데이타마이닝 기법을 이용한 경보데이타 분석기 구현

(Implementation of Analyzer of the Alert Data using Data Mining)

신 문 선 † 김 은 희 †† 문 호 성 ††† 류 근 호 †††† 김 기 영 †††††
(Moon Sun Shin) (Eun Hee Kim) (Ho Sung Moon) (Keun Ho Ryu) (Ki Young Kim)

요 약 최근 네트워크 구성이 복잡해짐에 따라 정책기반의 네트워크 관리기술에 대한 필요성이 증가하고 있으며, 특히 네트워크 보안관리를 위한 새로운 패러다임으로 정책기반의 네트워크 관리 기술이 도입되고 있다. 보안정책 서버는 새로운 정책을 입력하거나 기존의 정책을 수정, 삭제하는 기능과 보안정책 결정 요구 발생시 정책결정을 수행하여야 하는데 이를 위해서는 보안정책 실행시스템에서 보내온 경보 메시지에 대한 분석 및 관리가 필요하다. 따라서 이 논문에서는 정책기반 네트워크 보안관리 프레임워크의 구조 중에서 보안정책 서버의 효율적인 보안정책 수립 및 수행을 지원하기 위한 경보데이타 분석기를 설계하고 구현한다. 경보 데이타 저장과 분석을 위해서 데이타베이스 스키마를 설계하고 저장된 경보데이타를 분석하는 모듈을 구현하며 경보데이타 마이닝 엔진을 구현하여 경보데이타를 효율적으로 분석하고 이를 통해 경보들의 새로운 유사패턴그룹이나 공격시퀀스를 유추하여 능동적인 보안정책관리를 지원할 수 있도록 한다.

키워드 : 정책기반보안관리, 침입탐지, 경보데이타, 데이타마이닝, 연관규칙, 빈발사건, 클러스터링

Abstract As network systems are developed rapidly and network architectures are more complex than before, it needs to use PBNM(Policy-Based Network Management) in network system. Generally, architecture of the PBNM consists of two hierarchical layers: management layer and enforcement layer. A security policy server in the management layer should be able to generate new policy, delete, update the existing policy and decide the policy when security policy is requested. And the security policy server should be able to analyze and manage the alert messages received from policy enforcement system in the enforcement layer for the available information.

In this paper, we propose an alert analyzer using data mining. First, in the framework of the policy-based network security management, we design and implement an alert analyzer that analyzes alert data stored in DBMS. The alert analyzer is a helpful system to manage the fault users or hosts. Second, we implement a data mining system for analyzing alert data. The implemented mining system can support alert analyzer and the high level analyzer efficiently for the security policy management. Finally, the proposed system is evaluated with performance parameter, and is able to find out new alert sequences and similar alert patterns.

Key words : Policy-Based Network Management, Intrusion Detection System, Alert Data, Data Mining, Association Rule, Frequent Episodes, Clustering

· 이 연구는 한국전자통신연구원과 과학기술부 RRC(청주대ICRC)의 연구비 지원으로 수행되었음

† 정 회 원 : 충북대학교 전자계산학과
msshin@dblab.chungbuk.ac.kr
†† 비 회 원 : 충북대학교 전자계산학과
ehkim@dblab.chungbuk.ac.kr
††† 비 회 원 : 가림정보기술
hsmoon@dblab.chungbuk.ac.kr
†††† 종신회원 : 충북대학교 전기전자및컴퓨터공학부 교수
khryu@dblab.chungbuk.ac.kr
††††† 비 회 원 : 한국전자통신연구원 네트워크정보보호연구본부 연구원
kykim@etri.re.kr

논문접수 : 2003년 4월 3일
심사완료 : 2003년 10월 22일

1. 서 론

정책기반의 네트워크 관리는 네트워크 보안관리를 위한 새로운 패러다임으로 네트워크 전반에 대해 일관된 정책을 수행하고 적절한 정책을 수립하며, 관리자의 요구에 대해 정책의 용이한 변경을 제공함으로써, 네트워크 전반의 중앙 집중적인 관리를 가능케 하는 메커니즘[1]이다. 네트워크 환경에서 동적으로 용이하게 네트워크의 운영 방침을 적용하여 효율적인 네트워크를 운용하는 것이 정책기반 네트워크 관리의 목적이다. 정책에

의해 운영자는 손쉽게 네트워크를 관리할 수 있으며 상세 구현과 상관없이 일관성 있고 통합적이면서도 이해하기 쉬운 네트워크의 관리[2]를 가능하게 한다. 그럼에도 불구하고 인터넷 위협에 대응하기 위한 네트워크 전반에 걸친 완벽한 관리 메커니즘[3,4]이나 완벽한 침입 대응 시스템은 없다. 또한 네트워크 기반 침입 탐지시스템[5,6]에서 갈수록 다양해지는 침입에 능동적으로 대응하는 데에는 어려움이 많다[3,7]. 따라서 최근에는 침입 탐지 시스템에 데이타 마이닝 기법을 적용하여 많은 양의 감사데이타를 효율적으로 분석하거나 자동화된 침입 탐지 모델을 구축[8-10]하는 연구가 국내외에서 활발히 이루어지고 있다. 또한 침입탐지시스템의 성능향상과 알려지지 않은 공격에 대응하기 위한 방안으로 경보 상관관계 분석에 대한 연구[11,12]가 진행되고 있다. 이는 보안서비스의 질을 향상시키며 능동적인 보안이라는 측면에서 중요한 의미를 가진다. 경보 데이타 간의 연관성 분석을 통해, 새로운 공격의 탐지나 보다 정확한 탐지를 위한 침입 탐지 모델을 구축하고, 사용자에게는 보다 이해하기 용이한 정보를 제공할 수 있기 때문이다.

그런데 기존의 경보 상관관계 분석은 다음과 같은 몇 가지 문제점을 가진다. [13]의 개연적 경보 상관관계 분석은 경보 데이타의 속성간의 유사성을 이용하여 경보 데이타 간의 상관관계를 분석하는 기법으로 속성간의 유사성을 이용하여 공격 타입 간의 유사성을 정의하고 이를 이용하여 공격 타입 간의 연관 관계를 추출한다. 그러나 이 방법은 선택된 속성에 의존적이며, 경보 데이타 간의 인과 관계를 완벽하게 탐사하기에 적당하지 못하다는 단점을 가지고 있다. [14]에서는 발견 학습을 이용한 접근 방법을 "stealthy portscan"을 탐지하기 위해 적용하였다. 비록 발견 학습을 경보 데이타 상관관계 분석에 이용하였지만, 이 방법 또한 경보 데이타 간의 인과 관계를 완벽하게 분석하지 못하였다. [11]은 경보 데이타의 통합과 상관관계 분석 기법을 제안하였다. 특히, [11]에서 제안된 상관관계 분석 방법은 어떤 타입의 경보가 주어진 경보 유형의 다음에 오는지를 기술하기 위한 결과 메커니즘을 이용하였다. 이것은 오용 탐지 기법과 유사하다. 그러나 이 결과 메커니즘은 단지 경보의 유형과 경보에 의한 probe, 보안 레벨, 결과 정의에 포함된 두 경보 간의 시간 간격만을 사용하며, 가능한 모든 경보 데이타들이 서로 관련되기 위한 충분한 정보를 제공하지 않는다는 단점이 있다. 게다가 공격자가 어떻게 공격의 시퀀스를 조정할 것인지 예측하는 것 또한 쉽지 않다. 또 다른 접근 방법으로 공격 시나리오가 포함되어 있는 학습 데이타 집합에 기계 학습 기법을 적용하여 경보 상관관계 모델을 학습하는 기법[15]이 있다. 이 방법은 경보 데이타의 상관관계 분석을 위한 모델을 자동적으로 생성할 수 있는 장점이 있지만, 매 적용마다 학습이 필요하며, 결과 모델이 학습 데이타에 의존적이므로 학습 데이타에 포함되지 않은 공격 시나리오는 탐지할 수 없는 단점이 있다.

기존의 경보 상관관계 분석은 나름대로 여러 가지 방법으로 시도되었으나 특정 공격 시나리오에 대한 학습 방법을 주로 적용한 것으로 학습 시나리오에 없는 공격들에 대한 탐지에 어려움이 많다.

또한 보안정책 서버의 경보데이타 관리에는 기하 급수적으로 증가하는 경보 데이타의 양으로 인해 효율적인 관리에 어려움이 따른다. 따라서 이와 같은 문제점들을 해결하기 위한 새로운 접근방법으로 데이타 마이닝 기법을 적용한 경보 상관관계 분석을 제안한다. 데이타 마이닝 기법은 많은 양의 데이타로부터 알려지지 않은 유용한 지식을 추출해내는 기술이며 데이타 필터링의 효과도 있으므로 알려지지 않은 공격에 대한 시퀀스추출에 활용 가능하며 또한 많은 양의 경보데이타를 감소시키는 역할을 할 수 있기 때문이다. 따라서 이 논문에서는 정책기반 네트워크 보안관리 프레임워크의 구조 중에서 보안정책서버의 효율적인 경보관리를 지원하기 위하여 데이타마이닝 기법을 적용한 경보데이타 분석기를 구현한다. 구현된 경보데이타 분석기는 경보데이타마이닝 엔진을 포함하고 있어 보안정책 실행시스템으로부터 보안정책서버에 보내어지는 경보데이타에 대해 마이닝을 수행 후 생성된 경보 빈발패턴이나 경보 유사성 및 경보시퀀스 추출 등의 최종규칙을 활용, 경보 분석 및 불량호스트나 악의적인 사용자 등을 관리할 수 있는 기능을 제공한다. 보안정책서버는 이러한 경보데이타 분석 결과 정보들을 활용하여 전체 네트워크망의 위험 진단 및 새로운 보안정책을 수행한다. 논문의 구성은 2장에서는 네트워크 보안제어 시스템의 프레임워크에 대해 간략히 기술하고 3장에서는 경보 데이타 분석 모듈 구현을 기술한다. 4장에서는 경보 데이타 분석기를 지원하는 마이닝 시스템의 구현을 설명하고 5장에서는 구현된 마이닝 시스템의 적용 및 평가를 한다. 마지막으로 결론 및 향후 연구를 기술하고 끝을 맺는다.

2. 정책기반 네트워크 보안관리

정책기반 네트워크 보안구조(Policy-Based Network Management for Network Security: NS-PBNM)는 네트워크 보안을 위한 정책기반의 네트워크 관리 기법으로서 정책기반 네트워크 보안구조를 지칭한다.

정책기반 네트워크 보안관리의 프레임워크의 구성요소는 보안 정책을 생성하고 관리하는 PMT(Policy Management Tool), 보안 규칙에 따라 보안 행위를 결정하는 PDP(Policy Decision Point), 보안 규칙을 저장

하는 PR(Policy Repository)과 보안 행위를 수행하는 PEP(Policy Enforcement Point)와 PDP와 PEP간의 보안 정책 전달을 위한 통신 프로토콜[16,17]로 구성된다.

네트워크 보안 정책을 위한 프레임은 정책기반 네트워크 보안 구조의 계층적인 구성을 가지며 적어도 두개의 계층으로 구성한다. 하나는 관리 계층에 해당하는 보안 정책 서버 시스템이며 다른 하나는 실행 계층에 해당하는 접속점에서의 해킹 트래픽 감지 및 대응을 위한 침입탐지 기반의 보안정책 실행 시스템이다.

보안 정책 서버 시스템은 크게 PMT 블록과 PDP 블록, 보안정책 실행 시스템으로부터 전달된 경보를 처리하는 AM(Alert Manager)과 HA(High-level Analyzer)블록과 PR을 위한 디렉토리로 구성된다.

보안 정책 실행 시스템은 네트워크 접속점에서 입력 패킷에 대한 탐지와 분석을 제공하는 Sensor/Analyzer 블록과 보안정책 실행기능을 제공하는 PEP 블록으로 구성된다. 그림 1은 정책기반 네트워크 보안관리의 구성요소와 상호간의 관계를 나타나고 있다.

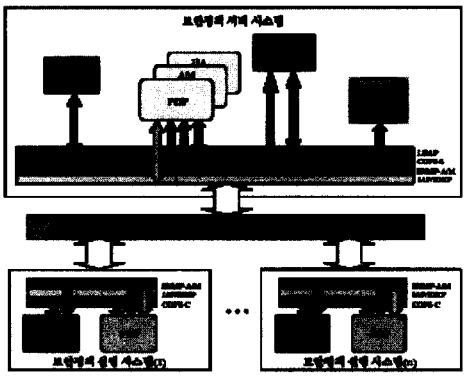


그림 1 정책기반 네트워크 보안관리의 프레임워크

이 논문에서는 보안정책서버 시스템의 HA를 지원하는 경보데이타 분석기를 구현한다. 먼저 경보데이타의 저장을 위한 데이타베이스 스키마를 설계하고 경보데이타를 데이타베이스에 저장하는 Alert Manager모듈을 구현한다. 또한 Black List Watcher 모듈과 Fault User/Host Watcher모듈은 불량사용자와 불량호스트를 관리한다. 그리고 경보데이타의 효율적인 분석을 위해서 마이닝 엔진을 구현한다.

3. 경보데이타 분석기

보안정책 서버의 주요 기능 중 하나가 침입이 발생한 경우 보안정책 실행 시스템에서 보낸 경보 메시지에 대한 포괄적이고 광범위한 침입탐지 분석 및 대응 기능이다. 최근의 네트워크 공격은 사전 예측 및 정상 트래픽과의 구분이 어려우며, 그 피해 영역이 단순히 한 시스템이 대상이 아닌 네트워크 전체를 대상으로 하고 있다. 이와 같이 네트워크 전반에 걸쳐서 발생되는 침입에 대응하기 위해서는 전체 네트워크에서 발생하는 경보의 포괄적인 분석이 필요하다. 이러한 분석 기능을 제공하고, 다양하고 복잡한 형태의 침입을 탐지하기 위하여 경보 데이타의 상관 관계를 분석하는 모듈이 요구된다.

경보의 상관 관계 분석은 경보 데이타 간의 상호 연관성 추출을 의미한다. 이는 경보의 유사 반복성을 분석하는 유사성 분석과 행위의 상관 관계를 분석하는 행위 분석을 포함한다. 경보의 상관 관계 분석을 위해 요구되는 세부적인 기능은 반복성 분석, 유사성 분석, 잠재성 분석, 행위 분석 등이다. 반복성 분석은 동일한 경보의 반복적인 발생에 대한 분석 기능을 제공한다. 유사성 분석은 동일 근원지 및 목적지에 대해 유사 경보 발생에 대한 분석 기능, 동일 근원지에서 발생하는 유사 경보에 대한 분석 기능과 특정 목적지에 대해 유사 경보 발생에 대한 분석 기능을 제공한다. 잠재성 분석은 잠재적 가능성을 분석하는 기능으로 동일 근원지와 동일 목적지, 동일 근원지, 동일 목적지에 대한 행위에 관련된 분석 기능을 제공한다. 행위 분석은 경보 시퀀스 분석, 공격 시퀀스 분석, 근원지 시퀀스 분석, 근원지 및 목적지 시퀀스 분석 기능이 포함된다. 경보 시퀀스 분석은 특정 경보 이후에 발생 가능한 경보의 통계적 가능성을 분석하는 기능을 제공한다. 공격 시퀀스 분석, 근원지 시퀀스 분석, 근원지 및 목적지 시퀀스 분석은 특정 공격이나 근원지, 근원지 및 목적지의 확률적 시퀀스를 의미하며, 기존의 발생 시퀀스를 기반으로 임의의 이벤트 발생시에 다음 단계에서 발생 가능한 이벤트를 확률적으로 분석하는 기능[15,19]을 제공한다.

보안정책 실행 시스템으로부터 전달되는 침입 정보 등을 구조적으로 저장하기 위하여 보안정책 서버는 RDBMS를 이용한다. 이는 경보 데이타와 같은 반복적이고 대량의 데이타를 관리하기 위해서는 안정적이고 고속의 RDBMS가 효과적이기 때문이다. 경보 데이타를 저장하기 위한 스키마는 표 1과 같다.

보안정책 실행 시스템이 지역적인 침입탐지 메커니즘에 의해 침입이 탐지가 되면 정책에 의한 그 대응과 처리를 수행하고 침입과 대응의 요약 정보를 위의 표 1의 스키마 형태로 보안정책 서버 시스템에 전달을 한다. 이렇게 전송된 경보 데이타는 Alert Manager에 의해 데이타베이스에 저장 되고, 저장된 경보데이타는 경보 데이타 분석기와 불량 사용자 및 호스트 관리기에 의해 분석되고 관리된다. 분석한 결과에 의해 전체 네트워크에 영향을 미치는 침입이 탐지가 되면 분석기는 그 결과를 보안 관리자와 대응 결정 시스템에 전송을 하고,

표 1 경보 데이타의 구성

속성이름	설 명
ALID	경보데이타 ID
SID	보안정책 실행시스템 ID
ATID	해킹 ID
ATTYPE	해킹 유형 형태
DDATE	해킹 탐지 일시
SADDR	근원지 IP 주소
DADDR	목적지 IP 주소
SPORT	근원지 포트 번호
DPORT	목적지 포트 번호
PROTO	프로토콜 유형
ICMPTYPE	ICMP 종류
ICMPCODE	ICMP 코드
IMPACT	임팩트 수준
ACTYPE	대응 유형
ACRESULT	대응 결과
ACDATE	대응 일시

이 결과는 정책 결정 과정이나 침입 대응을 지원하게 된다.

경보데이타 분석기는 경보 데이타의 상관 관계 분석을 위해 통계적인 방법을 이용하였다. 즉, 경보데이타 분석 모듈에서는 단순히 정량적인 수치에 의해서 불량 호스트와 불량사용자를 구분하고 관리하였다. 불량 사용자 및 호스트 탐지 및 관리 모듈은 보안 서버에 의해 관리되는 영역 내의 침입의 근원지 및 피해 호스트의 체계적인 관리를 제공하는 것이 그 목적이다. 이러한 목적을 위해 경보 데이타를 이용하여 구축된 특정 불량 사용자와 호스트를 감시하고 특별 관리 대상을 주목함으로써 네트워크의 자원에 대하여 사전 방어 및 조기 대응에 도움을 줄 수 있었다.

불량 사용자 및 호스트 관리 모듈은 축적된 경보 데이타로부터 추출된 정보를 가공하여 불량 사용자와 호스트의 리스트를 구축하고 이를 관리하기 위한 관리 모듈(Black List Manager)과 위험한 불량 사용자 및 호스트를 보안 정책에 따라 탐지하기 위한 탐지 모듈(Fault User/Host Function)로 구성이 된다. Black List Manager는 경보 데이타를 분석하여 불량 사용자와 불량 호스트의 리스트를 구축하고 이를 관리하는 기능을 한다. Fault User/Host Function은 구축된 불량 사용자와 호스트 리스트를 이용하여 불량 사용자 혹은 호스트의 위험도를 판단하고 정해진 기준을 넘어서는 경우 이를 관리자에게 통보하는 기능을 한다. 이러한 사건의 탐지는 임계 수치를 운용하여 결정하며, 임계 수치는 사용자나 호스트에 대해 총 해킹 건수와 각 임팩트별 해킹 건수 별로 설정된다. 임계 수치의 값은 관리자의 보안 운용수준에 따라 사전에 학습된 수치를 기준으로 하였다.

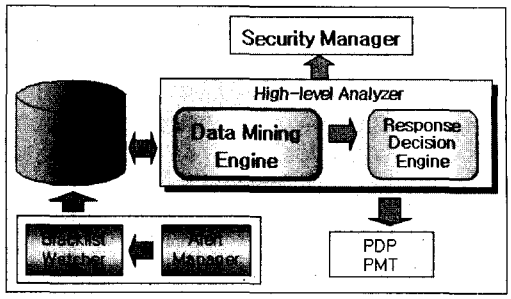


그림 2 경보데이타 분석기 구성 및 관계

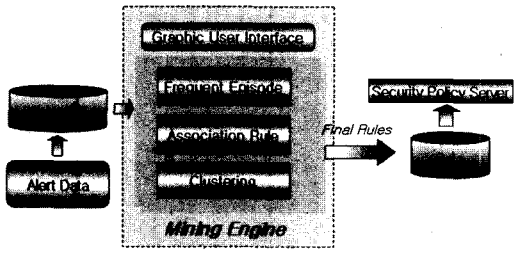


그림 3 경보데이타 마이닝 엔진 구성도

그러나 이 장에서 설계된 모듈의 지능적인 분석과 탐지를 위해 데이타 마이닝 기법을 적용함으로써 고수준의 분석 기능을 제공할 수 있도록 개선하였다. 다음 장에서 경보 데이타 마이닝 시스템의 구현과 적용에 대하여 기술할 것이며 보안정책서버의 정책지원을 위한 경보데이타 분석 모듈과 마이닝 엔진 등의 시스템의 구성과 주변 모듈과의 상관관계는 그림 3과 같다. 즉 그림 2에서 나타난 것처럼 고수준의 분석기에서 마이닝의 결과를 넘겨받아서 보안정책서버의 정책 결정을 지원하게 된다. 불량호스트 및 불량사용자관리 모듈은 고수준 분석기에 포함되어 있다.

4. 경보데이타 마이닝

이 논문에서 구현된 경보데이타 마이닝 시스템은 세 가지로 구성된다. 연관규칙 마이너는 레코드내의 속성간의 연관성을 탐사하고 빈발마이너는 레코드간 이벤트들의 패턴을 탐사하는데 사용된다. 그리고 클러스터링 마이너는 경보데이타들 중 유사성을 가지는 경보데이타들을 그룹화하여 유사한 공격패턴들을 찾아내고자 하였다. 또한 연관규칙마이너와 빈발마이너 모두 경보데이타의 특성을 고려하여 관심 있는 항목들만이 포함된 후보항목집합을 생성하도록 기존의 마이닝 알고리즘을 확장하였다. 즉 마이닝하고자 하는 항목들을 선택하여 관심 있는 항목들에 대해 마이닝을 수행하도록 하였다.

일반적으로 침입탐지시스템을 위한 데이타 마이닝기법으로 분류기법을 적용하여 침입인지 아닌지를 판단하는데 활용할 수 있다. 그러나 경보데이타는 일반적인 감사데이타와는 그 특성이 다르다. 즉, 경보데이타는 보안정책 실행시스템으로부터 보내진 메시지로 그 자체가 이미 침입에 대한 정보를 가지고 있다. 따라서 분류기법을 적용하기보다는 군집화(clustering) 기법을 이용하여 유사한 경보 데이타들을 같은 공격데이타로 군집화하여 동일 공격패턴으로 간주할 수 있도록 하는 것이 더 필요하다. 이는 대량의 경보데이타를 필터링하는 효과를 가지고 오며 또한 경보데이타의 유사도에 따라 순서를 예측하여 유사한 공격 패턴들의 시퀀스로 간주하여 공격에 대한 대응을 할 수 있도록 한다. 즉 클러스터링 마이너는 경보데이타의 유사성 분석을 수행하여 고수준의 의미를 추출 한다. 그림 3은 경보데이타 분석을 위한 마이닝 엔진의 구성도를 나타내고 있다. 구현된 마이닝 엔진의 특성을 살펴보면 다음과 같다. 연관규칙마이너의 경우는 기본 Apriori알고리즘에 키항목 제약조건을 추가하여 구현하였다. 이는 불필요한 후보항목의 생성을 줄이며 관심 있는 속성들에 대한 연관규칙 탐사를 하기 위함이다. 관심있는 속성은 사용자 인터페이스 부분에서 보안관리자가 선택하도록 하였다. 또한 빈발에피소드 마이너의 경우는 row vector라는 개념을 추가 확장하였다. 이는 빈발항목을 찾기 위해 사용되는 자료구조로써 아이템 집합이 포함된 트랜잭션을 기록한 비트를 포함시킨 것을 말한다. 이를 이용하여 속성들 간의 상관관계보다는 튜플들간의 상관관계만을 고려할 수 있다는 이점이 있으며 또한 기준 속성을 적용함으로써 후보항목 생성 시 기준속성을 포함하고 있는 항목만을 고려할 수 있게 하였다. 클러스터 마이너의 경우는 경보데이타가 다차원의 속성을 가지며 또한 경보간의 시퀀스를 추출하기 위하여 CURE 알고리즘을 확장하였다. 즉 CURE 알고리즘에 클러스터간의 시퀀스를 유추할 수 있도록 하는 preCluster 속성 값을 유지하도록 하여 클러스터간의 순서정보를 유추할 수 있도록 하였다.

4.1 기본 알고리즘을 확장한 연관규칙 마이너

기존의 연관규칙 탐사 알고리즘은 트랜잭션 데이타베이스를 그 대상으로 하고 있어 항목들을 T_ID로그룹핑하여 연관규칙을 탐사하지만 이 논문에서 마이닝의 대상은 경보데이타로써 이는 트랜잭션 데이타베이스와는 개념이 다른 형태로 다차원의 속성을 가지고 있고 또한 직관적으로 보안관리자에 의해 중요속성을 선택하는 것이 필요하다. 즉, 특정 공격에 대한 가능성 여부를 판단하기 위해, 예를 들면 특정 호스트에 대한 DOS(서비스 거부)공격에 대한 탐지를 위해서는 목적지IP 주소를 키 속성으로 선택하여 연관 규칙을 탐사할 수 있도록 하는 것이다. 따라서 이 논문에서는 Apriori 알고리즘을 확장 변형한 형태로 키항목(axis attribute)개념을 적용하여 관심 있는 속성들을 선택하여 연관규칙 탐사를 수행하며 최종 규칙을 생성하게 된다. 확장된 알고리즘의 수행은 3단계로 이루어진다

1) 빈발항목 집합을 생성하는 단계

이 단계에서는 로딩된 테이블에서 관심 있는 속성들로 이루어진 항목들에 대해서 후보항목을 생성한 후 후보항목의 지지도(support)가 전체 레코드(D)에서 최소지지도(minsupp)를 만족하는 후보항목에 대한 빈발 항목 집합을 생성한다. 이때 최소 지지도를 만족하지 못하는 항목들은 pruning 단계를 통해서 제거된다. 만일 항목집합 A, B에 대하여 A≥B이면, B를 지지하는 D의 모든 항목들이 필연적으로 A 또한 지지하므로 support (A) = support(B)이다.

하지만, 항목집합 A 가 D에서 최소지지도에 미치지 못한다면, 즉 support(A) < minsupp, support(B) = support(A) < minsupp이기 때문에 A의 모든 상위집합 B는 빈발하지 않게 된다

2) 연관규칙을 생성하는 단계

이 단계에서는 pruning 단계를 거쳐 최소 지지도를 만족하는 항목들로 이루어진 빈발항목집합에 대해서 연관규칙을 생성하게 된다. 이때 빈발항목들간의 최소지지도를 가지고 최소 신뢰도(minconf)를 계산하여 연관규칙을 생성한다. 최소 신뢰도는 최소 지지도를 만족하는 항목에 대해서 얼마나 지지하는 지에 대해 예측하는 확률을 말하며, 빈발 항목에 대한 신뢰도 계산은 Conf(R) = $p(X \subseteq D \mid Y \subseteq D) = p(X \subseteq D \wedge Y \subseteq D) / p(X \subseteq D)$ support (X U Y) / support(X)으로 지지하는 항목에 대한 신뢰도를 구할 수 있다.

3) 최종 룰 생성 단계

이 단계에서는 이전 단계에서 만들어진 룰에 대해서 최소 신뢰도(minconf)를 만족하는 최종 룰, 즉 Conf(R) ≥ minimum confidence, 만을 생성하여 룰 테이블에 저장하게 된다.

위와 같이 3단계로 속성들간의 연관성을 마이닝 하여 많은 양의 경보데이타를 효율적으로 분석할 수 있으며 키 속성제약사항에 따라 관심 있는 속성들간의 연관성을 분석하며, 불필요한 룰의 생성을 줄일 수 있다.

4.2 빈발 에피소드 마이너

마이닝 기법 중 빈발 에피소드 탐사는 일련의 사건 시퀀스로부터 빈번하게 발생하는 에피소드를 찾는 기법이다. 에피소드는 빈번하게 발생하는 특정 사건 시퀀스로 정의되며 시퀀스를 구성하는 사건은 서로 밀접하게 관련된 사건이다. 탐사 문제는 사용자가 지정한 단위 윈도우 크기를 갖는 시간 윈도우들의 집합에서 에피소드

가 발생한 윈도우의 비율이 최소 빈발도 이상을 만족하는 모든 빈발 에피소드를 찾는 것이다. 순차 패턴과 에피소드는 순차적으로 패턴을 탐사한다는 점에서는 거의 유사하지만, 순차 패턴은 전체 데이타베이스를 대상으로 패턴을 탐사하는데 비해 에피소드는 윈도우를 이용해 가면서 탐사한다는 점에서 차이점이 있다. 침입 탐지 시스템에서는 이를 이용하여 자주 반복되는 패턴을 탐지하고 이를 룰에 적용시키거나, 서비스 거부 공격의 지침으로 이용할 수 있다.

경보 데이타로부터 유용한 패턴을 찾기 위해 데이타 마이닝을 적용하는데 있어 기존의 알고리즘을 이용할 경우 속성들 간의 상관관계를 고려해야 하는 문제점이 발생한다. 경보데이타는 여러 가지 속성들로 이루어져 있으며 또한 각 속성들은 많은 값을 가지게 된다.

이 모든 데이타들을 binary 데이타베이스로 변환 시킬 수 없기 때문에 이 논문에서는 row vector를 이용한 확장된 알고리즘을 제시한다. row vector라는 것은 빈발항목을 찾기 위해 사용되는 자료구조로써 아이템 집합이 포함된 트랜잭션을 기록한 비트를 포함시킨 것을 말한다. 이를 이용하여 속성들 간의 상관관계보다는 튜플들간의 상관관계만을 고려할 수 있다는 이점이 있으며 또한 기준 속성을 적용함으로써 후보항목 생성 시 기준속성을 포함하고 있는 항목만을 고려할 수 있게 하였다.

이는 규칙 생성 시 불필요한 에피소드 항목들이 많아지는 것을 줄일 수 있다. 빈발 에피소드 마이닝은 다음과 같이 3 단계로 수행된다.

1) 후보 에피소드 생성단계

이 단계에서는 로딩된 테이블에서 관심 있는 속성들로 이루어진 투플들에 대해서 주어진 time window 단위에 의해서 투플들을 정렬한다. 윈도우 내의 time은 윈도우의 time 범위에 포함되어 있어야 한다. 즉, win = T_e - T_s + width(w), win_start time <= time < win_end time

윈도우 단위로 정렬된 테이블을 가지고 후보 에피소드 집합을 생성하게 된다. 에피소드는 (V, ≤, g) 노드들의 집합인 V와 V에 대한 부분 순서, 그리고 각 노드를 사건 형태와 연결하는 매핑인 g: V → E로 구성된다. 즉 g(V) 내의 사건들은 '≤'에 의해 표현되는 순서대로 발생해야 한다.

에피소드 α가 이벤트 시퀀스 s 내에서 빈발하게 나타나면 모든 서브에피소드 $\beta \leq \alpha$ 역시 빈발하게 나타난다. 후보항목 집합은 빈발한 작은 서브에피소드들로 구성되어지며 이것으로서 빈발하지 않을 수 있는 에피소드들에 대해 안전하게 제거할 수 있는 기준이 된다.

2) 빈발 에피소드 생성 단계

생성된 후보 에피소드 집합에서 최소빈발도(minimum frequent)를 만족하는 에피소드들을 추출하여 빈발한 에피소드집합을 한다. 에피소드들의 빈발도들에 대해서 신뢰도를 계산을 하게 된다. 에피소드들에 대한 신뢰도 계산은 다음과 같이 구할 수 있다.

Conf(R) = p(X ⊆ D | Y ⊆ D)

= p(X ⊆ D ∧ Y ⊆ D) / p(X ⊆ D)

⇒ frequency (X U Y) / frequency(X)

3) 최종 에피소드 생성 단계

생성된 빈발 에피소드들로부터 최소 신뢰도(minconf)를 만족하는 빈발 에피소드를 생성해 낸다.

⇒ Conf(R) >= minconf

위의 단계로 마이닝을 수행함으로써 규칙 생성 시 불필요한 에피소드 항목들이 많아지는 것을 감소시킬 수 있다.

4.3 클러스터링 마이너

클러스터링 분석은 유사도가 높은 데이타들을 같은 그룹으로 분류하여 주어진 데이타의 분포나 패턴을 찾아내는 기법[20]이다. 이와 같은 클러스터링 분석 기법을 경보 데이타의 분석에 적용함으로써, 효율적으로 데이타를 분석할 수 있으며, 데이타의 그룹화를 통해 고수준의 의미를 추출할 수 있다.

이 기법은 개체들의 집합을 개체의 클래스들로 그룹화하는 절차이다. 이때, 동일한 클러스터에 속하는 개체들은 유사성을 가지고, 다른 클러스터에 속하는 개체간에는 상서성을 가진다. 클러스터링에서 개개 데이타 개체 간의 유사성은 근접 지수에 의해 정의된다. 이 근접 지수는 두 데이타 개체 간의 유사성이나 연관성을 측정할 수 있는 함수이다. 개체 간의 유사성을 측정하는 방법에는 여러 가지가 있지만 주로 거리 개념을 이용하여 측정한다. 이 논문에서 구현된 클러스터 마이너는 개체 간의 유사성을 정의하기 위해서 유클리드 거리 함수를 이용한다. 이는 동일한 속성 값들을 가지는 데이타 개체는 유사하다는 가정을 기반으로 한다. n개의 속성을 가지는 데이타 개체의 계층적인 클러스터링 기법은 개체들의 계층적인 분해를 통해 클러스터링을 수행하는 기법이다. 이 논문에서는 경보데이타가 다차원의 속성을 가지고 있으며 경보간의 시퀀스 추출을 하기 위하여 CURE알고리즘을 확장 구현하였다. 즉 CURE 알고리즘에 클러스터간의 시퀀스를 유추할 수 있도록 하기 위해 preCluster 속성 값을 유지하도록 하여 클러스터간의 순서정보를 유추할 수 있도록 하는 모듈이 추가되었다.

경보데이타 클러스터의 목적은 입력 데이타를 그룹화하고, 결과 그룹 간의 순차적인 의미를 추출하는 것이다. 경보 데이타의 유사성을 분석하는 마이닝 시스템은 Data Preprocessor, Alert Cluster, Cluster Analyzer,

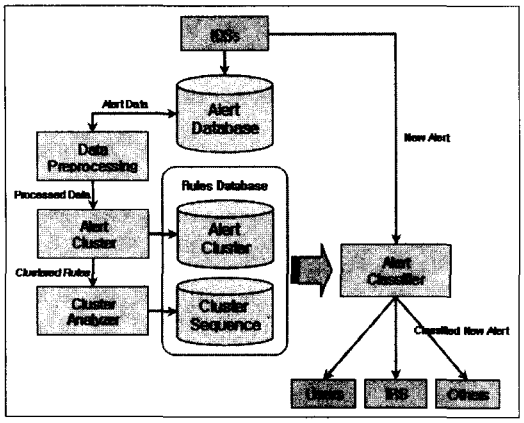


그림 4 경보데이타 클러스터링 마이너

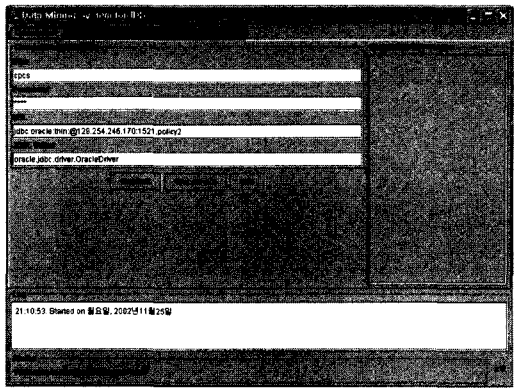


그림 5 사용자 인터페이스

Alert Classifier로 구성된다. 경보 데이타의 유사성을 분석하기 위한 클러스터링 마이너 시스템의 구조는 그림 4와 같다.

Data Preprocessor는 입력된 데이타 집합에 대해 Alert Cluster가 클러스터링을 수행할 수 있도록 전처리를 한다. 여기에서 효율적이고 보다 정확한 클러스터링을 위하여 도메인 지식에 의한 확장 속성을 추가하고 선택된 속성에 대해 정규화를 수행한다. Alert Cluster는 Data Preprocessor에 의해 처리된 데이타에 대해 실제 클러스터링을 수행한다. 이 모듈의 최종 결과는 그룹화된 데이타의 집합들이다. 그 결과는 룰 데이타베이스에 저장이 되고, 이는 이후에 새로운 경보의 자동적인 분류나 생성된 클러스터 간의 연관 관계 분석에 이용된다. Cluster Analyzer는 클러스터링의 수행을 통해 생성된 클러스터의 생성 원인을 분석한다. 이것에 의해 수행된 결과는 클러스터의 시퀀스로 표현된다. 이를 이용하여 우리는 클러스터간의 연관 관계를 분석할 수 있으며, 특정 경보에 대한 차후 가능한 경보의 집합의 예측에 이용할 수 있다. Alert Classifier는 Alert Cluster에 의해 생성된 클러스터 모델을 이용하여 새로운 경보를 적절한 클러스터로 분류하고, Cluster Analyzer의 결과로서 생성된 시퀀스를 이용하여 차후 발생 가능한 경보들을 추출하는 역할을 수행한다.

4.4 구현

구현 환경은 OS는 윈도우XP, 개발언어는 자바, 데이타베이스는 오라클8i를 사용하였다. 구현된 시스템은 사용자 인터페이스, 연관규칙 마이너, 빈발에피소드 마이너, 클러스터마이너 등으로 구성된다. 먼저, 사용자 인터페이스 부분에서는 마이닝할 데이타를 선택하도록 되어 있다. 데이타베이스에 저장된 테이블중 마이닝을 수행할 테이블을 선택한 후 어떠한 속성 값에 대해서 마이닝을 수행할 것인지를 선택한다. 이는 모든 항목을 대상으로 마이닝을 수행하지 않고 중요한 항목들에 대해서만 마이닝을 수행할 수 있도록 하기 위해서이다. 이렇게 특정 속성값을 선택함으로써 불필요한 많은 양의 후보항목집합을 줄일 수 있으며 또한 연관성 없는 항목들에 대한 연관규칙이나 빈발 에피소들 규칙을 줄일 수 있다. 그림 5는 사용자 인터페이스 부분이다. 연관규칙 마이너 탭을 선택하면 지지도와 신뢰도를 사용자가 입력할 수 있으며 관심 있는 항목 속성값을 선택한 후 마이닝을 수행할 수 있다. 또한 빈발에피소드마이너는 빈발도와 신뢰도 외에 타임윈도우의 값을 입력하여야 한다. 입력된 타임윈도우에 따라 최종 룰은 달라지게 된다. 클러스터링 마이너의 경우는 마이닝을 수행하고 그 결과를 뷰어를 통해서 볼 수 있다. 각각의 클러스터는 클러스터 ID를 가지고 데이타베이스 테이블에 저장된다. 다음 장에서 실제 경보데이타를 가지고 실험한 예를 설명하고 클러스터 마이너의 경우는 클러스터링의 정확도를 측정하기 위한 실험 및 클러스터 시퀀스생성에 대한 실험을 수행한 결과를 기술한다.

5. 적용 및 평가

5.1 적용

이 절에서는 실제 경보데이타를 대상으로 마이닝을 수행한다. 경보데이타는 시뮬레이션 결과로 임의로 얻어진 데이타이며 표 2는 저장된 경보데이타의 일부이다.

이 경보데이타를 대상으로 연관규칙 마이너를 수행시킨 결과는 그림 6과 같다. 또한 생성된 연관규칙 룰들의 의미를 정리하면 표 3(a)와 같은 결과를 얻을 수 있다. 또한 빈발 에피소드 마이너의 결과는 그림 7과 같으며 빈발에피소드 최종 룰의 의미를 정리하면 표 3(b)와 같다.

이러한 룰들은 지지도와 신뢰도에 근거한 신뢰성 정

표 2 예제 경보 데이타

[illegible]	ATID	AT_TYPE	DDATE	SCR_IP	DST_IP	SCR_PORT	DST_PORT	PROTO
1	50	7	2002.07.20 22:10:10	203.255.71.10	210.155.167.10	9158	21	TCP
2	50	6	2002.07.20 22:10:11	210.115.167.79	210.155.167.10	9159	21	TCP
3	50	5	2002.07.20 22:10:12	211.115.161.19	210.155.167.10	9160	21	TCP
...	...	...	...	...	...	...	...	...
3	50	5	2002.07.20 22:10:12	211.115.161.19	210.155.167.10	9160	21	TCP

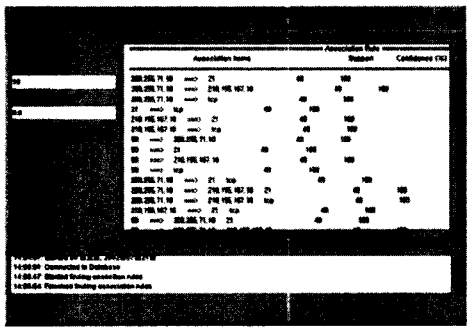


그림 6 연관규칙 마이너 실행결과

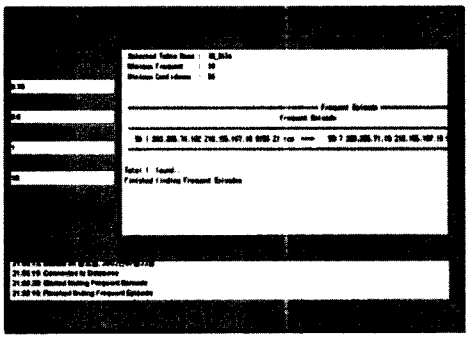


그림 7 빈발 에피소드 마이너 실행결과

보라고 할 수 있다. 예를 들면 attack id가 50은 목적지 포트번호 21번과 연관되어 있다는 것을 알 수 있다. 즉 연관규칙 마이너를 통해서 attack id속성과 destination port속성은 서로 밀접한 관계가 있다는 사실을 추출할 수 있다.

또한 빈발 에피소드 마이너를 실행시킨 경우도 최종 규칙을 살펴보면 5001공격 다음에 5007 공격이 일어난다는 것을 알 수 있다. 이 실험에서 사용된 경보데이타는 단순한 시뮬레이션 데이타이다. 따라서 추가적으로 구현된 시스템의 성능을 평가하는 작업이 필요하다. 그러나 분명한 것은 이 경보데이타 마이닝 시스템이 보안정책서버에게 신뢰성 정보를 제공해줄 수 있다는 것이다.

5.2 연관규칙마이너와 빈발에피소드마이너 실험 평가

실험에서는 생성된 데이타를 이용하여 연관규칙과 빈발에피소드의 최소 지지도 변경, 윈도우 폭에 따른 빈발에피소드 수행시간 등을 평가한다.

성능평가를 위해서 32000개의 레코드를 가지고 최소지지도를 줄여 가면서 성능 평가를 하였다. 최소지지도는 20%, 15%, 10%, 5% 씩 줄여가면서 실험을 하였다. 그림 8은 32000개의 데이타집합에 연관규칙과 빈발에피소드의 최소지지도를 줄여 가면서 실행시간을 측정한 결과를 보여준다.

표 3 최종 규칙의 의미

(a) 연관규칙

Association Rule	Meaning
50<=>21 (supp : 49, conf : 100%)	Attribute 50 (Atid) correlated with attribute 21 (dsc_port)
21<=>tcp (supp : 49, conf : 100%)	Attribute 21 (dsc_port) correlated with attribute tcp (protocol)
...	...

(b) 빈발에피소드 규칙

Frequent Episode Rule	Meaning
5001:210.155.167.10:21:tcp ⇒ 5007:210.155.167.10.21:tcp (fre : 10, conf : 100%, time : 10sec)	If 5001(Ftp Buffer Ovrflow) occur, then 5007(Anonymous FTP) occur together.
...	...

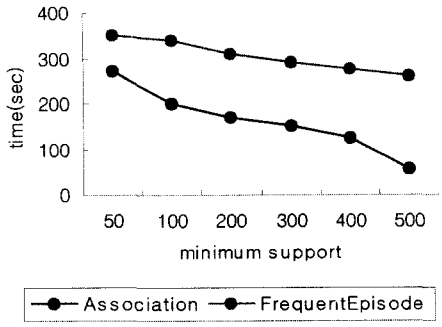


그림 8 최소 지지도 변경에 따른 수행시간

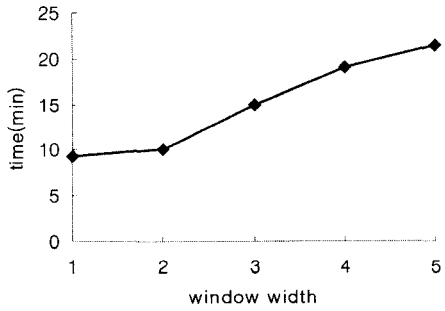


그림 9 윈도우 폭 변경에 따른 빈발에피소드 수행시간

그 결과 연관규칙에서는 지지도가 작을수록 데이타베이스에서 생성되는 후보 항목 수가 많아지므로 그만큼 수행시간도 오래 걸린다. 기본 데이타베이스 테이블을 스캔하면 다음 단계부터는 각 단계에서 생성된 테이블을 스캔하므로 기존 알고리즘보다는 수행시간이 적게 든다.

빈발에피소드에서는 빈발도(즉, 지지도)가 윈도우 테이블 수에 따라 영향을 받도록 되어 있기 때문에 빈발도가 변경됨에 따른 수행시간에는 변화가 적다. 그리고 연관 규칙과 빈발에피소드는 각 단계별로 생성되는 테이블이 최종 규칙이 생성되면서 나머지 테이블들을 삭제를 해버리기 때문에 규칙을 탐사하는 수행시간은 길지 않지만 전체 마이너를 수행하는 시간은 다른 방법에 비해 조금 더 걸리게 된다.

그림 9는 빈발에피소드 마이너에서 윈도우 폭을 1부터 하나씩 늘려가면서 수행 시간을 측정한 결과이다. 윈도우의 폭에 따라 윈도우 테이블이 생성되기 때문에 여기에서는 윈도우 폭을 1로 주었을 때 윈도우 테이블이 2199개가 생성이 되었다. 그래서 윈도우 폭을 증가시킴으로서 많은 수의 윈도우 테이블을 생성하기 때문에 시간이 오래 걸린다는 단점이 있었다.

5.3 클러스터링 마이너 실험 평가

클러스터링 마이너의 경우 두 가지 실험을 수행하였다. 첫 번째는 구현된 시스템의 클러스터링 성능을 평가하기 위한 실험이다. 이 실험은 구현된 시스템에 의해 생성되는 각 클러스터의 정확도를 평가하는 것이다. 두 번째는 생성된 클러스터에 대해 각 클러스터의 이전 클러스터를 정의하고 이를 기반으로 클러스터의 시퀀스를 생성할 수 있는지의 여부를 평가하기 위한 실험이다. 실험을 위해 이 논문에서 사용한 실험 데이타는 KDD Cup 1999 데이타 집합[21]이다. 이 데이타 집합은 DARPA 1998[22]를 이용하여 몇가지 속성을 추가하여 생성한 데이타 집합이다. [22]의 트레이닝 데이타는 7주간의 네트워크 트래픽으로 구성된 TCP Dump 데이타이다. 이 데이타 집합은 약 5,000,000 개의 데이타 인스턴스로 구성되어 있으며, 네트워크 환경 상에서 가능한 다양한 형태의 침입을 포함하고 있다. 또한 테스트 데이타 집합은 약 2,000,000개의 데이타 인스턴스로 구성되어 있으며, 2주간의 네트워크 트래픽을 기반으로 생성된 데이타 집합이다.

트레이닝 데이타 집합을 이용하여 형성한 모델에 대해 테스트 데이타 집합을 이용하여 실험한 결과는 표 4와 같다. 테스트 데이타의 경우 공격 레이블을 참조하지 않고 클러스터링을 수행한 후 공격레이블과 비교하여 각각의 공격 인스턴스별 클러스터링 정확도를 계산하였다. 표 4에 나타나 있는 것처럼 테스트 데이타에 대해서 트레이닝 데이타 집합에 비교적 많이 분포한 클러스터인 DOS, Probing 같은 공격 타입의 경우에는 정확하게 클러스터를 할당하였다. 그러나 작게 분포한 R2L, U2R 공격 타입은 비교적으로 클러스터링 정확도가 떨어졌다. 이처럼 초기 클러스터링 모델을 설정하는 입력 데이타 집합이 다른 요소에 비해 새로운 데이타의 클러스터링의 결과에 많은 영향을 미쳤다.

또한 생성된 클러스터에 대해 각 클러스터의 이전 클러스터를 정의하고 이를 기반으로 클러스터의 시퀀스를 생성할 수 있는지의 여부를 평가하기 위한 실험이다. 입력 데이타 집합은 실제로 생성된 경보 데이타를 이용하였으며 사용자 정의 변수는 앞의 실험에 의해 구하여진 값을 사용하였다. 생성된 각과 클러스터에 대한 각 이전 경보 데이타의 분포는 표 5와 같이 나타났다. 그림 10은 표 5에 나타난 결과의 생성된 시퀀스를 나타내고 있다.

표 4 테스트 데이타의 분류 결과

공격타입	클러스터링 정확도
DOS	98.34%
R2L	47.52%
U2R	51.37%
Probing	83.84%

표 5 클러스터링 분석 결과

Cluster1	54.85%	0.23%	1.54%	10.45%	32.34%
Cluster2	10.40%	7.46%	1.39%	71.23%	9.21%
Cluster3	29.09%	3.10%	32.67%	12.72%	11.88%
Cluster4	0.93%	2.49%	20.10%	67.86%	10.23%
Cluster5	2.43%	5.45%	10.44%	1.93%	76.06%

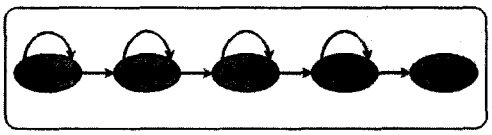


그림 10 생성된 클러스터의 시퀀스

이 실험은 결과 클러스터의 생성 원인이 되는 이전의 경보의 분포를 분석하여 클러스터 간의 시퀀스를 생성하였고, 생성된 각각의 클러스터 시퀀스를 통합하여 클러스터들의 시퀀스를 추출하여 발생한 경보의 향후 가능한 경보 타입을 예측하기 용이한 방법을 제공할 수 있다는 것을 확인시켜 주었다.

6. 결 론

이 논문에서는 보안정책 서버의 효율적인 보안정책 수립 및 수행을 지원하기 위한 경보데이타 분석기를 설계하고 구현하였다. 경보 데이타 저장과 분석을 위해서 데이타베이스 스키마를 설계하고 저장된 경보데이타를 분석하는 모듈을 구현하였다. 또한 경보 데이타 마이닝 엔진을 구현하여 경보데이타를 효율적으로 분석하고 마이닝 결과 생성된 의미 있는 규칙들은 보안정책 서버가 활용할 수 있도록 하였다.

경보상관관계 분석을 위해서 기존의 마이닝 기법들 중 연관규칙, 빈발 에피소드 및 클러스터링 기법을 확장 적용하였다. 연관규칙 마이닝과 빈발에피소드 마이닝 결과 생성된 최종 규칙들은 빈발 경보시퀀스분석과 빈발 공격시퀀스 분석에 활용할 수 있다. 또한 클러스터 마이너는 경보 데이타의 특성을 고려하여 다차원의 속성을 가지는 데이타 집합에 대해서도 클러스터링이 가능한 CURE 알고리즘을 확장하여 구현하였다. 새로운 경보를 적절한 클러스터에 할당하기 위하여, 클러스터의 대표값을 이용하여 개개 데이타 인스턴스와 생성된 클러스터들과의 유사도를 측정하는 방법을 정의하고 구현하였으며, 클러스터에 포함된 경보의 이전 경보를 분석하여 클러스터간의 시퀀스를 추출하기 위한 방법 또한 정의하였다. 특히 경보데이타 분석기를 구현함에 있어 클러스터링 기법을 적용한 것은 데이타간의 유사성을 이용한 경보 데이타의 그룹화를 통해 생성된 모델을 이용하여 새로운 경보 데이타에 대한 분류를 자동화할 수 있다. 이것은 과거에 탐지된 공격의 형태뿐만 아니라 새로운 혹은 변형된 경보의 분류나 분석에도 이용할 수 있다. 또한 생성된 클러스터의 생성 원인의 분석을 이용한 클러스터 간의 시퀀스의 추출을 통해 사용자가 공격의 순차적인 구조나 그 이면에 감추어진 전략을 이해하는데 도움을 줄 수 있으며, 현재의 경보 이후에 발생 가능한 경보들을 예측할 수 있다.

경보데이타의 효율적이고 능동적인 분석을 위해서 마이닝 기법을 적용하였으나 향후 연구로는 통계적기법을 기반으로 경보들 간의 상관관계분석에 따른 경보특성과 속성간의 상관도를 추출하여 이를 기반으로 상관도가 높은 항목들에 대한 속성 선택이 가능하도록 하는 연구와 실제 보안 제어시스템에 적용하고 평가하는 작업이 이루어질 것이다.

참 고 문 헌

[1] IPHIGHWAY, Inc., Introduction to Policy-based Networking and Quality of Service.

[2] B. Moore, E. Ellesson, J. Strassner, and A. Westerinen, Policy Core Information Model Version 1 Spec., IETF RFC3060, Feb. 2001.

[3] D.Schnackenberg, H. Holliday, R. Smith, K. Djahandari, and D. Sterne, Cooperative Intrusion Traceback and Response Architecture (CITRA), DISCEX01, Anaheim, California, June 2001.

[4] S. M. Lewandowski, D. J. Van Hook, G. C. OLeary, J. W. Haines, and L. M. Rossey, SARA: Survivable Autonomic Response Architecture, DISCEX01, Anaheim, California, June 2001.

[5] D. Anderson, T. Frivold, A. Valdes, "Next-generation Intrusion Detection Expert System (NIDES)," Technical Report SRI-CLS-95-07, May 1995.

[6] R. Heady, G. Luger, A. Maccabe, and M. Servilla. "The Architecture of a Network Level Intrusion Detection System," Technical report, University of New Mexico, Department of computer Science, Aug. 1990.

[7] D. Schnackenberg, K. Djahandari, and D. Sterne, Infrastructure for Intrusion Detection and Response, Proceedings of the DARPA Information Survivability Conference and Exposition, SC, Jan. 2000.

[8] W. Lee, W. Fan "Mining System Audit Data: Opportunities and Challenges," College of Computing Georgia Institute of Technology Atlanta, GA 30332-0280, IBM T.J. Watson Research Center Hawthorne, NY 10532.2000.

[9] W. Lee, S. J. Stolfo, K. W. Mok "A Data Mining Framework for Building Intrusion Detection Models," 2001.

[10] W. Lee, S. J. Stolfo. "Data Mining Approaches for Intrusion Detection," Columbia University, Computer Science Department, 1998.

[11] H. Debar and A.Wespi, "Aggregation and Correlation of Intrusion-Detection Alerts," In Recent Advances in Intrusion Detection, number 2212 in Lecture Notes in Computer Science, pages 85~103, 2001.

[12] M. S. Shin, H. S. Moon, K. H. Ryu, J. O. Kim and K. Y. Kim, "Applying Data Mining Techniques to Analyze Alert Data," APWeb2003, Xi'an, China, Apr. 2003.

[13] A. Valdes and K. Skinner, "Probabilistic Alert Correlation", In Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection (RAID 2001), pages 54~68, 2001.

[14] S. Staniford, J.A. Hoagland, and J.M. McAlerney, "Practical Automated Detection of Stealthy Portscans," In ACM Computer and Communications Security IDS Workshop, pages 1~7, 2000.

[15] O. Dain and R.K. Cunningham, "Fusing a Heterogeneous Alert Stream into Scenarios," In Proceedings of the 2001 ACM Workshop on Data Mining for Security Applications, pages 113~, Nov. 2001.

[16] E. Lupu and M. Sloman, Conflicts in Policy-based Distributed Systems Management, IEEE Transactions on Software Engineering, Vol. 25, No. 6, Nov. 1999.

[17] A. Westerinen, J. Schnizlein, J. Strassner, M. Scherling, B. Quinn, S. Herzog, A. Huynh, M. Carlson, J. Perry, and S. Waldbusser, Terminology for Policy-Based Management, IETF, July 2001.

[18] H. Jiawei and K. Michelle, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000

[19] P. Ning and Y. Cui., "An Intrusion Alert Correlator based on Prerequisites of Intrusions," Technical Report TR-2002-01, Department of Computer Science, North Carolina State Univ., Jan. 2002.

[20] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim, "CURE: An Efficient Clustering Algorithm for Large Databases," In Proceedings of the International Conference on Management of Data, (SIGMOD), Vol. 27(2), Seattle, WA, USA, 14, ACM Press, Jun. 1998

[21] KDD99Cup, ttp://kdd.ics.uci.edu/databases/kddcup-99/kddcup99.html, 1999.

[22] Lincoln Lab MIT. DARPA 2000 Intrusion Detection Evaluation Datasets. http://ideval.ll.mit.edu/2000.

[23] H. S. Moon, M. S. Shin, K. H. Ryu and J. O. Kim "Implementation of Security Policy Server's Alert Analyzer," In Proceedings of the International Conference on Computer and Information Science, pages 142-147, Seoul, Korea, Aug. 2002.

[24] M. S. Shin, E. H. Kim, H. S. Moon, K. H. Ryu and K. Y. Kim, "Data Mining Methods for Alert Correlation Analysis," Submitted for publication, International Journal of Computer and Information Science, USA, June, 2003.

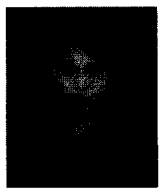
[25] 김은희, 신문선, 문호성, 류근호, 김기영 "감사데이타 분석 마이너 설계 및 구현", 정보과학회 춘계학술발표, 2002년 4월.

신 문 선

1988년 충북대학교 전산통계학과 학사
1997년 충북대학교 전자계산교육 석사
1999년~현재 충북대학교 전자계산학과 박사과정 수료. 관심분야는 시공간 데이타베이스, 데이타 마이닝, 데이타베이스 보안, 침입 탐지 시스템

김 은 희

2001년 삼척대학교 정보통신공학과 학사
2003년 충북대학교 대학원 전자계산학과 석사. 2003년 충북대학교 대학원 전자계산학과 박사과정 재학. 관심분야는 데이타베이스 보안, 시간 데이타 마이닝, 침입 탐지 시스템

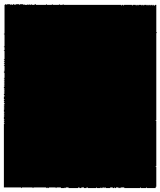


문 호 성

2001년 충북대학교 컴퓨터공학과 학사
2003년 충북대학교 대학원 전자계산학과 석사. 2003년~현재 가림정보기술. 관심분야는 데이타베이스 보안, 네트워크 보안, 데이타 마이닝, 침입 탐지 시스템

류 근 호

1976년 숭실대학교 전산학과(이학사) 1980년 연세대학교 공학대학원 전산전공(공학석사). 1988년 연세대학교 대학원 전산전공(공학박사). 1976년~1986년 육군군수 지원사 전산실(ROTC장교), 한국전자통신 연구원(연구원), 한국방송통신대 전산학과(조교수) 근무. 1989년~1991년 Univ. of Arizona Research Staff (TempIS 연구원, Temporal DB) 1986년~현재 충북대학교 전기전자 및 컴퓨터공학부 교수 관심분야는 시간 데이타베이스, 시공간 데이타베이스, Temporal GIS, 객체 및 지식기반 시스템, 지식기반 정보검색 시스템, 데이타마이닝, 데이타베이스 보안 및 Bio-Informatics



김 기 영

1988년 전남대학교 전산통계학과 학사 1993년 전남대학교 전산통계학과 석사 2000년 충북대학교 대학원 전자계산학과 박사. 1989년~현재 한국전자통신연구원 책임연구원. 네트워크 정보보호 연구본부 네트워크 보안구조 연구팀 관심분야는 Network Security, Protocol Engineering, Policy-based Secure Routing, Network Security Architecture