情報保護學會論文誌
第13券 第4號, 2003. 8

과탐지를 제어하는 이상행위 탐지 방법

조 혁 현*, 정 희 택*, 김 민 수**, 노 봉 남***

Anomaly Detection Method Based on The False-Positive Control

Hyug-Hyun Cho*, Hee-Taek Ceong*, Min-Soo Kim**, Bong-Nam Noh***

요 약

인터넷이 일반화되면서, 컴퓨터 시스템을 침입으로부터 효과적이면서 종합적으로 보호하기 위해 침입 탐지 시스템이 필요하게 되었다. 본 연구에서는 이상행위 탐지 기법을 이용한 침입 탐지 시스템을 구축할 때, 수행하는 정상행위 프로파일링 과정에서 발생하는 자기설명모순이 존재함을 제시하고 이를 제어할 수 있는 침입 탐지 방안을 제안하였다. 또한, 연관규칙을 적용한 프로파일링 과정의 결과는, 많은 정상행위 패턴이 생성될 수 있기 때문에, 이를 위해 군집화를 통한 효과적인 적용방안을 제시한다. 마지막으로, 사용자의 행위 패턴에 대해 군집화된 정상행위 패턴 데이터베이스로부터 이상행위 여부를 판단할 수 있는 유사도 함수를 제안하였다.

ABSTRACT

Internet as being generalized, intrusion detection system is needed to protect computer system from intrusions synthetically. We propose an intrusion detection method to identify and control the contradiction on self-explanation that happen at profiling process of anomaly detection methodology. Because many patterns can be created on profiling process with association method, we present effective application plan through clustering for rules. Finally, we propose similarity function to decide whether anomaly action or not for user pattern using clustered pattern database.

keyword : *침입 탐지 시스템(Intrusion detection system), 과탐지(False-positive), 프로파일링(Profiling), 연관규칙 탐사(Association rule mining), 군집화(Clustering)*

Ⅰ. 서 론

인터넷이 일반화되면서, 많은 시스템들이 정보공유를 위한 유용한 수단이 되지만, 아울러 정보 유출, 전산망 침해 등과 같은 부작용이 늘고 있다. 이러한 부작용을 막고 신뢰할 수 있는 컴퓨팅 환경을 마련하기 위한 한 방안으로써, 침입 탐지 시스템이 제안 구현되고 있다. 침입이란 컴퓨팅 자원의 무결성, 기밀성 그리고 활용성을 저해하는 임의의 행위 집합을 의미한다.[1,2,3,4,5,6] 이러한 침입 자체만을 막는 것은 시스템의 설계나 프로그래밍 과정에 포함된 오류와 더욱 복잡해지는 시스템 구성 때문에 충분하지 않다. 즉, 컴퓨터 시스템을 침입으로부터 효과적이면서 종합적으로 보호하기 위해 침입 탐지 시스템이 필요하다.

침입탐지 시스템은 오용탐지(misuse detection)와 이상행위 탐지(anomaly detection) 방법으로 구분한다. 오용 탐지는 알려진 침입 정보를 축적하고, 이를 기반으로 임의의 행위 집합에 대해 침입 여부를 결정

* 여수대학교 정보기술학부(hhcho@yosu.ac.kr, htceong@yosu.ac.kr)
** 전남대학교 정보보호협동과정 객원교수(phoenix@athena.chonnam.ac.kr)
*** 전남대학교 컴퓨터정보학부(bongnam@chonnam.ac.kr)

하는 기법이다. 오용 탐지는 전문가 시스템이나 상태 전이 분석 기법을 적용하여 구현한 시스템이 있으며, 잘 알려진 침입에 대해 성능이 좋으나, 변종 침입 패턴을 탐지 할 수 없는 단점이 있다.[3] 이상 행위 탐지는 정상적인 행위 정보를 축적하고, 이를 기반으로 임의의 행위가 정상 행위와 다름을 기준으로 침입여부를 결정하는 기법이다. 이상행위 탐지는 통계적 접근 방법이나, 뉴럴 네트워크 기법을 적용하여 구현한 시스템이 있으며, 알려지지 않는 침입을 탐지 할 수 있으나, 잘 알려진 침입에 대해 성능이 좋지 않다.[3] 오용 탐지 및 이상행위 탐지 방법이 갖는 이러한 단점들은, 기반으로 하고 있는 침입정보 및 정상행위 정보의 불완전성 문제에 기인한다. 침입 정보의 불완전성이 침입행위 패턴을 탐지하지 못하는 오탐지(false-negative) 오류를, 정상행위 정보의 불완전성이 정상행의 패턴을 침입으로 간주하는 과탐지(false-positive) 오류를 야기한다.

본 연구에서는 불완전성 문제를 제어 할 수 있는 침입 탐지 시스템을 제안한다. 완전한 정상행위 및 침입 패턴 구축을 통한 완전한 침입 탐지 시스템 구축은 불가능하다. 현재 이루어지는 모든 침입 패턴을 알 수 없을 뿐만 아니라 새로운 변종들이 생성되고 있기 때문이다. 그러나 침입 탐지 시스템을 구축하는 과정에서 발생하는 내재적 오류 가능성을 최소화함으로써, 알려진 정보를 기반으로 점진적으로 완전한 지식정보를 구축할 수 있다. 이를 위해 먼저 과탐지 오류의 발생이유를 실험을 통해 제시한다. 즉, 정상행위 정보를 생성하는 프로파일(profile) 생성과정에서 내재하고 있는 이런 오류의 발생 이유를 제시하고 이를 해결할 수 있는 방안을 제시한다. 생성된 프로파일 정보로부터 침입 탐사를 효과적으로 수행하기 위해, 프로파일 정보의 군집화(Clustering)를 적용하며, 이를 기반으로 침입 탐지과정에서 효과적인 역할을 수행할 수 있음을 제시한다.

본 논문의 구성은 2장에서 침입 탐지 시스템의 대상이 되는 침입에 대한 특성을 분석하고, 프로파일 생성과정 및 내재한 모순을 제시한다. 이러한 모순을 제거하기 위한 방안을 제시한다. 3장에서는 앞서 제안한 기법을 기반으로 군집화한 패턴 데이터베이스를 이용하여 침입 탐지 방안을 제시한다. 4장에서는 본 연구와 관련된 기존 연구들을 분석한다. 마지막으로, 5장에서는 본 연구의 결론 및 진행 중인 연구에 대해 제시한다.

II. 자기설명모순이 최소화된 프로파일 생성

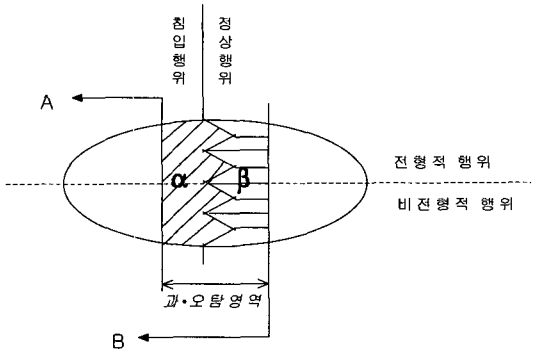
시스템 상에서 이루어지는 사용자의 행위는 전형적(normal) 행위와 비 전형적(abnomal) 행위로 구분할 수 있다. 비 전형적 행위로는 계정이 없는 사용자가 일반 사용자의 자원을 접근하는 것과 같은 것이다. 그러나 침입이 이러한 비 전형적 행위들로만 구성되지는 않을 수 있다. 이를 다음 [그림 1]과 같이 구분할 수 있다.

또한 사용자의 행위는 침입 행위와 정상 행위로 구분할 수 있다. 이러한 행위에 대해 침입 탐지 시스템에서 탐지된 결과는 A와 B 영역으로 구분할 수 있다. A영역 결과는 α영역에 해당하는 침입 행위를 발견하지 못한 경우이며, B영역 결과는 β영역에 해당하는 정상행위를 침입 행위로 간주하는 결과를 야기한다. 물론, 침입 탐지 시스템의 목적은 침입 행위만을 탐지하는 것이 목적이어야 한다. 그러나, 현재 이루어지는 모든 침입 패턴을 알 수 없을 뿐만 아니라 새로운 변종 패턴들이 지속적으로 생성되고 있기 때문에 한계를 갖는다. 이런 이유로 기존에 개발된 침입 탐지 시스템은 A 및 B 영역으로 구분할 수 있는 결과를 생성한다. α영역은 침입 탐지 시스템에 의해 침입행위를 정상행위로 간주하는 오탐지를 의미하며, β영역은 정상행위임에도 침입으로 구분하는 과탐지를 의미한다. 본 연구에서는 α와 β영역을 다음과 같이 과·오탐지 영역이라 정의한다.

정의 1 과·오탐지 영역

침입 탐지 시스템에서 정상행위를 침입으로 간주하거나 침입을 정상행위로 간주하는 영역.

과·오탐지는 알려지지 않는 침입 패턴이나, 정상행위를 학습할 때 정상행위로 모델링 되지 못한 정



〔그림 1〕 침입 탐지 시스템에서 행위

보로써, 침입 탐지 시스템에게 알려지지 않은(unknown) 정보 때문에 발생한다. 본 연구에서는 이상행위 탐지 기법을 기반으로 과탐지를 최소화하기 위한 방안을 제안한다.

2.1 프로파일 생성 및 자기설명모순

이상행위 탐지 방법을 이용한 침입 탐지 시스템에서 기초가 되는 정보는 사용자의 정상행위 패턴이다. 정상행위를 모델링하기 위한 과정인 프로파일링(profiling)은 정상행위들로만 구성된 자료로부터 트리 구조나 규칙 집합과 같은 정상행위 모델을 생성하고 이를 정상행위 패턴 데이터베이스에 유지한다. 새로운 사용자의 행위에 대해 패턴 데이터베이스의 임의의 패턴과 유사도를 측정하여 정상 패턴으로부터 어느 정도 벗어났는지를 판별함으로써, 침입 여부를 구분한다.

정상행위 패턴 데이터베이스를 구축할 때, 모델링되지 않은 정상행위에 의해 발생하는 과탐지 오류는 외부 및 내부 원인에 의해 발생한다. 외부 원인은 프로파일링 과정에 필요한 정상행위 데이터가 불완전함으로써 발생한다. 즉, 정상행위 데이터에 모든 가능한 정상행위를 포함하지 못할 수 있다. 둘째로 내부 원인은 정상행위 데이터로부터 생성된 패턴 데이터베이스가 입력 데이터의 모든 정보를 포함하지 못함으로써 발생한다. 이를 다음과 같이 외부 데이터 불완전성 및 내부 데이터 불완전성이라 정의한다.

정의 2 외부 데이터 불완전성

프로파일링 과정의 입력 데이터는 모든 정상행위 정보를 포함하지 못함.

정의 3 내부 데이터 불완전성

프로파일링 과정의 입력 데이터가 포함하고 있는 모든 정상행위 정보를 패턴 데이터베이스로 생성하지 못함.

외부 데이터의 불완전성은 얼마나 많은 정상행위 정보를 획득할 수 있느냐의 문제로 본 연구에서는 연구대상에서 제외하고, 내부 데이터의 불완전성을 보이기 위해 연관규칙 생성 알고리즘[7,8]을 적용하여 프로파일링 실험을 수행하였다. 실험에 사용한 데이터는 1999년 DARPA Intrusion Detection Evaluation Data Set을 이용하여 정상행위를 프로파일링 하기 위해 공격행위가 포함되지 않은 1주와 3주 데이터를 실험 데이터로 사용하여 프로파일링 하였다. 연관 규칙 생성 알고리즘을 사용한 이유는 지지도와 신뢰도를 통해 다양한 특성을 지정할 수 있고 많은 데이터 집합에 대해 효과적으로 규칙을 생성할 수 있기 때문에 적용하였다. 실험에서 15% 지지도, 40% 신뢰도, 사용자 명령어 길이 6을 기반으로 수행하였다. 정상 및 이상행위 판정을 위해 명령어의 일치 여부를 기준으로 유사도 함수를 정의하였다. 정의한 유사도 함수를 이용하여 이상 행위도는 1-*Similarity*(P,UP^k)에 의해 산출한다.

정의 4 유사도 함수

명령어 집합 $C=\{c_1, c_2, \ldots, c_z\}$, 사용자 행위 패턴 $UP^k=\{c_i^k, c_{i+1}^k, \ldots, c_j^k\}$, 패턴 데이터베이스 $P=\{p_1, p_2, \ldots, p_m\}$, $p_g=\{c_s, c_{s+1}, \ldots, c_t\}$에 대해 패턴 UP^k의 유사도 함수 *Similarity*(P,UP^k)는 다음과 같다. $1\le i,j,s,t\le z$, $1\le g\le m$.

$$Similarity(P,UP^k)=\max_{g=1}^{m}(w_1 FCF_{UP^k}\times w_2 SCF_{UP^k})$$

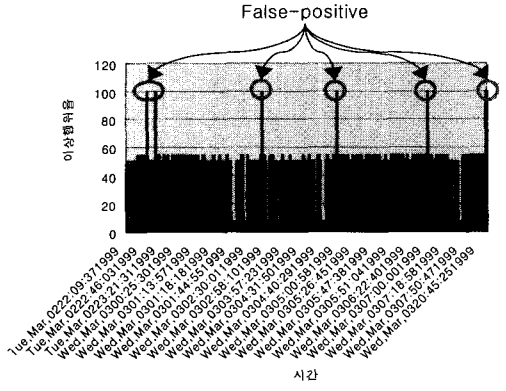
여기서 w는 가중치로서 $w_1+w_2=1$이다. FCF_{UP^k}(Frequent Count Factor)는 명령어의 반복 빈도를 고려한 요소이며, SCF_{UP^k}(Support Confidence Factor)는지지도와 신뢰도를 고려한 요소로 이들 간의 논리곱으로 표현한다. 각각은 다음과 같다. 패턴 데이터베이스에 있는 각 패턴과 사용자 행위 패턴간에 동일한 명령어가 반복되는 경우는 *Frequentcount*(p_g, UP^k)로, 그렇지 않는 경우는 *MatchCount*(p_g, UP^k)로, 명령어 개수를 의미한다. *SupportValue*(p_g, UP^k)와 *Confidence Value*(p_g, UP^k)는 패턴 데이터베이스에 있는 각 패턴과 사용자 패턴간에 $p_g \cap UP^k \neq \varnothing$ 일때, 해당하는 지지도와 신뢰도이다. *Length*(UP^k)는 사용자 행위 패턴의 길이를 의미한다.

$$FCF_{UP^k}=\begin{cases}\dfrac{MatchCount(p_g,UP^k)}{Length(UP^k)} & \text{명령어가 반복하지 않는 경우}\\[2ex] \dfrac{FrequentCount(p_g,UP^k)}{Length(UP^k)} & \text{명령어가 반복하는 경우}\end{cases}$$

$$SCF_{UP^k}=\frac{Length(UP^k)}{SupportValue(p_g,UP^k)+ConfidenceValue(p_g,UP^k)}$$

유사도 함수에서 SCF_{UP^k}는 패턴 간에 단순히 지지도만을 고려하지 않고 연관성을 의미하는 신뢰도를 고려하였다. 낮은 신뢰도를 갖고 전체 순서(total order) 관계가 없는 패턴이지만, 각 사건 간에 존재하는 높은 연관성을 지정할 수 있다. 예를 들어 사용자의 패턴이 {A,B,C,D}이고 이와 관련된 패턴 데이터베이스에 {A -> B : (0.7, 0.8)}, {A,B -> C : (0.5, 0.7)}, {A,C -> B : (0.5, 0.4)}, {A,B,C,E,F -> D : (0.1, 0.2)} -{규칙 : (지지도, 신뢰도)}- 이 존재 할 때, A,B,C,E,F -> D인 패턴에 대해 높은 유사도가 계산된다. 이는 A,B,C,E,F -> D가 낮은 지지도와 신뢰도를 갖지만, {A,B,C,D,E,F}의 모든 부분 집합이 패턴으로서 존재함을 내재하고 있기 때문이다. 즉, 각 사건 간에 많은 연관성 정보를 포함하고 있기 때문이다. 이런 이유로 반비례 계산을 채용하였고, FCF_{UP^k}와 SCF_{UP^k}간에 논리곱으로 정의한 이유는 패턴의 많은 부분이 일치하면서 일치한 패턴 간에 사용자 패턴과 패턴 데이터베이스의 패턴들 간에 다양한 요소를 고려한 유사도를 계산하기 위함이다.

정상행위 데이터에 대한 프로파일링에 의해 생성된 정상행위 규칙 집합을 이용하여, 정상행위만을 포함한 1주 수요일 데이터를 탐지하였을 때, 다음 [그림 2]와 같이 과탐지 오류가 발생한다.



[그림 2] 과탐지 오류

프로파일링 과정에서 사용한 정상행위 데이터로부터 생성된 정상행위 집합을 이용하여, 정상행위 데이터를 탐사 했을 때, 이상행위로 탐사되는 것을 다음과 같이 자기설명모순(Contradiction on Self-Explanation)으로 정의한다. 자신의 데이터로부터 추출된 정보를 근간으로 자신을 탐사 했을 때, 이상행위가 발견됨을 의미한다.

정의 5 자기설명모순

정상행위 집합을 생성하기 위해 사용된 정상행위 데이터를 대상으로 침입 탐사를 수행 했을 때, 이상행위가 발견되면 이를 자기설명모순이라 한다.

자기설명모순에 의한 이상행위는 침입을 의미하지 않는다. 이러한 모순의 발생원인은 자신의 정보를 완전히 포함하지 못하는 내부 데이터 불완전성이 존재하기 때문이다. 즉, 연관 규칙 생성 알고리즘에서 지지도와 신뢰도 값에 의해 임계 값 이하의 패턴을 프로파일링 과정에서 생성하지 않기 때문이다.

2.2 자기설명모순의 제어

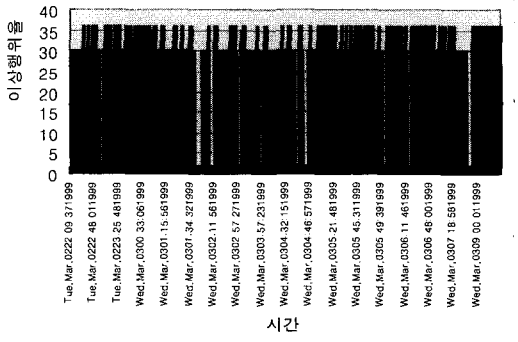
자기설명모순을 해결하기 위한 두 가지 방법이 존재한다. 첫째, 직관적인 방법으로 정상행위 데이터의 모든 정보를 정상행위 패턴으로 정의한다. 이는 무수히 많은 패턴 정보를 생성하게 된다. 이로 인해 침입 탐지 시스템의 성능을 저하하게 하고 실시간 탐지를 불가능하게 할 수 있다. 둘째는 자기설명모순을 허용하되 적절한 임계 값을 설정하게 하고, 이를 만족하는 정상행위 패턴을 생성하도록 한다. 즉, 각 항목에 대한 지지도와 신뢰도를 전문가에 의해 지정하게 하지 않고 일정 임계 값 이상을 만족하는 지지도와 신뢰도를 자동으로 결정하게 함으로써, 자기설명모순을 최소화한다.

사용자에 의해 정의된 임계 값 이하의 자기설명모순을 허용하기 위해, 즉 임계 값 이하의 과탐지 오류를 허용하기 위해 다음과 같은 알고리즘에 의해 프로파일링 과정을 수행한다. 다음 알고리즘에 의해 적절한 지지도와 신뢰도를 결정하게 되고, 이를 만족하는 정상행위 패턴은 [그림 2]와 같이 이상행위율 100%로 탐지되는 오류를 [그림 3]과 같이 제어할 수 있다. 자기설명 모순 제어 임계 값은 허용할 수 있는 이상 행위율을 의미한다.

제시한 알고리즘에서 초기 지지도와 신뢰도를 100%에서 감소하도록 지정하였으나 실험을 통해 50%에서 시작할 때 적절한 성능을 얻을 수 있었다. 물론, 사용자 정상행위 데이터 집합의 크기에 종속되지만, 연관 규칙 탐사 방법 중 DHP[8] 방법을 채용함으로써 수행시간을 최소화 할 수 있다.

```
Algorithm 자기 설명 모순 제어
입력 : 사용자 정상행위 데이터 ND
       자기설명 모순 제어 임계값 CSECV
출력 : 정상행위 패턴 NP

Find = 0;
for(Minsupport=100;Minsupport>0;Minsupport--){
 for(Minconfidence=100;Minconfidence>0;  Minconfidence--){
  NP=AssociationRuleMining(ND,Minsupport,Minconfidence);
  if( NP ==  ∅ ) continue;
  MaxAnomalDegree = Detection(NP,ND);
  if(CSECV >= MaxAnomalDegree ) {
    Find = 1;
    break;
  }
 }
 if( Find == 1 ) break;
}
```



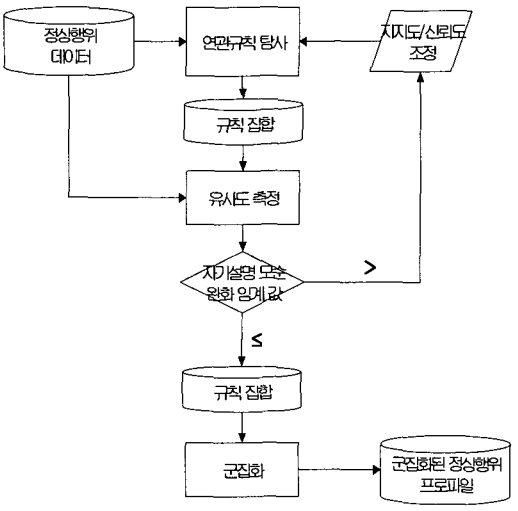
〔그림 3〕 과탐지 오류의 제어

앞에서 제시한 실험 데이터를 자기설명 모순 제어 임계값으로 40%를 지정했을 때, 지지도 11, 신뢰도 32때, 다음 [그림 3]과 같은 실험 결과를 생성할 수 있었다.

2.3 정상행위 프로파일 생성과정

앞에서 제시한 과정에 의해 최종 정상행위 프로파일 생성과정은 다음 [그림 4]와 같다. 정상행위 데이터로부터 연관규칙 알고리즘에 의해 규칙 집합을 생성한다. 다음으로, 생성된 규칙과 정상행위 데이터 간의 유사도 측정을 수행하여 자기설명모순 제어 임계 값 이하를 만족하는지 검사한다. 이를 만족하지 않으면 지지도와 신뢰도를 재조정하여 규칙 생성을 반복 수행한다.

마지막으로, 임계값 이하의 규칙 집합에 대해 군



〔그림 4〕 프로파일 과정

집화를 통해 군집화된 정상행위 프로파일을 생성한다. 이는, 연관 규칙 탐사에 의해 생성된 정상행위 패턴은 낮은 지지도와 신뢰도에 의해 많은 규칙을 생성할 수 있고 자기설명모순을 제어하기 위한 과정에서 많은 정상행위 모델을 생성할 수 있기 때문이다. 본 연구에서 실험한 데이터는 데이터 집합의 크기가 크기 때문에 적은 수의 규칙-[그림 3]의 결과에서는 23개의 규칙-을 생성하지만, [9,10]에서 제시한 것처럼 많은 규칙생성이 일반적이다. 많은 규칙 집합은 사용자의 행위가 침입인지를 판단하는데 많은 자원과 시간을 요하게 한다. 본 연구에서는 많은 규칙을 군집화하기 위한 방안으로 군집화 기법을 채용한다.

본 연구에서는 군집화 방안으로 SOM(Self- Organizing feature Maps)[11] 방법을 채용한다. 이는 SOM이 구조상 상당히 빠른 모델이며, 훈련이 빠르고 자기 조직화를 통해 정확한 통계적 모델을 생성할 수 있기 때문이다. 임계 값 이하의 과탐지를 만족하는 규칙 집합에 대해 사용자 명령을 기준으로 군집화 한다. 유사한 작업을 하는 사용자 그룹의 군집이 생성되면 이를 정상 행위 프로파일로 저장한다.

Ⅲ. 침입 탐지 방안

정상행위 프로파일 정보를 기반으로 사용자의 행위를 다음과 같은 유사도 측정에 의해 이상행위 여부를 판별한다. 군집화된 정상행위 프로파일과 사용자 행위 간의 유사도를 계산하기 위해 군집의 무게중심(centroid) 및 유사도함수를 다음과 같이 정의한다.

정의 6. 군집 C의 무게중심

$$Centroid(C) = \{ \langle p_g, \tau_{Centroid(C)}(p_g), \gamma_{Centroid(C)}(p_g) \rangle \mid p_g \in P,$$
$$\tau_{Centroid(C)}(p_g) = \max_{g=1}^{m}(SupportValue(p_g)),$$
$$\gamma_{Centroid(C)}(p_g) = \max_{g=1}^{m}(ConfidenceValue(p_g))\}$$

$SupportValue(p_g)$와 $ConfidenceValue(p_g)$는 동일 패턴 집합의 지지도와 신뢰도를 의미한다.

예제 1. [표 1]과 같이 사용자 패턴들이 하나로 군집 되었다고 가정하면 다음과 같은 무게 중심을 생성한다.

〔표 1〕 사용자패턴

사용자	패턴 ID	패턴	지지도 (%)	신뢰도 (%)
U1	P1	cd -> vi	14	34
U1	P2	vi -> cp	15	12
U2	P1	cd -> vi	17	20
U2	P3	cd, ls -> vi	19	34
U3	P1	cd -> vi	12	55
U3	P4	cp -> vi	15	22

Centroid(C)={<cd->vi,0.17,0.55>,<vi->cp,0.15,0.12>,<cd,ls->vi,0.19,0.34>, <cp->vi,0.15,0.22>}

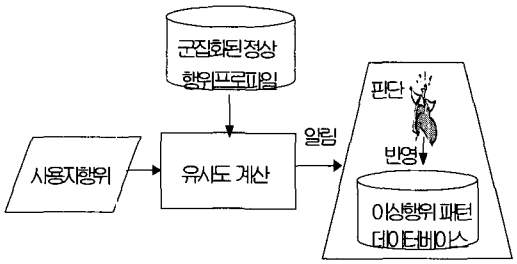
사용자의 행위 패턴에 대한 이상 행위 여부를 판단하기 위한 정의 4의 유사도 함수를 확장한 군집의 무게중심에 대한 유사도는 다음과 같다.

정의 7. 사용자 행위 패턴 UP^k와 군집 C의 무게중심간의 유사도

$$SimilarityOnCluster(C, UP^k) = \frac{\sum_{\forall sUP^k \in \phi(UP^k)} Similarity(C, sUP^k)}{\sum_{g=1}^{m}(SupportValue(p_g) + ConfidenceValue(p_g))}$$

여기서 $\phi(UP^k)$는 UP^k을 구성하는 사건들의 모든 부 서열 들의 집합이다.

예제 2. 예제 1에서 제시한 군집과 사용자 행위 패턴 {cd,ls,telnet}에 대한 유사도는 다음과 같이 계산한다.



〔그림 5〕 이상행위 탐지

$$SimilarityOnCluster(C, cd, ls, telnet) = \frac{0.33+0.94}{(0.17+0.55+0.15+0.12+0.19+0.34+0.15+0.22)} = 0.67$$

이상행위 탐지는 군집화된 정상행위 프로파일 정보와 사용자 행위 패턴간의 유사도 계산을 통해 [그림 5]와 같이 이상행위를 탐지한다. 탐지된 이상행위는 관리자에게 통보되고 이상행위 패턴 데이터베이스에 반영된다.

Ⅳ. 관련 연구

침입 탐지 시스템에 관한 연구들은 활발히 이루어지고 있다. 이장에서는 제안된 방안과 비교하여 본 연구와 관련된 기존 연구들을 두 가지 관점에서 분석한다. 첫째는 자기설명모순 관점에서 과탐지 오류를 고려하고 있느냐를 기준으로 하였다. 두 번째는 과·오탐지에 해당하는 알려지지 않은 영역으로부터 정보를 추출하는 방안을 기준으로 하였다. 전자는 침입 탐지 시스템에서 기반 지식이 되는 정보를 어느 정도까지 학습할 것이냐는 문제이다. 후자는 알려지지 않는 정보를 발견하기 위해 데이터 탐사 방법을 채용한 침입 탐지 시스템들을 분석 대상으로 하였다.

먼저, 침입 탐지 시스템을 학습하기 위한 방안으로 신경망,[12] 인공지능 기반,[13] 면역학 기반,[14] 데이터 탐사 기반[15,16,17,18,19,20] 등의 방법을 적용하여 다양한 기계 학습 방안을 제안하고 있다. 그러나 기존 연구들은 정상 행위를 모델링하기 위한 방안만을 제안하였을 뿐, 본 연구에서 제안한 자기설명모순을 고려하고 있지 않다. 또한 이를 제어하기 위한 방안을 고려하고 있지 않다. 한편 [15,16,17,18]에서는 데이터 탐사 기법을 적용한 침입 탐지 시스템을 제안하였고 어느 정도까지 학습해야 하는지를 제시하였다. 즉, [16]에서 제시한 학습 정도의 구분은 지속적인 규칙

생성을 통해 더 이상 새로운 규칙이 생성되지 않을 때를 기준으로 하였다. 이러한 방법은 지속적인 실험을 야기하고 본 연구에서 제안한 자기설명모순을 구분하지 못하였으며, 이러한 모순의 제어 방안도 고려하지 않고 있다.

다음으로 데이터 탐사 방법을 채용한 침입 탐사 시스템은 연관 규칙, 순차 패턴, 군집화, 분류 기법 등을 채용하였다. 특히, [15,16,17,18]에서는 연관 규칙과 빈발 에피소드(episode) 방법[21]을 채용하였다. 제안된 방안은, 연관 규칙에 의해 생성되는 많은 규칙에 대해, 특정 속성을 지닌 규칙만을 생성하거나, 공통 속성을 단순화하는 방안으로 데이터 탐사 방법을 응용하였다. 더욱이 지지도는 낮지만 의미 있는 규칙을 생성하기 위해, 빈발 항목 생성과정에서 지지도를 1/2씩 감소시키는 임의적인 방안을 제안하고 있다. 그러나, 제안된 방안은 본 연구에서 채용한 지지도와 신뢰도를 고려한 종합적인 유사도 계산 방안을 제안하고 있지 못하며, 특정 속성을 포함한 규칙을 생성한다 하더라도 지속적인 학습을 통해 많은 규칙이 생성되게 되는 문제가 있다. 또한, 어떤 속성을 포함하고 있는 규칙을 생성할 것인가와 관련된 속성 선택 문제가 존재한다. [19,20]에서도 데이터베이스 로그와 네트워크 패킷 데이터에 연관 규칙 기법을 적용한 침입 탐지 시스템을 제안하였으나, 본 연구에서 제안한 자기설명모순을 고려하고 있지 않으며, 많은 규칙의 생성에 의한 문제를 고려하고 있지 않다.

V. 결론 및 향후 연구

인터넷이 일반화되면서, 컴퓨터 시스템을 침입으로부터 효과적이면서 종합적으로 보호하기 위해 침입 탐지 시스템이 필요하게 되었다. 본 연구에서는 이상행위 탐지 기법을 이용한 침입 탐지 시스템을 구축할 때, 수행하는 정상행위 프로파일링 과정에서 발생하는 자기설명모순이 존재함을 제시하고 이를 제어할 수 있는 침입 탐지 방안을 제안하였다.

본 연구에서는 먼저, 이상행위 탐지 기법을 이용한 침입 탐지 시스템을 구축할 때, 수행하는 정상행위 프로파일링 과정에서 발생하는 자기설명모순이 존재함을 제시하였다. 프로파일링 과정에 사용되는 정상행위 데이터에 대해 이상행위를 탐지 할 때, 이상행위가 존재함을 제시하였고 이것의 원인을 제시하였다. 둘째, 이를 제어 할 수 있는 방안을 제안하였다. 사용자가 지정한 임계값을 기준으로 그 이하의 자기설명모순을 갖는 프로파일 생성 방안을 제안하였다. 셋째, 연관규칙을 적용한 프로파일링 과정의 결과는 많은 정상행위 패턴이 생성될 수 있기 때문에, 군집화를 통한 효과적인 적용방안을 제시하였다. 마지막으로, 사용자의 행위 패턴에 대해 군집화된 정상행위 패턴 데이터베이스로부터 이상행위 여부를 판단할 수 있는 유사도 함수를 제안하였다. 제안한 군집에 대한 유사도 함수는 정상행위 패턴이 갖는 지지도 및 신뢰도를 종합적으로 고려한 방안이다.

현재 동적인 프로파일 생성 및 유지할 수 있는 효율적인 방안을 연구 중에 있으며, 이상행위로 결정된 패턴을 모델링하기 위한 방안 및 이러한 패턴의 동적 유지 방안에 대해 연구하고 있다.

참 고 문 헌

[1] H. Debar, M. Dacier, and A. Wespi, "Towards a Taxonomy of Intru sion-Detection Systems", *Research Report of IBM Research Division, Zurich Research Lab.*, Jan. 1998.

[2] R. Heady, G. Luger, A. Maccabe, and M. Servilla, "The architecture of a network level intrusion detection systems", *Technical Report, Computer Science Dept. Univ. of New Mexico*, Aug. 1990.

[3] R. G. Bace, *Intrusion Detection*, Mac Millan Technical Pub., 2000.

[4] Denning, D. E, "An Intrusion-Detection Model", *IEEE Transaction on Software Engineering*, 13, pp.222~232, 1987.

[5] C. Kahn, P. A. Porras, S.Staniford-Chen, and B. Tung, "A Common Intrusion Detection Framework", 1998.

[6] A. K. Gjosh, J. Wanken and F. Charron, "Detection Anomalous and Unknown Intrusions Against Programs", *In Proc. of the Annual Computer Security Application Conference*, Dec. 1998.

[7] R. Agrawal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases", *In Proc. of the ACM SIGMOD Conference on Management of Data*, pp.207~216, 1993.

[8] Jung-soo Park, Ming-syan Chen, and P. S. Yu, "An effective hash-based algorithm for mining association rules", *In Proc. of ACM SIGMOD Conference on Management of Data*, pp.175~186, May 1995.

[9] B. Lent, A. Swami and J. Widom, "Clustering Ass-

ociation Rules", *In Proc. of the 13th International Conference on Data Engineering*, 1997.

[10] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Closed Set Based Discovery of Small Covers for Association Rules", *In Proc. 15emes Journees Bases de Donnees Avancees*, pp.361~381, 1999.

[11] T. Kohonen, *Self-Organizing Maps*, Springer, New York, 2001.

[12] K. L. Fox, R. R. Henning, J. H. Reed, and R. Simonian, "A Neural Network Approach Towards Intrusion Detection", *In Proc. of the 13th National Computer Security Conference*, pp.125~134, Oct. 1990.

[13] J. Frank, "Artificial Intelligence and Intrusion Detection : Current and Future", *In Proc. of the 17th Computer Security Conference*, Oct. 1994.

[14] S. A. Hofmeyr, "An Immunological Model of Distributed Detection and its Application to Computer Security", *Ph.D. Thesis, Univ. of New Mexico*, May 1999.

[15] W. Lee and S. J. Stolfo, "Adaptive Intrusion Detection: a Data Mining Approach", Kluwer Academic Pub., 2000

[16] W. Lee, S. J. Stolfo and K.W. Mok, *Algorithms for Mining system audit data*, Data Mining, Rough Sets, and Granular Computing, T. Y. Lin, Y. Y. Yao, and L. A. Zadeh (eds), Physica-Verlag, 2002.

[17] W. Lee and S. J. Stolfo, "Data Mining approachs for intrusion detection", *In Proc. of the 7th USENIX Security Symposium*, Jan. 1998.

[18] W. Lee, S. J. Stolfo and K.W. Mok, "A Data Mining Framework for Building Intrusion Detection Models", *In Proc. of the 1999 IEEE Symposium on Security and Privacy*, May 1999.

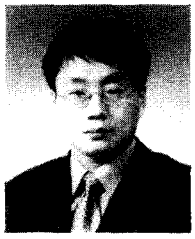
[19] 오세훈, 이원석, "페킷간 연관 관계를 이용한 네트워크 이상행위 탐지", *정보보호학회논문지*, 12 (5), pp.63~73, 2002.

[20] 박정호, 오상현, 이원석, "데이터베이스 시스템에서 연관 규칙 탐사 기법을 이용한 이상 행위 탐지", *정보처리학회 논문지*, 9(6), pp.831~840, 2002.

[21] H. Mannila and H. Toivonen, "Discovering generalized episodes using minimal occurrences", *In Proc. of the 2nd International Conference on Knowledge Discovery in Databases and Data Mining*, Aug. 1996.

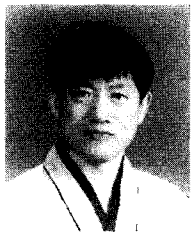
〈著者紹介〉

조 혁 현 (Hyug-Hyun Cho)
1984년 2월 : 홍익대학교 전자계산학과 졸업
1989년 2월 : 전남대학교 전산통계학과 석사
1997년 2월 : 전남대학교 전산통계학과 박사과정수료
1989년 3월~현재 : 여수대학교 정보기술학부 교수
<관심분야> 데이터베이스, 정보보안, 시스템 및 네트워크 보안 등



정 희 택 (Hee-Taek Ceong)
1992년 2월 : 전남대학교 전산통계학과 졸업
1995년 2월 : 전남대학교 전산통계학과 석사
1999년 8월 : 전남대학교 전산통계학과 박사
1999년 9월~현재 : 여수대학교 정보기술학부 조교수
<관심분야> 워크플로시스템 및 보안, 데이터마이닝, 분산처리시스템 등

김 민 수 (Min-Soo Kim)
1993년 : 전남대학교 전산통계학과 졸업(학사)
1995년 : 전남대학교 대학원 전산통계학과(이학석사)
2000년 : 전남대학교 대학원 전산통계학과(이학박사)
2000년~2001년 : 한국정보보호진흥원 선임연구원
2001년~현재 : 전남대학교 리눅스시스템보안연구센터 객원교수
<관심분야> 시스템 보안, 네트워크 보안, 정보보안, 신경망 등



노 봉 남 (Bong-Nam Noh) 정회원
1978년 : 전남대학교 수학교육과 졸업(학사)
1982년 : KAIST 대학원 전산학과 졸업(석사)
1994년 : 전북대학교 대학원 전산과 졸업(박사)
1983년~현재 : 전남대학교 컴퓨터정보학부 교수
2000년 : 리눅스 보안 연구센터 소장
<관심분야> 컴퓨터와 네트워크 보안, 정보보호시스템, 전자상거래 보안, 사이버사회와 윤리