연관마이닝 기법을 이용한 침입 시나리오 탐지를 위한 상태전이 알고리즘

김창수* · 황현숙**
*부경대학교 전자계산학과
**부경대학교 정보시스템학과
E-mail: cskim@dolphin.pknu.ac.kr

State Transition Algorithm for Penetration Scenarios Detection using Association Mining Technique

Chang-Soo Kim* · Hyun-Suk Hwang**
*Dept of Computer Science, PuKyong National University
**Dept of Information System, PuKyong National University

요 약

현재 인터넷 환경에서 크래킹은 보편화되어 있다. 이러한 크래킹을 탐지하거나 방어하기 위한 기법들은 대부분 기존의 불법 침입 유형을 분석하여 대응 알고리즘을 개발하는 것이 대부분이다. 현재 알려진 침입 탐지 기법은 비정상 탐지(Anomaly Detection)와 오용 탐지(Misuse Detection)로 분류할 수 있는데, 전자는 통계적 방법, 특징 추출 등을 이용하며, 후자는 조건부 확률, 전문가 시스템, 상태전이 분석, 패턴 매칭 등을 적용한다.

본 연구에서는 상태전이 기반의 연관 마이닝 기법을 이용한 침입 시나리오 탐지 알고리즘을 제안한다. 이를 위해 본 연구에서는 의사결정지원시스템에서 많이 적용한 연관 마이닝 기법을 여러 가지 불법 침입과 연관된 상태 정보를 분석할 수 있는 수정된 상태전이 알고리즘을 제시한다.

I. 서 론

인터넷 환경의 크래킹 발생은 계속적으로 증가하고 있으며, 새로운 유형의 침입 또한 증가하고 있다. 이러한 알려진 혹은 알려지지 않은 침입을 탐지하기 위한 연구들이 많이 진행되고 있다.

침입탐지는 크게 비정상 탐지(Anomaly Detection)와 오용 탐지(Misuse Detection)로 구분할 수 있으며, 전자는 컴퓨터 자원의 비정상적인 사용에 근거한 침입을 탐지하는 방법으로써 통계적 방법, 특징 추출, 신경망 등을 이용하고 있다. 후자는 시스템이나 응용 소프트웨어의 약점을 이용한 침입에 대해 탐지하는 방법으로써 전문가 시스템, 상태전이 분석, 패턴매칭 기법 등[7,8]이 있다. 현재 개발된 침입 탐지 시스템들은 대부분 위의 방법을 적용하거나 혼용된 방법을 사용하고 있다. 그러나 알려지지 않은 새로운 침입에 대해 실시간 추적이 가능한 침입탐지 시스템에 대한 연구들은 많이 진행되고 있으며, 잘 알려진 기법으로는 전자상거래 분야에서 많이 연구된 데이터 마이닝(Data Mining) 기법[9]이 있다. 데이터 마이닝은 정리되지 않은 다양한 데이터들에 대해 특정 용도에 부합하도록 가공하여 유용한 정보를 발견하기 위한 작업이다.

본 논문에서는 이러한 데이터 마이닝 기법 중 연관기법을 침입탐지 알고리즘에 적용하여 수집된 자료로부터 침입판정 정보를 생성할 수 있는 상태전이 알고리즘을 제시하고자 한다.

II. 관련연구

2.1 시스템 침입 유형

시스템 침입 유형은 크게 호스트 기반의 침입과 네크워크 기반의 침입으로 구분할 수 있다.

2.1.1 호스트 기반 침입

(1) 내부 서비스 거부 공격

시스템 내부 자원의 과도한 사용을 유도함으로

써 시스템의 서비스 속도를 떨어뜨리는 방법으로 예를들면 파일을 계속해서 open하여 사용 공간을 차지하게 하는 공격 방법 등이 있다. 이 외에도 한 사용자가 아주 많은 프로세스를 생성하여 CPU의 과부하로 인한 시스템 속도를 현저하게 떨어뜨리거나, 사용자가 매우 큰 파일을 생성하여 디스크 공간을 모두 사용하게 하는 디스크 공격 등이 있다.

(2) 시스템 내부 취약성

시스템에서 실행중인 여러 프로세스들이 가지고 있는 약점을 파악하여 이를 이용한 공격들이 여기에 해당된다. 예를들면 ps명령은 UNIX 시스템에서 프로세스들의 상태를 보여주는 root setuid 프로그램으로, 실행도중 /tmp/ps_data라는 임시파일을 생성한다. 이때 /tmp 디렉토리의 퍼미션에 sticky bit가 설정되어 있지 않을 경우 지워질 수 있는데, psrace란 프로그램으로 이 파일을 지우고 미리 만들어둔 쉘을 링크 시킴으로 소유자가 root인 setuid root 쉘을 가지게 된다.

이 외에도 프로그램 구현 시 버퍼의 한계 값을 검사하지 않는 함수를 사용하여 다른 스택의 영역을 이용한 버퍼 오버플로우 방식이나, 패스워드 변경시 생성되는 임시파일을 이용하여 공격하는 등의 다양한 방식[1,6]이 존재한다.

2.1.2 네트워크 기반 침입

(1) 외부 서비스 거부 공격

응용프로그램이나 네트워크 자원을 공격대상으로 속도를 느리게 하거나 서비스를 멈추게 하는 방법으로 finger 프로그램의 redirection 기능을 이용하여 근원지 주소를 속이거나, DNS 서비스를 방해하여 해당 서버로 네임서비스를 받는 호스트들의 네트워크 기능을 마비시킨다든지 하는 다양한 방법이 있다.

(2) 원격지 시스템 침입

네트워크 자원이나 응용 프로그램의 취약점을 공격대상으로 하여 접근권한을 취득하거나, 트로이 목마나 백도어 프로그램을 이용하여 시스템에 접근하는 방법이다. 이러한 공격에는 FTP 서버 공격과 Login 백도어 등이 있으며, 네트워크 상에서 전송되는 패킷을 훔쳐보는 스니퍼링, 시스템에 피해를 주는 바이러스 등이 있다.

2.2 침입 탐지 분석 기법

침입탐지 기법에는 크게 비정상적인 방법으로 통계학적 방법과 상태전이 방법 등이 있다.

2.2.1 통계학적 방법

침입탐지 시스템 개발에 있어서 가장 초기 방법 중의 하나인 통계학적 침입 탐지 방법[1]은 시스템 상에서 생성된 감사 자료(audit data)의 양과 형태의 변화를 측정하여 침입을 탐지한다. 통계학적 비정상 탐지는 임계값 탐지와 프로파일 기반 비정상 탐지로 분류할 수 있다. 임계값 탐지의 경우 시스템의 정상 동작 동안 일어날 이벤트 예상치가 지정한 양을 능가할 경우 이를 검출하는 것으로 MIDAS(Multics Intrusion Detection and Alerting System)[2], NADIR (Network Anomaly Detection and Intrusion Reporter)[3] 등이 있다. 프로파일 기반은 시스템 내에 있는 감사 로그들의 감시를 통해 침입을 탐지하는 방법으로 SRI의 IDES(Intrusion Detection Expert System)[4], MIDAS[2], NADIR[3], Haystack[5] 등이 있다.

2.2.2 상태 전이 기법

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 명령어 기반으로 단계별 불법 침입 분석기법에 많이 적용되고 있다. 침입 탐지 분석을 위해 각 상태에 대해 이벤트가 주어지면 다음 단계로 전이된다. 이러한 상태 전이 기법은 STAT(State Transition Analysis Tool)[6], NetSTAT[7] 등의 시스템에서 적용되고 있다.

2.3 연관 마이닝 생성 알고리즘

본 연구에서는 연관 기법에 의한 최고 빈도의 후보 명령 집합을 구한 후, 구해진 명령집합에 대해 다양한 유형의 침입 유형 데이터베이스와 비교하는 과정을 수행한다. 이러한 과정을 수행하기 위해 우선 필요한 과정이 빈발 항목 집합과 후보 항목 집합을 생성하는 것이 필요하다. 첫 번째 단계에서는 미리 결정된 최소 지지도인 $s_{\min}$ 이상의 트랜잭션 지지도를 가지는 항목집합들의 모든 집합들인 빈발 항목집합을 찾는다. 두 번째 단계에서는 빈발 항목집합을 사용하여 데이터베이스로부터 연관규칙을 생성한다. 이때 모든 빈발 항목집합 L에 대해 공집합을 제외한 L의 모든 부분집합을 찾는다. 잠재적인 빈발 항목집합들의 수는 모든 항목들의 멱집합(power set)의 크기와 같으며, 고려될 항목들의 크기에 대하여 기하급수적으로 증가한다. 따라서 모든 알고리즘들은 실제로 빈발한 항목들을 찾기 위해 후보라 지칭하는 빈발 가능성이 있는 항목집합들을 생성한 후, 데이터베이스를 읽어가면서 각 후보 항목집합들의 지지도를 계산한다. (그림 1)과 (그림 2)는 후보 항목과 빈발 항목 집합을 생성하는 예를 나타낸 것이며, (그림 3)은 이러한 수행에 대한 빈발 항목

집합 생성 알고리즘을 나타내고 있다.

TID	Items
100	A C D
200	B C E
300	A B C E
400	B E

그림 1 생성된 데이터 항목

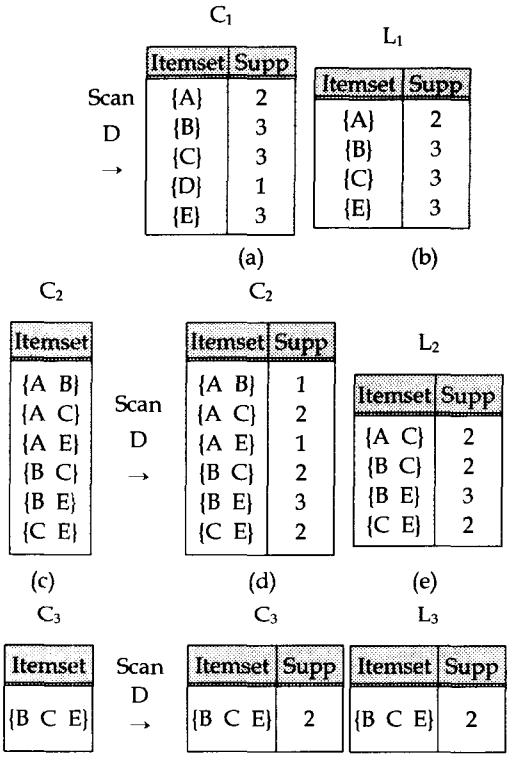

그림 2 후보 항목집합과 빈발 항목집합의 생성

1: L_1 = {large 1-itemsets}
2: **for** ($k=2; L_{k-1} \neq \emptyset; k++$) **do begin**
3: C_k=apriori-gen(L_{k-1}); //New candidates
4: **forall** transactions $t \in D$ **do begin**
5: C_t=subset(C_k, t);//Candidates contained in t
6: **forall** candidates $c \in C_t$ **do**
7: c.count++;
8: **end**
9: L_k = { $c \in C_k$ | c.count $\geq$ minsup}
10: **end**
11: Answer = $\bigcup_k L_k$;

그림 3 빈발 항목집합 생성 알고리즘

(그림 4)는 (그림 3)의 라인 3에 있는 apriori-gen()에 대한 호출 함수로 후보 항목집합을 생성하며, join 단계와 prune 단계로 구성된다. (그림 4)의 라인 2~4는 join 단계이로써 k번째 후보 항목집합을 생성하며, 라인 6~9는 prune 단계로써 생성된 후보 k-항목집합에서 필요없는 후보 항목집합을 삭제한다.

1: **Algorithm** Apriori-gen
2: **insert into** C_k // Join step
3: **select** $a.item_1, \cdots, a.item_{k-1}, b.item_{k-1}$
4: **from** $L_{k-1}a, L_{k-1}b$
5: **where** $a.item_1 = b.item_1, \cdots, a.item_{k-2} = b.item_{k-2}, a.item_{k-1} < b.item_{k-1}$;
//Prune step: now prune rules with subsets missing in L_{k-1}
6: **forall** itemset $c \in C_k$ **do**
7: **forall** (k-1)-subsets s of c **do**
8: **if** ($s \notin L_{k-1}$) **then**
9: **delete** c from C_k ;

그림 4 후보 항목집합 생성 알고리즘

III. 침입 시나리오 자동생성 알고리즘

본 논문에서 제안한 침입 시나리오 자동생성 알고리즘인 AGAPS(Automated Generation Algorithm of the Penetration Scenarios)는 State Description Table(SDT)을 (그림 5)의 수정된 Apriori 연관 알고리즘의 입력으로 사용하며, 그 출력인 후보 침입 시나리오는 후보 SDT인 CSDT (Candidate SDT)와 후보 rule chain에 저장된다. 후보 rule chain의 경우 STAT의 rule chain과 동일하다. CSDT는 후보 침입 시나리오가 발생하면 값이 증가되는 SDT의 구조와 동일하다. 그리고 Count 항목값 N이 설정된 임계값 이상이면 후보 침입 시나리오인 CP_1은 SDT와 rule chain에 저장된다.

1: L_1 = {SDT}
2: **for** ($k=2; L_{k-1} \neq \emptyset; k++$) **do begin**
3: $CSDT_k$ = apriori-gen(L_{k-1});
4: **forall** transactions $t \in SDT$ **do begin**
5: $CSDT_t$ = subset($CSDT_k$, t);
6: **forall** candidates $c \in C_t$ **do**
7: c.count++;
8: **end**
9: L_k={ $c \in C_k$ | c.count $\geq$ minsup}
10: **end**
11: Answer = $\bigcup_k L_k$;

그림 5 수정된 Apriori 알고리즘

1: **Algorithm** Modified Apriori-gen
2: **insert into** $CSDT_k$
3: **select** $a.item_1, \cdots, a.item_{k-1}, b.item_{k-1}$
4: **from** SDT
5: **where** $a.item_1 = b.item_1, \cdots, a.item_{k-2} = b.item_{k-2}, a.item_{k-1} \neq b.item_{k-1}$;

6: **forall** itemset $c \in CSDT_k$ **do**
7: **forall** (k-1)-subsets s of c **do**
8: **if** ($s \notin L_{k-1}$) **then**
9: **delete** c from $CSDT_k$

그림 6 수정된 Apriori-gen 알고리즘

Ⅳ. 비교 분석 및 평가

본 논문에서는 구현하기가 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 분석해보면 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 유사한 혹은 알려지지 않은 불법 침입에 대해 새로운 침입 유형을 분석하기 위해서는 연관 데이터 마이닝 기법을 적용하여 새로운 침입 탐지 분석 알고리즘을 제시하고자 하였다. [표 1]은 기존의 연구된 침입 탐지 시스템 유형들과 비교하고 있다.

[표 1] 기존 침입 탐지 시스템과의 비교

침입탐지 시스템	자료 수집원			침입 모델		특성
	호스트	멀티호스트	네트워크	비정상 탐지	오용 탐지	
IDES		o		o	o	전문가
STAT	o				o	상태전이, 규칙기반
MIDAS	o			o	o	전문가
NADIR			o	o	o	전문가/규칙
본 논문	o		o		o	상태전이, 연관마이닝

Ⅳ. 결 론

본 논문에서는 불법 침입 추론 정보를 생성하기 위해 오용 탐지 모델에 기반을 두고 상태 전이 분석과 연관 데이터 마이닝 기법을 적용하였다. 이를 위해 명령어 상태 전이 기법에 연관 규칙을 적용하여 새로운 침입 탐지 유형을 판정하는 알고리즘을 수정 제안하였다. 기존 연구들이 대부분 상태전이와 규칙 기반, 혹은 전문가 시스템과 규칙 기반을 적용한데 반해, 본 연구에서는 기존의 상태 전이 기법과 유사하지만 이러한 상태 테이블을 기반으로 연관 데이터 마이닝 기법을 응용한 새로운 접근 방법을 시도하였다.

그리고 이미 알려진 침입에 대해서는 상태 전이 분석 기법에 의해 실시간으로 탐지가 될 수 있으며, 유사하거나 새로운 유형의 침입에 대해서는 연관 기법의 최종 단계에 있는 빈도 항목 집합과 후보 항목 명령 집합을 이용하여 이미 알려진 침입 유형과 비교하여 침입 유형 판정이 가능함을 알 수 있다.

참고문헌

[1] 한국정보보호센터, 정보시스템 침해사고방지기술('97), pp.163-168, Jan. 1998

[2] 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

[3] B. Hubbard, T. Haley, N. McAuliffe, L. Schaefer, N. Kelem, D. Wolcott, R. Feiertag and M. Schaefer, "Computer System Intrusion Detection", Trusted Information Systems, Inc.,Final Technical Report, Dec, 1990

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

[5] K. Chen, S.C. Lu and H.S. Teng, "Adaptive Real-Time Anomaly Detection Using Inductively Generated Sequential patterns", presented at the Fifth Intrusion Detection Workshop, SRI Internation, Menlo Park, CA, May 1990

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

[7] G. Vigna and R. Kemmerer, "NetSTAT: A network-based intrusion detection approach", Proceedings of the 14th Annual Computer Security Applications Conference, Scottsdale, Arizona, Dec. 1998

[8] Berson A, Smith S, and Thearling. K, "Buolding Data Miniing Applications for CRM", Mc-Hill, 1999

[9] Michael G., and Gruenwald L., "A Suvery of Data Mining and Knowledge Discovery Software Tools", SIGMKDD Explanations, Vol. 3, Issue2, 2000. pp.20-33.