권한이동 모델링을 통한 은닉 마르코프 모델 기반 침입탐지 시스템의 성능 향상

(Performance Improvement of Intrusion Detection System based on Hidden Markov Model through Privilege Flows Modeling)

박 혁 장 † 조 성 배 ††
(Hyuk-Jang Park) (Sung-Bae Cho)

요 약 기존 침입탐지시스템에서는 구현의 용이성 때문에 오용침입탐지 기법이 주로 사용되었지만, 새로운 침입에 대처하기 위해서는 궁극적으로 비정상행위탐지 기법이 요구된다. 그 중 HMM기법은 생성메커니즘을 알 수 없는 이벤트들을 모델링하고 평가하는 도구로서 다른 침입탐지기법에 비해 침입탐지율이 높은 장점이 있다. 하지만 높은 성능에 비해 정상행위 모델링 시간이 오래 걸리는 단점이 있는데, 본 논문에는 실제 해킹에 사용되고 있는 다양한 침입패턴을 분석하여 권한이동시의 이벤트 추출방법을 이용한 모델링 기법을 제안하였고 이를 통하여 모델링 시간과 False-Positive 오류를 줄일 수 있는 지 평가해 보았다. 실험결과 전체 이벤트 모델링에 비해 탐지율이 증가하였고 시간 또한 단축됨을 알 수 있었다.

키워드 : 침입탐지시스템, 비정상행위 탐지, 은닉마르코프모델, 권한이동, 감사자료 축소

Abstract Anomaly detection techniques have been devised to address the limitations of misuse detection approach for intrusion detection. An HMM is a useful tool to model sequence information whose generation mechanism is not observable and is an optimal modeling technique to minimize false-positive error and to maximize detection rate. However, HMM has the short-coming of login training time. This paper proposes an effective HMM-based IDS that improves the modeling time and performance by only considering the events of privilege flows based on the domain knowledge of attacks. Experimental results show that training with the proposed method is significantly faster than the conventional method trained with all data, as well as no loss of recognition performance.

Key words : Intrusion Detection System, Anomaly Detection, Hidden Markov Model, Privilege Change, Audit Data Reduction

1. 서 론

전 세계 네트워크 환경의 확산과 고속화로 보다 신속하고 다양한 서비스가 가능해졌으며 개인의 업무 효율 또한 엄청난 향상을 가져 왔다. 하지만 컴퓨터에 대한 의존도가 높아짐에 따라 보안의 문제가 크게 대두되고 있다. 특히 최근 금융망이나 국방망, 전력망 등 중요기관들이 해킹당하는 사례가 늘고 있어, 불법적인 침입을 사전에 탐지하여 국가의 중요 정보통신 기반구조의 피해를 차단할 필요가 있다. 이에 따라 침입탐지에 대한 요구와 관심이 증가되었는데 이제까지는 주로 방화벽 등과 같은 시스템 보안 메커니즘 개발에 치중하였으며 불법적인 침입을 완벽하게 막을 수 있는 보안 시스템을 개발하기는 어렵다고 알려져 있었다. 하지만 최근 들어 많은 침입탐지 시스템들이 개발되고 있다.

침입탐지 시스템이란 내부자의 불법적인 사용, 오용 또는 외부 침입자에 의한 중요 정보 유출 및 변경을 알아내는 것으로서 각 운영체제에서 사용자가 발생시킨 키워드, 시스템 호출, 시스템 로그, 사용시간, 네트워크 패킷 등의 분석을 통하여 침입여부를 결정한다[1]. 요즘 사용되고 있는 대부분의 서버 컴퓨터 시스템은 시스템 내에서 발생한 이벤트에 관한 자세한 정보를 얻을 수

· 이 논문은 정보통신부 대학정보통신 연구센터 지원사업의 지원 및 한국소프트웨어진흥원의 관리로 수행되었음.
† 비 회 원 : 연세대학교 컴퓨터과학과
twinkler@candy.yonsei.ac.kr
†† 종신회원 : 연세대학교 컴퓨터과학과 교수
sbcho@candy.yonsei.ac.kr
논문접수 : 2002년 6월 25일
심사완료 : 2002년 10월 11일

있는 C2 이상의 보안 감사 프로그램을 자체적으로 지원하여 잠재적으로 보안에 영향을 주는 모든 이벤트를 기록하기 때문에 쉽게 정보를 얻을 수 있다[2].

침입탐지 기술의 세계적인 현황은 아직 기술적 미성숙 상태이지만 역동적으로 다양한 실험적 제품이 개발되고 있다. 일반적으로 침입탐지 시스템은 모델링 방법에 따라 두가지로 나눌 수 있다. 첫째는 공격에 관한 축적된 지식을 사용하여 침입의 증거를 찾는 방식으로 오용탐지(misuse detection) 또는 지식기반 기법이라고 한다. 두 번째는 사용자행위에 관한 정상행위모델을 생성한 후 여기에서 벗어나는 경우를 찾는 방식으로 비정상행위탐지(anomaly detection) 또는 행동기반 기법이라고 한다[3]. 오용탐지 기법은 기존에 잘 알려진 침입일 경우에 높은 탐지 성능을 보여 주지만 잘 알려지지 않은 침입일 경우에는 탐지할 수 없다는 단점이 있다. 반면에, 비정상행위 침입탐지시스템은 예상치 못한 취약점을 이용한 침입의 탐지가 가능하고 보안상의 취약점을 사용하지 않는 권한 남용형 공격탐지가 가능하다. 비정상행위 탐지를 위해 사용되는 방법은 크게 통계적인 방법(statistical approach), 규칙기반 전문가 시스템, 신경망(neural network), 면역 시스템, 은닉 마르코프 모델(hidden Markov model) 등이 있다.

통계적인 방법을 사용한 대표적 침입탐지 시스템은 NIDES를 들 수 있는데 이 시스템은 프로파일에 있는 장기간의 사건 유형과 최근에 있었던 사건 유형의 일치 정도 측정을 위해 감사레코드(Audit record)에 대한 분포상태를 측정하는 방식을 사용하여 모델링하였다[4]. 하지만 모델링할 수 있는 침입행위의 종류가 제한적이라는 문제 때문에 사용하기는 힘들다. Los Alamos National Laboratory에서는 전문가의 지식을 모델링하기 위한 도구로서 규칙기반언어를 사용하여 NADIR과 Wisdom & Sense라는 시스템을 개발하였다[5]. 이 시스템에서는 적절한 사용정책을 기술하는 규칙집합에 기반하여 사용자들의 행동을 검사하고 패턴에 맞지 않는 행동을 찾아낸다. 또한 보안정책이 적절히 적용되고 있는지 검증하는데도 사용된다. 하지만 전문가 시스템은 공격에 관한 지식, 정상행위에 대한 지식을 추출하기 쉽지 않기 때문에 구현이 매우 어려운 단점을 가지고 있다. 신경망을 사용한 대표적 침입탐지 시스템은 프랑스의 CS Telecom에서 개발한 Hyperview를 들 수 있다. 신경망을 통한 사용자의 행위 학습의 경우 시간의 흐름에 관한 데이타의 윈도우를 신경망 입력값으로 매핑하여 사용하는데 입력값으로 명령어이름, CPU 사용량, 메모리 사용량 등 60개의 감사자료를 이용한다. 하지만 입력과 출력간의 관계를 명시적으로 알 수 없고 모델링에 많은 계산량이 필요한 단점이 있다[6]. 또한 면역시스템 합법적인 행동을 통한 불법적인 접근을 탐지할 수 없는 단점이 있다.

HMM모델은 생성메커니즘을 알 수 없는 이벤트들을 모델링하고 평가하는 도구로서 뉴멕시코 대학의 S. Forrest에 의해 고안된 시스템 호출을 기반으로 한 침입탐지시스템이 대표적이다[7]. 미국의 아리조나 주립대학이나 뉴 멕시코 대학의 RAID 등에서 제안한 HMM 기반 침입탐지시스템은 다른 비정상행위 탐지 기법들에 비해 False-Positive 오류를 최소화하면서 탐지율을 높일 수 있다는 장점으로 많은 관심을 모았다. 또한 HMM을 이용하여 BSM 데이타에서 발생하는 정상행위 이벤트를 모델링하는 경우, 시스템 호출관련 Measure를 이용하면 좋은 결과를 얻을 수 있었다[8]. 하지만 높은 성능에 비해 정상행위 모델링과 침입 판정시 매우 많은 시간을 소요하고 오용탐지 기법에 비해 오류율이 상대적으로 높다는 문제 때문에 실시간 침입탐지에 직접 사용하기는 어렵다. 따라서 시스템 자체의 성능을 개선하거나, 정상행위 모델링을 위한 데이타의 효과적인 축약기법이 필요하다.

본 논문에서는 HMM 기반 비정상행위 탐지 시스템의 성능을 향상시키기 위하여 다양한 침입패턴들을 분석하여 불필요한 정보를 축약하는 최적의 Measure 추출 방법을 제안한다. 또한 실험을 통해 모델링 시간이 얼마나 단축되는지, 그리고 높은 탐지율과 낮은 False-positive 오류율을 얻을 수 있는지 평가해 본다.

2. 권한 이동 이벤트

2.1 침입유형

침입의 궁극적 목표는 루트 권한의 획득이라고 할 수 있는데 최근 유행하고 있는 버퍼 오버플로우(Buffer Overflow) 등이 대표적인 방법이다. 이러한 침입은 사용되지 않거나 패스워드가 없는 일반 유저의 권한으로 침입하여 고도화된 공격기법을 이용, 루트의 권한을 획득한다. 여기에서는 호스트기반으로 발생할 수 있는 침입유형을 분석해보고 이를 통하여 각 유형에 적절한 measure들에 대해서 알아본다.

현재 업체 및 기관에서 주로 사용하고 있는 해킹 분류안은 침입차단시스템(Firewall) 공개 SW를 최초로 개발한 Marcus J. Ranum에 의해 분류되었고 한국정보보호원 CERTCC-KR에서는 이 분류 방법을 일부 확장하여 10가지로 분류하였다[9]. 그 중 호스트에서 발생 가능한 유형은 버퍼오버플로우, S/W 보안오류, 구성설

정 오류, 서비스거부공격 등이 있다.

2.2 권한이동 시점의 measure 추출

CERTCC의 2002년도 1월과 2월 해킹 동향을 보면 CERTCC 취약성 분류 중 호스트 기반 침입이 가능한 취약성인 S/W 보안오류와, 버퍼오버플로우, 구성 설정 오류, 서비스 거부공격 등의 발생수가 1월달에는 44회, 2월달에는 46회가 발생된 것을 알 수 있다(표 1). 그 중 버퍼오버플로우 취약성을 이용한 방법과 사용자 권한 설정 오류를 이용한 방법이 거의 90%이상을 차지하고 있음을 알 수 있었다.

표 1 최근 해킹 동향

침입유형	2001년 12월	2002년 1월	2002년 2월
S/W 보안오류	2	0	0
버퍼오버플로우	16	35	38
구성,설정 오류	3	4	4
서비스 거부공격	3	4	6
총 갯수	24	43	45

본 논문에서는 공격행위의 특성을 알아보기 위해 4가지 침입유형에 대한 17개의 공격패턴을 수집하여 공격 유형별로 침입 시 발생되는 이벤트 정보를 분석하였다. 침입 분석에 사용된 Exploit는 표 2와 같다.

분석 결과 모든 버퍼오버플로우와 S/W 보안오류 침입 성공 시 실제 User ID(UID)와 Effective User ID (EUID)가 변한 후 EUID가 root의 권한으로 계속적으로 유지되는 것을 확인할 수 있었다. 대표적인 버퍼오버플로우 공격인 xlock은 X 윈도우 화면을 잠그는 프로그램으로 사용자가 입력한 인수에 대해 길이를 충분하게 검사하지 않기 때문에 프로그램의 내부 스택을 겹쳐 쓸 수 있는 버그를 가지고 있다. 이 공격 파일을 분석해 보면 setuid root로 설정되어 있기 때문에 침입자는 루트 관리자의 권한으로 임의의 프로그램을 실행시킬 수 있다. 정상적인 xlock 파일과 exploit 파일의 실행을 비교해보면 가장 큰 차이점은 EUID 권한의 변경에서 큰 차이점을 발견할 수 있다. 일반 파일의 경우 권한이동이 일어난 후 다시 exit 이벤트의 호출과 함께 EUID가 일반 유저로 바뀌지만 해킹시도 시에는 권한의 이동이 일어난 시점에서 execve 파일로 root 소유의 shell을 실행시켜 EUID는 root로 유지된다.

이렇듯 호스트 관련 침입의 대부분은 시스템의 버그나 사용자의 잘못된 사용을 통해 루트의 권한을 획득함으로써 발생하는 것을 알 수 있었다. 따라서 정상적인 루트 권한 획득에 관련된 정보를 모델링하여 공격행위를 탐지한다면 도스공격 등을 제외한 90%이상의 불법

표 2 분석에 사용된 침입 유형

Solaris 버전	침입 유형	침입 형태
2.5.1 sparc	버퍼오버플로우	Solaris ufsrestore vulnerability
	버퍼오버플로우	rpcbind file overwrite Vulnerability
	버퍼오버플로우	Libc (getopt() bug) stack overflow
	버퍼오버플로우	Eject exploit
	버퍼오버플로우	Passwd stack overflow
	버퍼오버플로우	Buffer overflow in /bin/fdformat
	구성 설정 오류	IFS환경 변수이용 침입
2.7 sparc	버퍼오버플로우	OpenView xlock Heap Overflow
	버퍼오버플로우	Lpset -r Buffer Overflow Vulnerability
	S/W 보안오류	DTMail Mail Environment Variable Buffer Overflow
	버퍼오버플로우	libc2_list_devices exploit
	버퍼오버플로우	ufsrestore Vulnerability
2.8	DOS 공격	디스크채우기, 메모리고갈, 프로세스 만들기
	버퍼오버플로우	whodo Buffer Overflow Vulnerability
	버퍼오버플로우	kcms_configure KCMS_PROFILES Buffer Overflow
	S/W 보안오류	mailx -F Buffer Overflow Vulnerability
	버퍼오버플로우	libsldap Buffer Overflow Vulnerability

적인 권한 이동 공격을 감시할 수 있으며, 그 결과 대부분의 호스트 기반 침입을 탐지할 수 있다. 게다가 침입탐지의 감시 대상을 효과적으로 줄여 침입탐지를 위한 시스템 자원의 사용을 획기적으로 줄일 수 있기에 실용화 가능성을 보일 수 있다.

3. 권한이동 모델링 기반 침입탐지시스템

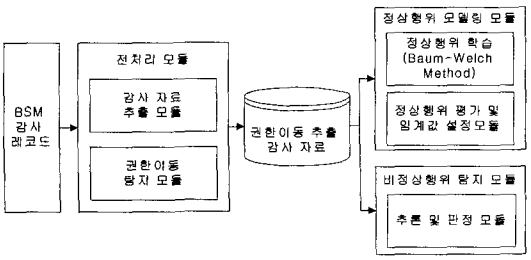


그림 1 시스템 전체 구조

그림 1은 사용자 권한이동 기반 침입탐지시스템의 전체 구조이다. 시스템은 크게 전처리모듈, 정상행위 모델링모듈, 침입 탐지모듈로 나뉜다. 정상행위 모델 모듈은 수집된 정상 감사 자료로부터 은닉 마르코프 모델 기반의 정상행위 모델을 생성하고, 침입탐지 모듈은 수집되는 탐지대상 감사 자료를 정상행위 모델과 비교해 침입을 판정한다. 제안하는 권한이동 침입탐지시스템의 핵심 기술은 정상행위 모델링을 위한 은닉 마르코프 모델의 사용과 권한이동 시점의 정보 추출을 통한 감사자료 필터링에 있다.

3.1 감사자료 수집

감사 자료를 수집, 축약하는데 가장 쉽게 이용할 수 있는 감사 자료는 시스템 로그파일이다. 유닉스 시스템의 경우 /var/log/message나 /var/adm/ 디렉토리의 wtmp, utmp, sulog와 같은 사용자 로그 정보가 존재하는데, 이와 같이 기본적인 로그 파일을 이용할 경우 특별한 작업 없이 손쉽게 감사 자료를 얻을 수 있다는 장점이 있지만 대부분의 버퍼 오버플로우 공격 도중의 권한 취득 증거를 찾기가 쉽지 않고 침입자들이 침입에 성공한 경우 주요 로그파일에서 자신의 흔적을 손쉽게 지울 수 있다[10]. 이러한 단점 때문에 사용되는 것이 시스템 호출레벨의 감사 자료이다.

기본보안모듈에서 제공하는 감사 자료는 그림 2와 같은 정보를 포함하고 있으며 감사자료 수집기에서는 관련 정보들을 기본보안모듈로부터 수집하고 감사자료 변환을 위한 형태로 메모리에 저장한다. 수집되는 시스템 감사 정보의 종류는 총 242가지인데, 이중 id 255번까지는 운영체제에서 제공하는 시스템 호출이고 id 1000번 이후부터는 사용자 정의 시스템 호출이다.

3.2 권한이동 시점의 정보 추출

유닉스는 여러 사용자가 한 시스템의 제한된 디스크와 메모리를 사용하기 위하여 각 사용자마다 권한을 부여하고 있다. 그런데 시스템 프로그램의 오류나 잘못된 사용으로 인해 사용자가 쉽게 슈퍼유저의 권한을 얻을 수 있는 문제가 있다. 정상적인 권한 이동은 관리자가 일반사용자의 권한으로 들어와 작업도중 SU 명령 등을 통하여 루트의 권한을 획득하거나 일반 사용자가 잠시 슈퍼유저 소유의 시스템 파일(SETUID)로 되어 있는 명령어를 실행시킬 때 발생한다. 이때 SETUID라는 것은 임시적으로 사용자의 권한을 바꿔줄 수 있는 규칙을 파일에 적용시켜주는 것으로서 어떤 사람이든 SETUID

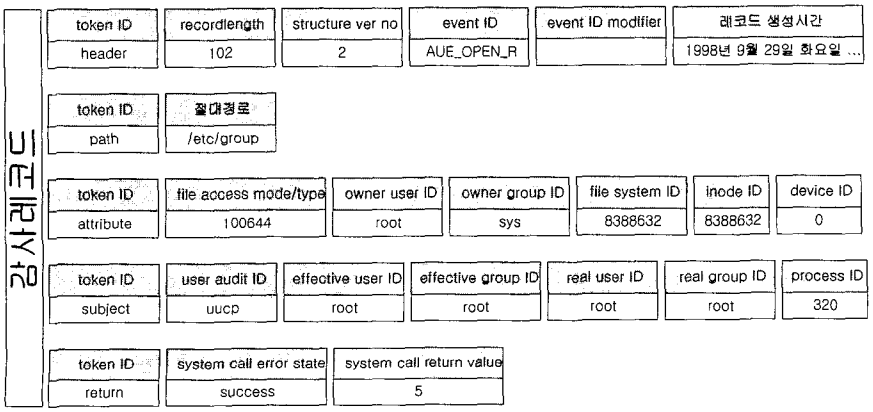


그림 2 BSM 감사 레코드

로 설정된 파일을 실행하면 그 파일의 소유 계정으로 프로그램이 실행된다[11].

일반적인 호스트 침입은 위와 같은 시스템 구조의 취약성을 이용한다. 그림 3은 버퍼 오버플로우 공격 시 권한이동의 예인데, 일반상태(Q0)에서 fdformat 명령을 사용하면 사용자는 잠시 슈퍼유저의 권한으로 작업을 수행하고(Q1) 해당 작업이 종료된 후 본래의 상태(Q0)로 돌아가야 한다. 그런데 fdformat 작업 도중 버퍼 오버플로우 공격이 발생되면 fdformat의 작업이 수행된 후 본래의 상태(Q0)로 돌아가지 않고 슈퍼유저의 권한이 그대로 유지되는 상태(Q2)로 전환된다. 결국 해커는 Q2에서 슈퍼유저의 권한으로 시스템에 접근하게 된다. 그런데 보안모듈이 fdformat과 관련된 정상적인 권한이동의 시퀀스를 알고 있다면, 버퍼 오버플로우에 의해 비정상적인 상태(Q2)가 되는 것을 발견할 수 있다. 사용자 권한이동 학습 모델은 이와 같은 정상적인 권한이 이동되는 시점 전의 정상 시스템 호출 시퀀스를 수집하고 모델링하여 정상모델을 생성한 후, 사용자들의 권한 이동상태를 비교하여 침입을 탐지한다.

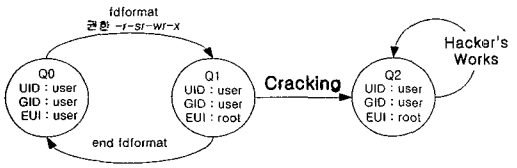


그림 3 버퍼 오버플로우 공격 시 권한이동 예

이외에도 유닉스의 파일 시스템 보안과 관련하여 허점을 가지고 있는 것으로는 링크(symlink) 메커니즘을 들 수 있다. 링크는 하나의 파일을 여러 개의 서로 다른 이름으로 접근할 수 있도록 하는 방법으로 하나의 파일을 복사하지 않고도 여러 경로를 통해 공유할 수 있도록 한다는 점에서 편리한 기능이다. 하지만 이 기능을 악용하여 프로그램 수행도중 생성되는 임시데이타를 SETUID가 설정된 프로그램과 연결시킴으로써 시스템에 심각한 문제를 일으킬 수 있다. 이렇듯 SETUID나 Symbolic Link 등을 통하여 잠시 바뀌는 권한 이동전의 행동들은 정상행위와 비정상 행위를 구분하는 중요한 시점이 된다.

제안하는 모델은 기본보안모듈 감사자료에서 EUID와 UID가 변경되었을 경우 그 시점을 기준으로 이전에 사용되었던 일정양의 데이타 시퀀스를 가지고 평가한다. 이때 기록의 범위는 임의적으로 변경하여 최적의 값을 찾게 된다. 권한이 이동되는 시점의 기본보안모듈 데이타를 분석하면 표 3과 같은 시스템 호출 이벤트들이 발생하는데, 분석 결과 침입이 시도되었을 경우에는 직접 실행되는 execve 이벤트 전에 symlink, setpgrp, seteuid, vfork 등이 많이 발생되었다. [2]에서는 평가를 통하여 어떠한 이벤트들이 권한 이동을 야기하는지와 해킹 시도 시에 발생하는 Process들과 이벤트들이 무엇인지를 파악하였는데, 이러한 차이점을 이용하여 권한이동 순간을 파악할 수 있다. 이러한 방법을 통해 EUID와 UID가 변경되었을 경우 그 시점을 기준으로 전에 사용되었던 일정양의 데이타 시퀀스를 따로 저장한 후 각각 은닉 마르코프 모델에 적용시킨다.

3.3 침입탐지

감사자료 수집기에 의해 수집되고 감사자료 변환기를 거쳐 필터링된 정보는 감사자료 평가기에서 수행되는 forward-backward procedure를 통해 정상행위 모델과 비교되어 침입이 판정된다. 침입 판정에서는 이미 구축되어 있는 은닉 마르코프 모델 기반 정상모델에 사용자 행위 시퀀스를 입력하여 각 정상행위에서 현재 행위가 생성되었을 확률을 구한다. 확률을 구하는 방법으로는 forward-backward procedure나 Viterbi 알고리즘을 사용할 수 있다. 각 모델별로 구해진 확률은 판정모듈에 전달되어 비정상행위인지 판정한다. 이때 정상행위 모델링으로부터 구해진 임계값(threshold)과 비교하여 더 낮

표 3 권한이동 관련 이벤트

이벤트 아이디	시스템 호출	이벤트 아이디	시스템 호출	이벤트 아이디	시스템 호출
2	fork	27	setpgrp	200	setuid
11	chown	38	fchroot	214	setegid
16	stat	39	fchown	215	seteuid
21	symlink	40	setreuid	237	lchown
23	execve	41	setrgid	6158	rsh
24	chroot	69	fchroot	6159	su

은 수치를 가질 경우 침입으로 간주하게 된다. forward-backward procedure에서는 forward 변수인 α를 사용해서 입력시퀀스가 해당 모델로부터 나왔을 확률 Pr $(O|\lambda)$를 계산한다. α는 시간 t에 부분관찰 시퀀스 $O_1, O_2,..,O_t$를 보고 상태 q_i에 있을 확률로 다음과 같이 정의된다.

$$\alpha_t(i) = \Pr(O_1, O_2, \ldots, O_t, i_t = q_i|\lambda)$$

$\alpha_T(i)$는 입력시퀀스 O의 모든 심볼을 순서에 맞게 가지고 있으면서 최종상태가 i인 확률을 나타낸다. $\alpha_T(i)$를 모든 상태 i에 대해 고려하면 $\Pr(O|\lambda) = \Pr(O_1, O_2, \ldots, O_T|\lambda)$를 구할 수 있다. $\alpha_t(i)$는 다음 절차에 의해 귀납적으로 구할 수 있다.

- 단계 1 (초기화)

$$\alpha_1(i) = \pi_i b_i(O_1)$$

- 단계 2 (귀납)

$$\text{for} \quad t = 1 \text{ to } T-1$$
$$\alpha_{t+1}(j) = \left[\sum_{j=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1})$$

- 단계 3 (종료)

$$\Pr(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

backward 변수 β는 $\beta_t(i) = \Pr(O_{t+1}, O_{t+2}, \ldots, O_T|i_t = q_i, \lambda)$로 정의되며 forward변수와 유사한 과정에 의해서 구할 수 있다.

- 단계 1 (초기화)

$$\beta_T(i) = 1$$

- 단계 2 (귀납)

$$\text{for} \quad t = T-1 \text{ to } 1$$
$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$$

3.4 정상행위 모델 생성

정상행위 모델링은 전처리 단계에서 생성된 정상행위 시퀀스를 기반으로 은닉 마르코프 모델의 파라메터를 결정하는 과정이다. 은닉 마르코프 모델은 실제적인 생성모델을 알 수 없고 단지 생성된 시퀀스에 의해서만 확률적으로 관측할 수 있는 이중의 확률 절차로서, 사용자의 행위 시퀀스를 모델링하기에 유용한 도구이다[12][13]. 이것은 고정된 값인 관찰 시퀀스의 길이, 상태 수, 심볼 수와 학습에 의해 조정되는 전이확률, 관측확률, 초기상태분포로 구성된다. 전이확률은 한 상태에서 다음 상태로 전이할 확률을 나타내며, 관측확률은 한 상태에서 특정 심볼이 관측될 확률을 나타낸다. 초기 상태 분포는 처음에 해당 상태에서 시작할 확률을 나타낸다. 은닉 마르코프 모델은 다음과 같이 표현되며, 모델 λ는 학습에 의해 조정되는 변수들로 간략히 (A, B, π)로 나타낸다. 은닉 마르코프 모델의 파라메터 결정은 주어진 시퀀스가 해당 모델 λ로부터 나왔을 확률인 $\Pr(O|\lambda)$값이 최대가 되도록 $\lambda = (A, B, \pi)$를 조정한다. 이를 계산하는 해석적인 방법은 알려져 있지 않고 반복적으로 λ를 결정하는 방법으로 Baum-Welch의 재추정식이 있다.

Baum-Welch 재추정식에서는 두 개의 변수가 추가로 사용된다. $\xi_t(i,j)$는 시간 t에 상태 q_i에 있다 시간 t+1에 상태 q_j에 있을 확률로 정의되며 다음과 같이 표현될 수 있다.

$$\xi_t(i,j) = \Pr(i_t = q_i, i_{t+1} = q_j|O, \lambda)$$
$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\Pr(O|\lambda)}$$

$\gamma_t(i)$는 시간 t에 상태 q_i에 있을 확률이며 다음 수식을 통해 구할 수 있다.

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$

두 값을 시간 t에 대해 각각 합을 취하면 하나의 시퀀스에서 각각 상태 i에서 j로 변할 기대값과 상태 i에 있을 기대값을 구할 수 있다. 위의 값이 구해지면 다음 수식에 의해서 새 모델 $\overline{\lambda} = (\overline{a}, \overline{b}, \overline{\pi})$를 구할 수 있다. 시퀀스 O를 관찰한 결과로 $\overline{\lambda}$를 구한 후 $\Pr(O|\lambda)$와 $\Pr(O|\overline{\lambda})$를 비교한다. $\Pr(O|\lambda)$가 더 크다면 우도 함수의 임계점에 다다랐으므로 재추정 과정을 종료한다. Pr $(O|\overline{\lambda})$가 더 큰 경우는 더 나은 모델이 생성된 경우이며 λ를 $\overline{\lambda}$로 대치한 후 재추정 과정을 반복한다.

$$\overline{\pi_i} = \text{시간 } (t=1) \text{에 상태 } S_i \text{에 있을 빈도}$$
$$= \gamma_1(i)$$

$$\overline{a_{ij}} = \frac{\text{상태 } S_i \text{에서 상태 } S_j \text{로 전이할 기대횟수}}{\text{상태 } S_i \text{에서 전이할 기대횟수}}$$
$$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\overline{b_j(k)} = \frac{\text{상태 } j \text{에서 심볼 } v_k \text{를 볼 기대횟수}}{\text{상태 } j \text{에 있을 기대횟수}}$$
$$= \frac{\sum_{t=1 \; s.t. \; O_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

4. 성능평가 및 결과

4.1 실험 환경

실험을 위해 10명의 사용자가 참여하였으며, 정상행위의 경우 한 달에 걸쳐 수집하였다. 주로 사용된 프로그램은 메일 서버, 홈페이지 접속, 유닉스 명령어 실행, FTP 데이타 전송 등이다. 다양한 상황 구성을 위해 매

번 다른 방법으로 프로그램을 실행시켰으며 총 수집된 데이타는 90메가바이트이고 시스템 호출의 개수는 767,237 개이다. 각 데이타는 2대의 Solaris 2.5.1 서버에서 수집되었다. 테스트에 사용된 침입은 총 6개, 시도된 침입은 18차례이며 3명이 각 Exploit에 대해서 3번씩 침입을 시도하였다. 실험에 사용된 침입탐지 시스템은 모든 정상행위를 단일 모델에 사용한 방법, 정상행위를 사용자 별로 나누어 모델링한 방법, 권한이동을 단일 모델에 사용한 방법, 사용자 별 모델링을 통하여 권한이동을 탐지하는 방법 등으로 구현하여 장단점을 비교한다. HMM의 상태값은 실험을 통하여 5, 7, 10, 15로 변화시켰고 시퀀스 길이도 20, 25, 27, 30으로 변화해 가며 실험하였다. 또한 정상행위 모두를 학습시켰을때의 모델링 시간에 비해 대상을 권한이동 순간으로 하였을 때 모델링 시간이 얼마나 단축되는지에 대해서도 실험해 보았다.

실험을 위한 침입 유형은 표 4와 같다. 침입 유형으로는 DARPA 프로젝트에서 사용되어진 침입과 그 외 다양한 침입패턴들을 적용시켜 보았는데 대표적인 침입유형으로는 버퍼 오버플로우, 레이싱 컨디션, CGI 취약점을 이용한 침입이다. 대부분의 호스트기반 침입은 일반 사용자에서 루트로 권한을 옮기는 U2R형태의 침입이고 기본보안모듈을 통한 네트워크로부터의 침입탐지 가능성을 알아보기 위해 R2U 침입형태도 시도하여 보았다.

표 4 실험에 사용된 침입유형

Solaris 버전	침입 형태
2.5.1 sparc	Solaris ufsrestore vulnerability
	Rpcbind file overwrite Vulnerability
	Libc (getopt() bug) stack overflow
	Eject exploit
	Passwd stack overflow
	Buffer overflow in /bin/fdformat
	CGI Exploit

4.2 실험 결과

수집된 정상행위를 은닉마르코프 모델을 통하여 학습시킨 후 정상적인 데이타와 취약성을 가진 xlock파일의 정상적인 실행 데이타, 해킹을 시도한 데이타 등, 3가지 데이타를 가지고 HMM 평가값을 비교하여 보았다. 정상행위 데이타는 임계값이 -73.2였으며 가장 기본적인 명령어인 Ls 명령어에 대한 평가값은 최소 -57.3까지 나왔고 평균 -39.7 정도였다. 실험을 통하여 정상행위 임계값에 비하여 매우 높음을 알 수 있다. 또한 xlock 파일은 정상행위의 경우 최소 -68.1이 나왔으며 해킹 시도 시 -84.3으로 임계값보다 낮은 수치를 보여 주었다. 실험을 통하여 HMM을 이용한 침입탐지가 매우 효과적임을 알 수 있었다.

(1) 정상행위 모델링 시간 측정

표 5는 4가지 모델링방법에서 소요된 시간과 실험에 사용된 시퀀스의 총갯수를 보여 주고 있다. 보다 신뢰할 수 있는 시간 측정을 위해 각각 5번씩 실험하였다.

실험 결과 전체 학습 데이타를 이용할 경우 이전연구와 마찬가지로 시간이 매우 많이 소요된다는 것을 알 수 있었다. 이는 하루에도 엄청난 양의 감사자료를 발생시키는 대규모 서버에서 학습하는데 몇 달이 걸릴 수 있고 실시간 탐지가 거의 불가능하다는 것을 의미한다. 하지만 침입의 핵심이 되는 권한 이동 순간만을 수집 모델링 했을 경우 시퀀스 개수가 약 1/132정도로 단축되었고 이로 인해 모델링 시간 또한 약 1/256로 단축되었다. 본 실험에서는 상태와 시퀀스의 길이에 따른 시간의 변화 또한 측정하였는데, 시퀀스의 길이에 따른 변화보다는 상태의 변화에 따라 시간이 차이가 많이 나는 것을 알 수 있었다. 즉 같은 탐지율을 얻을 수 있다면 상태를 최저로 하는 것이 시스템의 성능 향상에 도움이 될 것이다.

(2) 각 모델링 별 성능 비교

본 실험에서는 테스트 데이타 실험 시 정상행위 모델

표 5 각각의 모델링에 대한 학습 소요시간

모델링 기법	상태/시퀀스	시퀀스 갯수	소요 시간
전체-사용자별 모델링	5/20	754716	3시간 11분
	15/30	754694	4시간 35분
전체-단일 모델링	5/20	767218	5시간 07분
	15/30	767208	6시간 29분
권한이동-사용자별 모델링	5/20	5437	16.2초
	15/30	5287	35초
권한이동-단일 모델링	5/20	5950	26.3초
	15/20	5950	43.5초
	5/30	5792	28초
	15/30	5792	57.5초

링으로부터 추출된 값을 기준으로 임계치를 조정시켜 각각의 상태에 따른 탐지 성능과 시퀀스 길이에 따른 탐지 성능을 보이기 위해 False-Positive 오류율과 침입탐지율에 대한 ROC(Receiver Operating Characteristics) 곡선으로 그렸다. ROC 분석이란 의사결정의 임계점을 선택하여 가설검정의 정확도와 측정오차에 대해 시각적이고 정량적인 리포트를 얻는 방법으로서 침입탐지 시스템을 평가하는 방법으로 많이 사용되고 있다. 바람직한 침입탐지 시스템은 낮은 False-Positive 오류에서 높은 침입탐지율을 보여 주어야 하므로 곡선이 왼쪽에 있을수록 좋은 성능을 나타낸다. 실험결과는 고정된 시퀀스 사이에서 상태수 변화에 따른 탐지율의 변화와 고정된 상태수 사이에서 시퀀스 변화에 따른 탐지율의 변화를 ROC 곡선으로 보였다.

가. 권한이동 단일 모델링

표 6은 실험으로 얻어진 결과 중 각 시퀀스별 최적의 성능을 보인 값들을 추출하여 표로 보여 주고 있다. 실험 결과 시퀀스의 길이나 상태에 상관없이 대부분 좋은 탐지율을 보여 주고 있고 시퀀스 25에서 30사이에서 최적의 탐지율을 보여 주었다. 호스트 1에서는 상태 10, 시퀀스 30에서 100% 탐지율 이었을 때 0.602%의 false-positive 오류율을 보여 주었고, Host 2에서는 상태 7, 시퀀스 25였을 때 최적의 false-positive 오류율을 보여 주었다. 하지만 시퀀스가 20 이하로 내려갈 경우 탐지율에 따른 false-positive 오류율이 증가하는 것을 볼 수 있는데, 이는 권한 이동 탐지 특성상 권한 이동 순간만을 탐지 대상으로 보기 때문에 침입의 시도가 있었을 경우 최소한 20 이상의 시퀀스를 보아야 한다는 것을 의미한다. 또한 상태의 수가 큰 탐지율의 변화를 보여 주지 못하며 심지어 호스트 2에서 시퀀스 25, 27을 기준으로 상태의 변화에 따른 false-positive 오류율을 비교하였을 경우 모두 같은 값을 보여주었다. 하지만 시퀀스 길이가 작을 때는 상태가 적을 때 좋은 성능을 보여주고 시퀀스가 증가할 경우 상태수 10이나 15에서 더 좋은 false-positive 오류율을 보이는 것은 시퀀스 길이가 길어질수록 상태의 수도 증가하는 것이 더 좋은 성능을 보여준다는 것을 의미한다.

나. 권한이동 사용자별 모델링

다음은 사용자 별로 권한이동 순간을 따로 수집하여 각 사용자별로 모델링을 한 후 사용자마다 임계값을 적용하여 탐지율을 측정하였다.

테스트 데이타에서 실제 해킹을 시도한 사용자는 Host 1에서는 사용자 1과 4이고 호스트 2에서는 사용자 1과 2이기 때문에 본 그래프에서는 4명의 사용자에 대해서만 결과를 보인다. 실험결과 호스트 1에서 사용자 1의 경우 언제나 탐지율 100% 인 경우에 false-positive 오류율이 0%가 나왔다. 사용자 4의 경우에도 2~5%정도의 false-positive 오류율로 비교적 좋은 성능을 보여주고 있다. 호스트 2에서 마찬가지로 사용자 1의 경우에는 시퀀스 30이 넘어 갈 경우를 제외하고는 시퀀스와 상태에 상관없이 모두 false-positive 오류율 0%에 탐지율 100%를 보여주고 있고 사용자 2의 경우는 상태 5 시퀀스 25에서 최적의 성능을 보여주었다. 이렇듯 사용자 모델링이 비록 사용자의 특성에 따라 차이가 많이 날 수는 있겠지만 단일 모델링과 결합하여 사용한다면 매우 좋은 성능을 보일 것이라고 사료된다.

다. 전체 데이타를 이용한 단일 모델링

다음 모델링 방법은 각 사용자의 학습 데이타 시퀀스를 통합하여 모델링하는 것으로 모든 사용자의 정상행위를 포함한다. 이 방법을 이용하였을 경우 다양한 패턴들이 정상행위에 들어가 있어 특정한 패턴으로 치우치게 되는 단점을 해결할 수 있지만 정상행위에 포함된 사용자가 참여율이 매우 낮고 사용 패턴이 남들과 매우 틀리다면 테스트 시 그 사용자의 행동을 침입으로 간주

표 6 F-P error를 이용한 침입탐지시스템 성능비교

운영체제	상태/시퀀스	HMM 임계값	침입횟수/유형	탐지율	F-P error
Host 1	5/20	-53.8	10/local	100%	6.707%
	15/25	-65.5	10/local	100%	2.367%
	10/27	-73.2	10/local	100%	2.439%
	10/30	-81.3	10/local	100%	0.602%
Host 2	5/20	-53.8	8/local	100%	5.172%
	7/25	-65.2	8/local	100%	1.149%
	10/27	-73.1	8/local	100%	2.881%
	15/30	-81.2	8/local	100%	4.899%

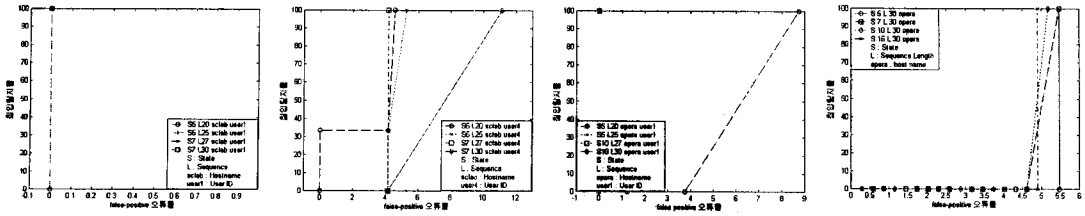


그림 4 권한이동 사용자별 침입탐지율

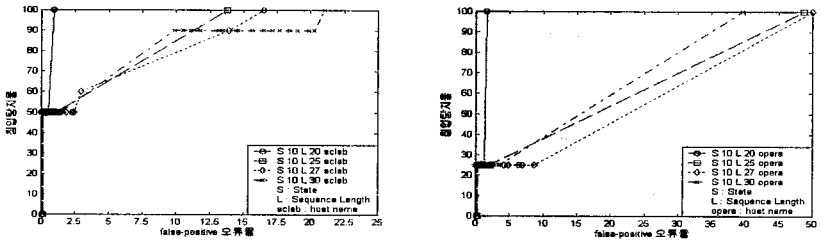


그림 5 전체 데이터-단일모델링 침입 탐지율

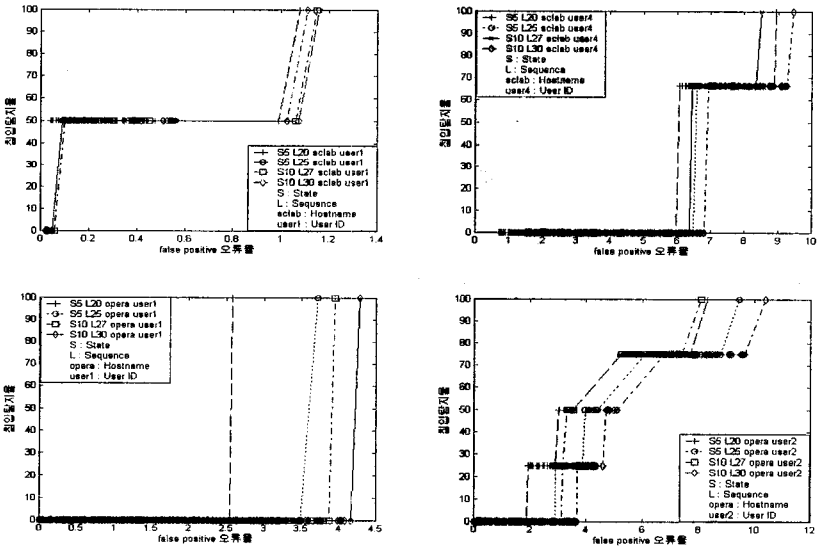


그림 6 전체 데이터-사용자별 모델링 침입탐지율

할 수 있는 가능성이 커지게 된다. 실험은 고정된 상태에서 시퀀스의 변화를 평가하는데 권한 이동 평가에서 상태에 따른 변화가 거의 없었기 때문에 고정된 시퀀스에서 상태의 변화는 측정하지 않았다.

실험 결과 호스트 1에서 최적의 탐지율은 상태 10, 시퀀스 20일 경우이며 100% 탐지율일 때 0.9453% false-positive 오류가 났다. 호스트 2에서도 마찬가지로 상태 10, 시퀀스 20에서 100% 탐지율일 때 1.6165%의 false-positive 오류율을 보여 주었다. 그밖에 시퀀스가 20일 경우 대부분 false-positive 오류율이 약 1%의 결과를 보여주었다. 하지만 시퀀스가 그 아래로 내려가든가 25이상으로 올라가면 그림 5에서 보듯이 false-positive 오류율이 현저하게 높아졌다. 즉 전체 데이타를 이용할 경우 권한이동에 비해 시퀀스 크기에 민감하게 작용하는 것을 알 수 있었다. 하지만 기존 실험과 마찬가지로 상태에 따른 변동은 크지 않다.

라. 전체 데이타를 이용한 사용자별 모델링

다음은 각 사용자의 학습 데이타 시퀀스를 추출하여 사용자별로 모델링하는 방법으로 권한이동 시퀀스를 통한 사용자별 탐지 방법과 전체 데이타를 이용한 단일 모델링 방법을 비교한다.

실험 결과 사용자에 따라 약간의 차이는 있지만 전체 데이타를 이용한 사용자별 모델링은 단일 모델링에 비하여 시퀀스의 크기에 적은 영향을 받는 것을 알 수 있다. 또한 탐지율이 단일 모델링에 비해 비교적 좋게 나온다. 최적의 탐지율은 상태 5, 시퀀스 20에서 1.076%의 false-positive 오류율을 보여 주고 있으며 그밖에 다른 유저들도 10% 안팎의 false-positive 오류율을 나타내고 있다.

5. 결론 및 향후 연구

본 논문에서는 HMM모델의 모델링 시간 지연문제를 극복하고 좀더 좋은 탐지율을 얻기 위해 권한이동 침입탐지 시스템을 제안하였다. 제안한 권한 이동 모듈은 BSM 데이타에서 발생되는 EUID와 UID가 변경되었을 경우 그 시점을 기준으로 전에 사용되었던 일정양의 데이타 시퀀스를 가지고 평가를 한다.

실험은 제안한 4가지 방법을 통하여 비교 평가 하였는데, 모델링의 경우 하나의 모델로 모델링한 것보다 사용자별로 모델링한 경우가 오류가 더 적음을 확인하였고, 사용자별 모델링이 단일 모델링에 비해 시퀀스나 상태에 민감하지 않아 평가하기가 쉬운 것을 알 수 있었다. 또한 권한이동 추출 방식이 기존의 전체 학습데이타를 이용하는 방식보다 시간상으로 약 1/256배 줄어들었고 성능은 오히려 더 좋아졌다. 특히 전체 학습데이타를 이용할 경우 비록 False-Positive 오류가 1%가 되어도 시퀀스의 개수 당 알람이 하루에도 수백 번 울릴 수 있지만, 권한이동관련 모델링시 시퀀스의 개수가 약 1/200로 줄어들기 때문에 관리하기가 수월하다.

하지만 권한이동 탐지 기법은 특성상 일반 사용자가 다른 사용자로 권한을 옮기는 것 이외에 DOS attack이나 CGI 버그를 이용한 정보 유출 시도 등 몇 가지 특정한 행위에는 취약함을 보여 줄 수있다. 이러한 문제를 위해 고수준의 문맥특성을 반영할 수 있는 모델링 방법에 대한 연구가 보충 되어야 할 것이다. 이외에도 좀더 정확한 탐지를 위해 각 프로세스의 CPU 사용량, I/O 정보를 이용한 다중 Measure 추출 방법 등의 연구가 필요하다. 마지막으로 다변적인 형태의 침입을 탐지하기 위한 IDS 간 연동성에 관한 연구와 서로간의 장점을 이용한 비정상행위-오용탐지 시스템 연계형 IDS의 연구 등 향후 통합보안 인프라 구성이 필요하다.

참 고 문 헌

[1] H.S. Vaccaro and G.E. Liepins, "Detection of anomalous computer session activity," Proc. IEEE Symp. on Research in Security and Privacy, pp. 280-289, 1989.

[2] K.E. Price, "Host_based misuse detection and conventional operating system's audit data collection," M.S. Dissertaion, Purdue University, Purdue, IN, December 1997.

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

[4] H.S. Javitz and A. Valdes, "The SRI IDES statistical anomaly detector," NIDES Technical Report, 1994.

[5] J. Hochberg, et al, "Nadir: An automated system for detecting network intrusion and misuse," Computers & Security, vol. 12, no. 3, pp. 235-248, 1993.

[6] H. Debar, M. Becker and D. Siboni, "A neural network component for an intrusion detection system," Proc. 1992 IEEE Computer Society Symposium on Research in Security and Privacy, pp. 240-250, Oakland, CA, May 1992.

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

[8] 최종호, 조성배, "은닉 마르코프 모델에 기반한 정상행위의 순서적 이벤트 모델링을 통한 침입탐지 시스템", 정보과학회, pp 306-308, 1999.

[9] CERTCC-KR, 한국정보보호진흥원, http://www.certcc.or.kr/

[10] S. Axelsson, "Research in Intrusion-Detection Systems: A Survey," Chalmers University of Technology, 1999.

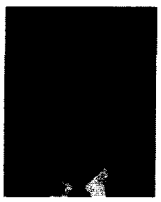
[11] B.A. Kuperman and Eugene H. Spafford. "Generation of application level audit data via library interposition," CERIAS TR 99-11, COAST Laboratory, Purdue University, West Lafaytte, IN, October 1998.

[12] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. of the IEEE, vol. 77, no. 2, 1989.

[13] L.R. Rabiner and B.H. Juang, "An introduction to hidden Markov models," IEEE ASSP Magazine, 1986.

박 혁 장

2000년 8월 ~ 2002년 8월 연세대학교 컴퓨터과학과 석사졸업. 관심분야는 침입탐지 시스템



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