 있는 장치가 마련되어야 할 것이다. 앞으로 변별력있는 사용자 이벤트 추출에 관한 연구와 사용자 및 사용패턴간의 유사성 판정에 관한 연구가 수행되어야 할 것이다.

참 고 문 헌

[1] E. H. Spafford, *Security Seminar*, Department of Computer Science, Purdue University, January 1996.

[2] 한국정보보호센터, *호스트기반 침입탐지시스템 개발에 관한 연구*, 1998년 12월.

[3] S. Kumar and E. H. Spafford, "An application of pattern matching in intrusion detection," *Technical Report CSD-TR-94-013*, 1994.

[4] T. F. Lunt, "A survey of intrusion detection techniques," *Computer & Security*, vol. 12, no. 4, June 1993.

[5] T. Lane and C. E. Brodley, "Temporal sequence learning and data reduction for anomaly detection," *ACM Trans. on Information and System Security*, vol. 2, no. 3, pp. 295~331, August 1999.

[6] A. K. Ghosh, A. Schwartzbard and M. Schatz, "Learning program behavior profiles for intrusion detection," *Proc. Workshop on Intrusion Detection and Network Monitoring*, Santa Clara, USA, April 1999.

[7] C. Warrender, S. Forrest and B. Pearlmutter, "Detecting intrusions using system calls: Alternative data models," *Proc. IEEE Symposium on Security and Privacy*, May 1999.

[8] H. Deba, M. Dacier and A. Wespi, "Towards a taxonomy of intrusion-detection systems," *Computer Networks*, vol. 31, pp. 805~822, 1999.

[9] P. Porras and R. Kemmerer, "Penetration state transition analysis - A rule-based intrusion detection approach," *Proc. 8th Annual Computer Security Applications Conf.*, pp. 220~229, November 1992.

[10] P. Helman and G. Liepins, "Statistical foundations of audit trail analysis for the detection of computer misuse," *IEEE Trans. on Software Engineering*, vol. 19, no. 9, pp. 886~901, September 1993.

[11] P. Helman, G. Liepins and W. Richards, "Foundations of intrusion detection," *Proc. Fifth Computer Security Foundations Workshop*, pp. 114~120, Franconic, USA, June 1992.

[12] H. Javitz et al., "Next generation intrusion detection expert system (NIDES) - 1. statistical algorithms rationale - 2. rationale for proposed resolver," *Technical Report A016-Rationales*, SRI International, March 1993.

[13] H. Vaccaro and G. Liepins, "Detection of anomalous computer session activity," *Proc. 1989 IEEE Symposium on Research in Security and Privacy*, pp. 280~289, 1989.

[14] T. Spyrou and J. Darzentas, "Intention modelling: Approximating computer user intentions for detection and prediction of intrusions," *Information Systems Security*, pp. 39~335, Chapman & Hall, May 1996.

[15] S. Forrest, S. A. Hofmeyr and A. Somayaji, "Computer immunology," *CACM*, vol. 40, no. 10, pp. 88~96, October 1997.

[16] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff, "A Sense of Self for Unix Processes," *Proc. IEEE Symposium on Computer Security and Privacy*, pp. 120~128, Los Alamos, USA, 1996.

[17] Sunsoft, *Solaris 2.5 Sunshield Basic Security Module Guide*, 1995.

[18] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257~286, February 1989.

[19] T. Dunigan and G. Hinkel, "Intrusion detection and intrusion prevention on a large network: A case study," *Proc. Workshop on Intrusion Detection and Network Monitoring*, Santa Clara, USA, April 1999.

최 종 호

1998년 연세대학교 컴퓨터과학과 졸업(학사). 2001년 연세대학교 대학원 컴퓨터과학과 졸업(석사). 관심분야는 인공지능, 정보보안, 침입탐지시스템

조 성 배

1988년 연세대학교 전산과학과(학사). 1990년 한국과학기술원 전산학과(석사). 1993년 한국과학기술원 전산학과(박사). 1993년 ~ 1995년 일본 ATR 인간정보통신연구소 객원 연구원. 1998년 호주 Univ. of New South Wales 초청연구원. 1995년 ~ 현재 연세대학교 컴퓨터과학과 부교수. 관심분야는 신경망, 패턴인식, 지능정보처리