# 점진적 마이닝 기법을 적용한 침입탐지 시스템의 오 경보 분석 프레임워크 설계

김 은 희[†] · 류 근 호[††]


## 요 약

침입탐지 시스템은 실시간으로 공격행위에 대하여 다량의 경보를 기록한다. 이들 경보 중에는 실제 공격 경보뿐만 아니라 공격으로 잘못 탐지하여 발생된 오 경보들도 있다. 오 경보는 침입탐지 시스템의 효율성을 저하시키는 주요요인이 되므로, 이 논문에서는 오경보 분석을 위한 프레임워크를 제안한다. 또한 지속적으로 증가하는 오 경보를 분석하기 위해 점진적 데이터 마이닝 기법을 적용한다. 제안한 오경보 분석 프레임워크는 GUI, DB Manager, Alert Preprocessor, False Alarm Analyzer로 구성되어 있다. 우리는 실험을 통해 증가하는 오경보를 분석하고, 분석된 오경보 규칙을 침입탐지 시스템에 적용하여 오 경보가 감소됨을 확인하였다.


**키워드 : 침입탐지 시스템, 오 경보, 점진적 데이터 마이닝**

# A Design of false alarm analysis framework of intrusion detection system by using incremental mining method

Eun Hee Kim[†] · Keun Ho Ryu[††]


## ABSTRACT

An intrusion detection system writes a lot of alarms against attack behaviors in real time. These alarms contain not only actual attack alarms, but also false alarms that are mistakes made by the intrusion detection system. False alarms are the main reason that reduces the efficiency of the intrusion detection system, and we propose framework for false alarms analysis in the paper. Also, we apply an incremental data mining method for pattern analysis of false alarms increasing continuously. The framework consists of GUI, DB Manager, Alert Preprocessor, and False Alarm Analyzer. We analyze the false alarms increasingly through the experiment of the proposed framework and show that false alarms are reduced by applying the analyzed false alarm rules in the intrusion detection system.


**Key Words : Intrusion Detection, False Alarm Data, Incremental Data Mining**

## 1. 서 론

지금까지 진행된 침입 탐지 시스템 연구 분야는 많은 연구가와 관련 업체들로 인해 발전을 이루어 왔다. 하지만 여전히 해결되지 않은 문제들이 남아있다. 그 중에 하나가 오 경보에 대한 문제이다. 오 경보에 대한 문제는 침입탐지 시스템의 신뢰성과도 연관이 있는 만큼 중요한 사항이다. 침입 탐지 시스템에서는 공격행위와 이벤트에 대해서 모두 경보를 발생시킨다. 이것은 침입탐지 시스템에서 미리 정의해 놓은 시그네처에 위배되면 어떠한 행위에 대해서도 공격여부로 간주하여 경보를 발생시킨다. 이렇게 잘못 탐지되어 발생된 경보가 전체 경보 중 60% 정도를 차지하고 있어 침입탐지 시스템의 성능을 저하시키는 요인이 되고 있다[22]. 보안 관리자는 전체 경보 중 오 경보를 분석하기 위해 수천 개에 달하는 경보를 직접 분석 해야 하기 때문에 많은 노력과 비용이 요구된다. 뿐만 아니라 불필요한 경보의 발생으로 인한 저장 공간 또한 낭비되고 있다. 최근에는 경보 분석을 위해 로그 데이터 및 경보 상관관계 분석[1, 3, 9, 16], 추론기법[14], 데이터 마이닝 기법[6-8, 10-13, 17-19]을 적용한 다양한 방법들이 제안되고 있다. 그 중에서도 데이터 마이닝 기법을 적용한 경보 간의 상관관계 분석을 통해 공격에 대한 시나리오를 예측하여 침입 탐지 율을 높임으로서 상대적으로 오 경보에 대한 발생률을 감소 시켰었다. 하지만 이들 연구는 오 경보에 대해서 완전히 제거하지는 못하였다. 따라서 오 경보를 줄일 수 있도록 오 경보에 대한 분석을 도와줄 수 있는 보조적인 도구가 필요하며, 또한 점진

※ 이 연구는 정보통신부 대학 IT 연구센터 육성·지원사업의 연구비 지원으로 수행되었음.
† 준 회 원 : 충북대학교 전자계산학과 박사수료
†† 종신회원 : 충북대학교 전기전자컴퓨터공학부 교수
논문접수 : 2005년 12월 31일, 심사완료 : 2006년 4월 20일

적으로 증가하는 오 경보 분석을 통해 지속적인 규칙 관리가 필요하다.

이 논문에서는 오 경보를 감소시키는데 도움을 줄 수 있는 오 경보 분석 프레임워크를 제안한다. 제안된 프레임워크에서는 시간이 지남에 따라 계속 증가하는 오 경보에 대한 패턴을 분석하기 위해 기존의 데이터 마이닝 기법을 기반으로 한 점진적 데이터 마이닝 기법을 적용한다. 아울러 오 경보에 대한 특성을 분석하기 위해 근원지 주소와 목적지 주소간의 연관성을 고려하였다. 구현된 프레임워크를 통해 침입탐지 시스템에서 발생되는 경보 중 오 경보 패턴 분석을 수행하여 실제 침입탐지 시스템에 적용함으로서 오 경보의 발생이 감소되는 것을 확인하였다.

이 논문의 효율적인 구성을 위해 2장에서는 관련연구에 대해서 기술하고, 3장에서는 제안된 오 경보 분석 프레임워크에 대해서 기술한다. 4장에서는 제안된 프레임워크를 통한 실험 및 분석 결과를 보여주며 마지막으로 5장에서는 결론으로 끝을 맺는다.

## 2. 관련연구

오 경보의 패턴 분석을 위해 기존에 연구되었던 기법들은 크게 4가지로 분류 할 수 있다.

첫째, 경보간의 유사성을 기반으로 분류하는 기법[2, 4, 8, 11, 16]이 있다. 이 기법은 경보 간의 유사성을 계산하기 위해 개념적 클러스터링과 일반화 계층 구조를 사용하여 각기 다른 클러스터로 경보를 분류 하는 기법으로서, 경보간의 유사성을 측정하는 척도를 어떻게 결정할 것인가 하는 것이 중요하게 고려된다. 둘째, 미리 정의된 공격 시나리오에 따라 경보의 상관관계를 분석하는 기법[3]이 있다. 이 기법은 각 공격 단계마다 구성된 공격의 알려진 시퀀스 패턴을 이용하여 예측하기 때문에 알려지지 않은 공격 시나리오는 발견할 수 없지만, 새로운 공격 시나리오 가설을 위해 주로 사용 되었다. 셋째, 공격의 사전 조건과 사후 조건을 명시하여 분석하는 기법[1, 9, 15]이 있다. 이 기법은 공격이라고 판단되는 경보와 공격이 발생된 이후에 나타나는 경보들을 각각 사전 조건과 사후 조건이라고 명시하여 공격의 전략을 분석하는 기법으로 공격과 경보에 대한 많은 사전 지식이 필요하다. 여기에서는 공격의 사전 조건이 다음 공격의 사후 조건을 최소 하나 이상 만족 한다면 두 공격에 대한 경보는 서로 연관성이 있다고 본다.

마지막으로 다중의 이질적인 정보 소스로부터 경보의 상관관계를 분석하는 기법[7, 10]이 있다. 이 기법에서 다중의 이질적인 정보는 IDS 센서, 방화벽, 취약성 스캐너, 안티-바이러스 등과 같이 네트워크로부터 수집된 데이터와 시스템으로부터 수집된 모든 데이터를 포함한다. 보안 분석가는 다중 소스로부터 경보 상관관계를 유용하게 분석하기 위해 다양한 정보 시스템의 보안과 관련한 개념 및 관계를 기술한 모델을 사용한다. 특히 이 기법에서는 소스에 의해 제공된 정보는 각 소스들 간의 충돌로 인하여 의미적으로 다르게 제공되기도 하기 때문에 아주 중요한 부분이다. 서로 다른 센서로부터 수집된 경보들은 경보 자체가 침입 탐지 시스템에 매우 의존적이므로 서로 통합 관리하는데 매우 어려움을 겪고 있다. 그 결과로 유사한 공격에 대해서 상이한 경보를 발생 시킬 수도 있고, 상이한 공격에 대해서 유사한 경보를 발생 시킬 수도 있으며, 또한 어떤 침입탐지 시스템에서는 정상이라고 처리하는 것을 다른 침입탐지 시스템에서는 공격이라고 처리하는 경우 또는 그 반대의 경우가 많기 때문이다.

실제 보안 관리자들이 오 경보를 줄이기 위한 방법은 크게 두 가지로 분류 해 볼 수 있다. 첫 번째 각 조직의 보안 정책 및 네트워크 구성에 맞게 침입 탐지 시스템의 환경 설정 및 침입에 대한 규칙의 초기치 값을 설정 하는 것이다. 이 방법은 알려진 공격 행위에 대해서만 적용이 가능하다는 점과 보안 관리자는 모든 경보 발생 원인에 대한 사전 지식을 가지고 있어야 하며, 이들 정보를 통해서만 오 경보를 분석해 낼 수 있다는 단점이 있다. 두 번째는 시스템의 취약성 정보를 이용하여 공격을 판별함으로서 오 경보를 식별해 낼 수 있다. 이 방법은 미리 정의되어 있는 취약성 정보에 어떠한 행위가 시도되었다고 판단될 때 경보를 발생시키는 방법이다. 따라서 이 방법은 취약성에 대한 정보 관리를 통해 오 경보뿐만 아니라 전반적인 경보의 수를 감소시키는 방법으로 실제 관리에 있어서 첫 번째 방법과 같이 사용하면 유용하지만, 비교 하고자 하는 정보가 취약성 정보에 없다면 실제 공격임에도 불구하고 정상으로 간주하는 경우가 초래하게 된다. 위의 두 가지 방법들은 네트워크의 환경설정을 통해 줄이는 방법이기 때문에 수동적인 요소를 배제할 수 없고 보안정책이나 네트워크 구성이 변경되었다면, 수시로 업데이트가 이루어져야하는 부담감을 가지고 있다. 따라서 수동적이고 임의적인 요소를 최대한 배제한 자동화된 분석 기법이 요구된다.

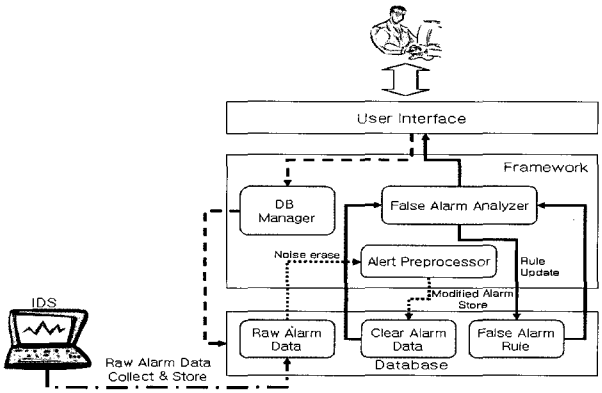
이 논문에서는 이러한 문제점을 해결하기 위해서 점진적 데이터 마이닝 기반의 오 경보 분석 프레임워크를 제안한다. 점진적 데이터 마이닝기법이란 기존의 마이닝 기법을 확장한 것으로서 지속적으로 증가하는 데이터간의 규칙을 유지하는 기법으로 이전에 탐사된 규칙을 유지하면서 데이터가 증가된 이후의 규칙을 탐사하는 방법이다[20].

## 3. 오경보 분석 프레임워크

### 3.1 프레임워크 구조

이 장에서는 오 경보 분석을 위해 제안된 프레임워크의 구조에 대해서 기술한다. 제안된 프레임워크는 GUI, DB Manager, Alert Preprocessor, False Alarm Analyzer로 구성되어 있으며, 저장구조는 침입탐지 시스템에서 발생된 경보를 수집하여 저장하는 테이블인 Raw Alarm Data, 수집된 경보 데이터로부터 오경보로 분석하기 위한 전처리 과정으로서 노이즈 제거, 오류 정정 및 분석에 유효한 속성들만을 추출하여 저장하는 테이블인 Clear Alarm Data, 마지막으로 오경보로

서 분석된 규칙들을 관리하는 테이블인 False Alarm Rule로 구성되어 있다. 아래 (그림 1)은 전체 오 경보 분석 프레임워크를 보여준다.




(그림 1) 오 경보 분석 프레임워크

(그림 1)에서처럼 User Interface는 사용자가 오 경보에 대한 분석 수행 및 규칙 결과를 확인 할 수 있도록 인터페이스 역할을 수행하고, DB Manager는 데이터베이스 내의 모든 테이블에 접속 및 해제와 같은 데이터베이스를 관리하기 위한 모든 작업을 담당한다. Alert Preprocessor는 수집된 경보로부터 오 경보 분석을 위해 관심 있는 속성들만을 추출하여 오 경보 분석 테이블로 저장하는 역할을 수행한다. False Alarm Analyzer는 점진적 마이닝 기법을 기반으로 오 경보 분석 테이블로부터 오 경보 패턴을 분석하는 역할을 한다.

다음절에서는 오 경보 전처리과정과 분석과정에 대해서 자세히 기술한다.

### 3.2 오 경보 전처리 과정

오 경보는 따로 정의되어 있는 것이 아니라 실제 공격 경보 속에 포함되어 있다. (그림 2)는 발생된 경보의 한 예를

```
[**] [1:472:1] ICMP redirect host [**]
[Classification: Potentially Bad Traffic] [Priority: 2]
11/11-16:21:29.856720 210.125.146.126 -> 210.125.145.192
ICMP TTL:63 TOS:0x0 ID:55550 IpLen:20 DgmLen:56
Type:5  Code:1  REDIRECT HOST NEW GW: 210.125.146.1
[Xref => http://www.whitehats.com/info/IDS135]
[Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-1999-0265]
```

(그림 2) 경보 예제

보여준다. Snort[21]라는 공개된 침입탐지 시스템에서 발생시킨 경보 중 ICMP redirect host 라는 경보를 보여준다.

이 경보는 네트워크 호스트가 호스트 다이어그램을 위해 ICMP가 방향을 바꾸고자 할 때, 즉 짧은 경로를 찾고자 할 때 발생된다. 이 행위는 기본 적인 행위로서 공격으로 볼 수 없다. 하지만 침입 탐지 시스템에서는 이러한 행위를 공격으로 간주하여 경보를 발생한다. 이러한 경보들은 IP 별로 디렉토리가 생성되어 전체적으로 통합 및 분석이 어렵다. 우리는 생성된 경보를 통합 및 분석을 용이하게 하기 위해 경보의 모든 속성을 데이터베이스에 저장한 후 분석 이후의 결과는 다시 파일로 저장을 한다. 결과를 다시 파일로 처리하는 이유는 보안 관리자 입장에서 분석이 용이하게 하기 위한 자료로서 제공하기 위함이다.

아래 (그림 3)은 (그림 2)와 같은 형태의 경보를 데이터베이스로 변환하여 원하는 속성만을 추출한 경우이다. 경보가 기본적으로 포함하고 있는 속성은 경보ID, 경보이름, 위험 난이도, 발생날짜/시간, 근원지 주소, 목적지 주소, 프로토콜 프레그먼트 속성, 다른 참조 주소 등으로 구성되어 있으며, 경보마다 약간의 차이는 있다.

이 같은 경보들 중에서 오 경보를 분석하기 위해 우리는 먼저 근원지 주소와 목적지 주소간의 관계성(From ~ To)을 고려한다. 근원지와 목적지가 같은 경보들을 보면 거의 실제 공격이라기보다는 자체적인 이벤트 발생에 의한 오 경보일 확률이 높기 때문이다.

경보데이터의 속성은 네트워크 환경에 매우 의존적이기 때문에 실험을 위해 자체 구성된 네트워크 환경에서 발생된 특성을 기준으로 전처리 과정을 수행 하였다.

| no | sig_name | timestamp | srcip | dstip | protocol |
|---|---|---|---|---|---|
| 233 | ICMPPINGNMAP | 2004-11-11 15:30:06 | 210.115.161.89 | 210.125.145.192 | ICMP |
| 234 | ICMPredirecthost | 2004-11-11 15:33:34 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 235 | ICMPredirecthost | 2004-11-11 15:33:53 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 236 | INFOMSNchataccess | 2004-11-11 15:43:42 | 210.125.145.192 | 207.46.108.16 | TCP |
| 237 | ICMPredirecthost | 2004-11-11 15:43:52 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 238 | ICMPredirecthost | 2004-11-11 15:43:52 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 239 | ICMPEchoReply | 2004-11-11 15:43:52 | 210.125.145.40 | 210.125.145.192 | ICMP |
| 240 | ICMPredirecthost | 2004-11-11 15:43:52 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 241 | ICMPredirecthost | 2004-11-11 15:45:55 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 242 | ICMPredirecthost | 2004-11-11 15:57:55 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 243 | INFOMSNchataccess | 2004-11-11 15:58:05 | 210.125.145.192 | 207.46.108.85 | TCP |
| 244 | INFOMSNchataccess | 2004-11-11 15:58:39 | 210.125.145.192 | 207.46.108.85 | TCP |
| 245 | INFOMSNchataccess | 2004-11-11 15:59:17 | 210.125.145.192 | 207.46.108.85 | TCP |
| 246 | INFOMSNchataccess | 2004-11-11 15:59:24 | 210.125.145.192 | 207.46.108.85 | TCP |
| 247 | INFOMSNchataccess | 2004-11-11 16:00:14 | 210.125.145.192 | 207.46.108.85 | TCP |
| 248 | INFOMSNchataccess | 2004-11-11 16:00:17 | 210.125.145.192 | 207.46.108.85 | TCP |
| 249 | INFOMSNchataccess | 2004-11-11 16:00:21 | 210.125.145.192 | 207.46.108.85 | TCP |
| 250 | INFOMSNchataccess | 2004-11-11 16:01:53 | 210.125.145.192 | 207.46.108.39 | TCP |
| 251 | INFOMSNchataccess | 2004-11-11 16:02:32 | 210.125.145.192 | 207.46.108.39 | TCP |
| 252 | INFOMSNchataccess | 2004-11-11 16:02:59 | 210.125.145.192 | 207.46.108.39 | TCP |
| 253 | INFOMSNchataccess | 2004-11-11 16:03:01 | 210.125.145.192 | 207.46.108.39 | TCP |
| 254 | ICMPredirecthost | 2004-11-11 16:05:34 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 255 | ICMPredirecthost | 2004-11-11 16:09:52 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 256 | ICMPEchoReply | 2004-11-11 16:10:57 | 210.125.145.40 | 210.125.145.192 | ICMP |
| 257 | ICMPredirecthost | 2004-11-11 16:10:57 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 258 | ICMPredirecthost | 2004-11-11 16:21:54 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 259 | ICMPredirecthost | 2004-11-11 16:33:53 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 260 | ICMPredirecthost | 2004-11-11 16:37:34 | 210.125.146.126 | 210.125.145.192 | ICMP |
| 261 | ICMPPINGNMAP | 2004-11-11 16:42:04 | 210.115.161.89 | 210.125.145.192 | ICMP |
| 262 | INFOPossibleIRC... | 2004-11-11 16:42:59 | 210.125.145.192 | 211.115.11.95 | TCP |

(그림 3) 오 경보 분석을 위한 데이터 집합

### 3.3 오 경보 분석

경보는 침입탐지 시스템이 동작하는 동안에는 지속적으로 발생된다. 경보의 수가 점점 증가할수록 보안 관리자는 경보의 패턴 분석뿐만 아니라 포함된 오 경보를 찾아내기에는 역부족이다. 이들 경보로부터 오 경보 패턴을 분석하기 위해서도 오 경보 패턴 규칙 역시 지속적으로 갱신을 시켜주어야 한다.

이 논문에서는 점진적인 오 경보 분석을 위해 몇 가지 요구사항을 만족해야 한다.

첫째, 새로 추가된 항목에 대해서도 빈발 항목을 찾아야 한다. 둘째, 이전 항목에서는 빈발하였지만 항목이 추가된 이후에는 빈발하지 않을 수 있다. 반대로 이전에는 빈발하지 않았지만 항목의 수가 추가된 이후에는 빈발 항목이 될 수 있다. 셋째, 모든 발견된 빈발항목은 전체 항목의 빈발항목이어야 한다. 이러한 조건을 만족하기 위해 비빈발항목집합(Small itemset)이라는 개념을 적용하였다. 비빈발항목집합은 빈발항목 집합(Large itemset)의 반대 개념으로서, 후보 항목에 포함되지 않은 항목들을 버리지 않고 향후 추가될 항목들을 비교하기 위해 저장하는 데이터 구조이다. 추가된 비빈발항목집합을 통해 점진적인 관리를 위한 오 경보에 대한 후보 항목의 수를 감소시킨다.

제안된 프레임워크에서는 점진적으로 증가하는 오 경보 패턴을 분석하기 위해서 점진적 연관규칙 기법을 사용한다. 점진적인 연관규칙 기법은 데이터(D)로부터 주어진 시간(t)에 규칙탐사를 수행하고 분석된 시점(t′) 이후부터 추가적으로 저장된 데이터(D′)에 대해서 새로운 규칙을 탐사하기 위해서 이전의 규칙을 유지하는 기법을 말한다. 이때 가장 고려해야 될 사항이 추가로 분석을 할 경우 전체 데이터베이스를 다시 스캔 할 것인지 아니면 추가적으로 저장된 부분만을 스캔 할 것인지가 중요하다. 전체 데이터베이스를 매번 다시 스캔 한다면 I/O비용은 많이 들겠지만 정확한 분석 결과를 가져 올 것이다. 반대로 추가적인 부분만을 스캔하여 분석한다면 I/O 비용은 줄일 수 있지만 전체 데이터베이스를 스캔 했을 때처럼 결과의 정확성이 다소 감소할 수도 있다.

우리는 점진적으로 갱신되는 데이터 분석을 위해 점진적 연관규칙 기법을 기반으로 확장하여 적용 한다. 이 기법은 새로운 항목이 데이터베이스에 추가될 때 이전에 발견된 규칙을 유지하면서 갱신하는 기법으로서, 갱신이 일어날 경우 이전에 발견된 규칙을 이용하여 갱신된 데이터베이스의 크기를 줄이기 때문에 점진적으로 추가된 데이터를 가지고 분석을 하기 위해 I/O 비용과 계산 비용을 최소화할 수 있다.

오 경보 분석은 최초 규칙을 탐사하는 단계와 오 경보 테이블에 새로운 데이터의 추가가 일어난 후 새로운 규칙을 탐사하는 단계로 나누어 수행된다.

#### 3.3.1 최초 규칙 탐사 단계

최초 규칙 탐사 단계는 기존의 연관규칙 기법과 유사하다. 먼저 오 경보 테이블을 스캔 하여 읽어 들인 후 각 항목에 대해 빈발항목집합을 생성한다. 빈발항목집합이란 최소 지지도(minimum support)를 만족하는 항목만을 추출해 놓은 것을 말한다. 최소 지지도란 전체 오 경보 데이터의 개수 중에 임의의 한 오 경보 데이터가 포함되어 있는 확률 변수를 의미하는 것으로, 분석가의 경험에 따라 임의적으로 지정해 줄 수 있다. 빈발항목집합을 생성하기 위해 먼저 후보 항목 집합(Candidate itemset)을 생성한다. 후보 항목 집합이란 {(item_name$_i$, item_count),.....,(item_name$_n$, item_count)}로서 각각의 항목마다 (item_name, item_count)쌍으로 구성되어 있고, 이것은 각 항목이름과 항목들이 전체 데이터 개수 중에서 발생된 회수를 각각 나타낸다. 생성된 후보 항목 집합은 초기에 입력된 최소 지지도 값과 비교하여 후보항목 집합의 발생회수가 최소 지지도 보다 크면 ,item_count ≥ min_support, 빈발항목집합으로 분류되고, 그렇지 않으면 비빈발항목집합으로 분류된다. 비빈발항목집합 역시 후보항목 집합과 유사한 형태인 {(item_name$_i$, item_count),....., (item_name$_n$, item_count)}로 구성 되어 있다. 생성된 빈발항목집합은 최소 신뢰도(minimum confidence)와 비교하여 최종 규칙을 생성하게 된다. 최소신뢰도란 빈발항목으로 생성된 항목들 중 임의의 항목 A 와 B 가 있다고 가정 했을 때, 항목 A가 발생 했을 때 항목 A 와 항목 B 가 함께 나타날 수 있는 확률을 나타내는 변수이다. 빈발항목집합 내에 포함된 각 항목들의 개수가 최소 신뢰도와 비교하여 최소 신뢰도보다 크면 최종 규칙으로서 생성되고, 그렇지 않으면 역시 비빈발항목집합으로 분류된다.

#### 3.3.2 규칙 갱신 및 유지 단계

이 단계에서는 최초 규칙 탐사 이후부터 추가되는 데이터들에 대해서 지속적으로 규칙을 갱신 및 유지하는 단계이다. 우리는 전체 데이터베이스를 스캔하는 대신 추가된 부분에 대해서만 스캔 한다. 그리고 이전에 생성된 비빈발항목집합과 빈발 항목 집합을 이용하여 새로운 규칙을 갱신 및 유지한다. 추가 된 항목들은 새로운 규칙을 탐사하기 전에 다음 2가지 조건에 만족하는 지를 확인 한다.

• **조건 1)** 추가된 항목이 비빈발항목집합에 포함되어 있을 때

```
for each new item ∈ Small itemset
for each item_count = new item_count +Small itemset.item_count
      for (i=1; Small itemset ∉ Ø; i++) {
        if item_i_count ≥ min_supp
      then add itemi into Large itemset
      else update Small itemset (item_i_count)
}
```

• **조건 2)** 추가된 항목이 빈발항목집합에 포함되어 있을 때

```
for each new item  ∈ Large itemset
for each item_count = new item_count + Large itemset.item_count
   for (i=1; Large itemset ∉ Ø; i++) {
      if item_i_count ≥ min_supp
      then update Large itemset (item_i_count)
      else add itemi into Small itemset
}
```

여기서 new item은 새로 추가된 항목이고, item_count는 항목들의 빈발한 회수를 의미한다.

조건 1은 새로 읽어 들인 항목들이 기존에 이미 발생 되었던 항목인 경우로서 새로 읽어 들인 항목을 가지고 빈발항목집합을 생성하기 전에 비빈발항목집합과 비교를 하여 후보항목집합을 생성하지 않고 빈발항목집합으로 생성될 지 여부만을 파악한다. 새로 추가된 항목이 비빈발집합항목내의 개수를 더하였을 때 최소 지지도 이상을 만족하면 빈발항목집합으로 분류되고 그렇지 않으면 비빈발항목집합에서 해당 항목에 대한 개수만 갱신한다.

조건 2는 새로 읽어 들인 항목들이 빈발항목집합에 포함되어 있는 경우이다. 이 경우에는 빈발항목집합에 포함된 항목의 개수를 더하여 최소지지도를 만족하는 지 여부를 파악한다. 이때 최소 지지도를 만족하면 빈발항목집합의 개수를 갱신하고, 그렇지 않은 경우에는 최초에는 빈발항목집합이었을 지라도 항목이 갱신된 경우에 빈발항목집합이 되지 못하기 때문에 비빈발항목집합으로 분류된다. 여기서 규칙 탐사를 위해 사용된 최소지지도는 초기단계에서 사용하던 지지도와 증가된 데이터 개수에 따라 비례하여 증가된다. 즉, *min_supp(now) = min_supp(old_DB) + min_supp(new_db)*. 초기에 전체 데이터가 100개 일 때 지지도를 20%를 설정했다면, 즉 항목의 빈발 회수가 20번으로, 규칙을 탐사 한 후 새로운 데이터가 20개 더 추가되었다고 하면 20개 추가된 데이터에 대하여 20%의 지지도를 적용한다. 즉 4번의 빈발회수만 만족하면 되는 것이다. 결과적으로 전체 데이터 120개에 대한 각 항목들의 지지도는 24회로 조절되어 규칙을 갱신하게 된다.

## 4. 실험 및 분석

3장에서는 침입탐지 시스템에서 발생되는 많은 경보들 중에서 오 경보를 분류하기 위한 전처리 과정과 오 경보 분석 과정에 대해서 기술했다. 이 장에서는 프레임워크의 구현을 통해 실험 및 분석결과를 보여준다.

### 4.1 실험 환경

우리는 실험을 위해 허브를 중심으로 침입탐지 시스템, 공격 대상 머신, 공격자 머신으로 이루어진 작은 네트워크를 구성하였다. 각 머신의 운영체제는 모두 윈도우즈 기반으로 하였다. 침입탐지 시스템은 Snort라는 공개된 소스를 사용하였고, 우리의 프레임워크는 침입 탐지 시스템 컴퓨터에 같이 설치하였다. 경보 수집을 위해 15일 동안 오전 9시부터 저녁 12시까지 침입탐지 시스템을 운영하여 수집된 경보들 중 3500개를 사용하였다. 오 경보 분석 프레임워크는 자바 플랫폼을 이용하여 구현되었고, 침입 탐지 시스템의 경보 및 오 경보를 관리하기 위한 데이터베이스로서 MySQL을 사용하였다.

### 4.2 결과 분석

우리는 전체 발생된 경보들 간의 근원지 주소와 목적지 주소 등의 연관성을 통해 오 경보를 탐지 하였다. 발생된 경보데이터에 대해 "1) 누가 공격을 제일 많이 하고 있는가? 2) 누가 공격을 제일 많이 당하고 있는가?"를 통해 공격의 의도성을 분석할 수 있기 때문이다. 물론 공격자들은 근원지 주소를 변경하여 공격을 시도하기도 한다. 이런 행위 또한 자주 수행 된다면 근원지 주소를 분석하여 경보 데이터에 대한 발생을 줄여 나갈 수 있을 것이다. 이번 실험은 홈 네트워크 주소로 지정된 컴퓨터의 오 경보를 대상으로 한다.

규칙 탐사를 위한 임계치로서 지지도와 신뢰도에 대한 초기치 값은 여러 번 실험을 거쳐 최소 지지도 50%, 최소 신뢰도 50%로 설정하였다. 전체 경보에서 발생된 근원지 주소는 모두 52개였고, 목적지 주소는 32개였다.

#### 4.2.1 근원지 기반 분석

가장 빈발한 근원지 주소를 분석하는 이유는 오 경보는 자신의 컴퓨터에서 발생한 행위에 대해서도 공격으로 잘못 탐지 할 수 있기 때문에 근원지 주소의 분석이 필요하다. 먼저 전체 탐지된 근원지별 경보 발생에 대한 분석 결과는 <표 1>과 같이 나타났다. 전체 발생된 경보 중 가장 빈발하게 공격을 시도한 곳은 210.125.***.126 주소와 210.125.***.192 주소이다. 여기서 210.125.***.126주소는 공격자 머신이고, 210.125.***.192는 공격 대상 머신이다. 이 결과를 통해서 공격 대상 머신이 공격의 근원지로서 500회 이상의 경보를 발생 시켰다는 것은 자체적으로 발생된 이벤트에 대해 침입으로 간주하고 경보를 발생시킨 것으로 예측 할 수 있다. 물론 다른 관점에서는 우회 공격용으로 사용되었음을 배제 할 수는 없지만, 이 논문에서는 우회 공격용으로 사용되었다는 것은 고려하지 않는다.

〈표 1〉 근원지 별 경보 발생 빈도 수

| 근원지 주소 | 경보 발생 회수 | 근원지 주소 | 경보 발생 회수 |
|---|---|---|---|
| 210.115.***.89 | 51 | 210.125.***.221 | 9 |
| 210.125.***.1 | 78 | 210.125.***.46 | 11 |
| 210.125.***.151 | 11 | 210.125.***.126 | 1995 |
| 210.125.***.192 | 521 | 210.125.***.220 | 24 |
| 210.125.***.2 | 8 | 210.125.***.242 | 24 |
| 210.125.***.210 | 9 | 220.73.***.217 | 45 |
| 210.125.***.40 | 22 | 210.125.***.3 | 24 |

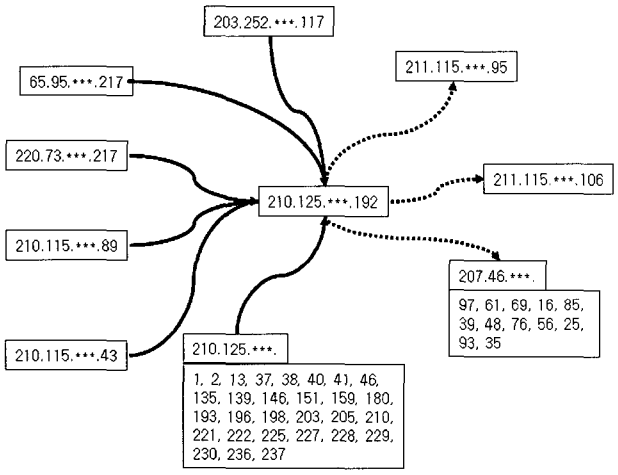
#### 4.2.2 목적지 기반 분석

목적지 주소를 분석하는 이유는 경보에서 목적지 주소의 의미는 공격을 당하는 대상을 말한다. 공격을 당한 회수가 높다는 것은 실제 대상의 취약성이 많다는 의미이거나 아니면 공격자의 공격 행위를 속이기 위한 중간 거점으로 사용되고 있다는 것을 의미하거나 또는 침입 탐지 시스템에서 잘못 탐지하는 수도 있기 때문에 목적지 주소의 분석이 필요하다. 먼저 전체 목적지별 발생된 경보에 대한 분석 결과는 <표 2>와 같이 나타났다.

〈표 2〉 목적지 별 경보 발생 빈도 수

| 목적지 주소 | 경보 발생 회수 | 목적지 주소 | 경보 발생 회수 |
|---|---|---|---|
| 207.46.***.22 | 9 | 207.46.***.61 | 69 |
| 207.46.***.3 | 44 | 207.46.***.63 | 85 |
| 207.46.***.30 | 51 | 207.46.***.69 | 12 |
| 207.46.***.33 | 15 | 207.46.***.88 | 10 |
| 207.46.***.42 | 21 | 207.46.***.93 | 59 |
| 207.46.***.5 | 9 | 207.46.***.97 | 13 |
| 207.46.***.52 | 37 | 210.125.***.192 | 2340 |

위 결과에서 보듯이 가장 많이 공격을 당한 머신은 210.125.***.192로서 공격 대상 머신으로 설정을 해 놓은 것이다. 따라서 근원지와 목적지간의 결과를 통해 근원지와 목적지 간의 관계성을 찾아 낼 수 있었다. (그림 4)는 근원지와 목적지간의 연관성을 나타낸 것으로서 실선은 공격 대상 머신으로 공격을 당한 것을 나타낸 것이고, 점선은 공격 대상 머신이 공격을 시도한 것을 나타낸 것이다. 우리는 공격 대상 머신으로 210.125.***.192(이후 HomeNet)라는 주소를 설정하고, 이 머신으로 공격을 시도하는 패킷과 이 머신으로부터 나가는 패킷들을 분석 할 수 있었다.




(그림 4) 근원지와 목적지 간의 상관관계

우리의 프레임워크를 통해 분석된 결과 중 유사하게 발견된 패턴들 중에 From HomeNet (210.125.***.192)로부터 발생된 경보와 To HomeNet (210.125.***.192)이 공격을 당했을 경우에 발생된 경보의 규칙을 얻을 수 있었다. 규칙의 유형은 연관규칙과 유사하다. 우리는 이들 규칙 들 중 가장 지지도와 신뢰도가 가장 높은 2개의 규칙을 살펴본다. 먼저 To HomeNet의 경우를 나타내는 규칙이다.

- **규칙 1:** 공격 대상 머신이 공격을 당한 경우 (To HomeNet)
  *ICMP redirecthost, 210.125.***.126 --> 210.125.***.192 [513, 100]*

규칙 1은 전체 분석된 오 경보 패턴 중 HomeNet이 공격의 대상이 되어 발생된 경보의 패턴을 보여준다. 위 규칙은 210.125.***.126이라는 근원지에서 210.125.***.192로 ICMP 패킷을 보냈을 경우 발생된 것으로 전체 발생된 경보 중 똑같은 행위가 513회 발생했음을 의미한다. 오 경보 분석을 위해서는 공격과 경보에 대한 사전 지식을 기반으로 분석을 해야 한다. 위 패턴은 침입탐지 시스템에 기술된 시그네처 중 "icmp.rules"과 매칭 하여 발생된 것으로, "icmp.rules"라는 시그네처에는 icmp 프로토콜과 관련하여 발생될 수 있는 공격 시그네처들에 대해서 기술해 놓은 것이다. 대표적으로 DoS 공격이나 라우팅 문제 등으로 인해 발생 할 수 있는 공격들에 대해 기술해 놓은 것이지만, 호스트에서 라우팅 신호에 대한 응답이 없을 경우에도 발생되는 것으로 이러한 경우 오 경보가 된다. 위 경우는 후자의 경우에 해당하므로, Icmp redirecthost 라는 오 경보의 발생을 제거하기 위해 시그네처를 수정해야한다. 하지만 오 경보를 제거하기 위해 시그네처를 수정할 때에는 특별히 주의를 해야 한다. 간혹 실제 공격으로 들어오는 경우를 탐지 못할 수도 있기 때문이다. 이 경보를 제거하기 위해서 위 시그네처를 찾아서 주석(#) 처리를 해 줌으로서 이러한 내용의 경보를 발생시키지 않게 된다. ICMP redirect host라는 경보는 대부분 호스트에서 정상 트래픽인 경우에도 잘못 탐지하여 경보를 발생시킨다. 우리는 위 시그네처를 침입탐지 시스템에 적용한 후 icmp redirect host 라는 이름의 경보 데이터가 발생되지 않음을 확인 하였다.

다른 경우로서 HomeNet 으로부터 출발하여 공격이라고 탐지된 경우들이다. HomeNet 컴퓨터가 근원지가 되어 발생되는 경보 중 오경보를 분석한 규칙이다.

- **규칙 2:** 공격 대상 머신이 공격을 시도한 경우 (From HomeNet)
  *Info MSN chat access, 210.125.***.192 -> 207.46.***.61 [53, 63]*

규칙 2는 전체 분석된 오 경보 패턴 중 HomeNet 으로부터 발생된 패턴에 대해서 유사한 형태를 가지고 "Info MSN chat access"라는 메시지에 대해서 요약을 해 놓은 것이다. 위 패턴은 침입탐지 시스템에 기술된 시그네처 중 "chat.rules"와 매칭 하여 발생된 것이다. "chat.rules"라는 시그네처에는 사용자들이 사용하는 여러 가지 메신저 프로그램들과 관련한 공격 시그네처들에 대해서 기술해 놓은 것이다. Info MSN chat access는 사용자가 MSN 메신저 프로그램을 이용하였을 때 발생된 것으로 MSN 채팅 시 클라이언트가 MSN 서버를 찾기 위해 계속 시도하는 것을 의미한다. 즉, 사용자는 MSN 메신저 프로그램을 이용하였을 뿐인데 공격으로 탐지하여 경보를 발생시킨 것이다. 이렇게 Info MSN chat access라는 이름으로 발생된 경보 데이터가 전체 2870개 중 518개로서 30% 정도의 경보를 발생시키고 있었다. 이 경보의 위험성은 거의 없기 때문에 Misc activity 라고 분류되어 있다. 위 시그네처를 찾아서 주석(#) 처리를 해 줌으로

서 이러한 내용의 경보를 발생시키지 않게 된다. Info MSN Chat access 라는 경보가 대부분 오 경보로 분석되어 경보를 발생 시키지 않지만, 만약이라도 MSN 메신저를 통해서 공격을 시도 했다면, 실제 공격 또한 경보로서 남기지 못할 수도 있다.

위 시그네처를 침입탐지 시스템에 적용한 후 실제 MSN 메신저 프로그램과 관련한 경보데이터가 생성되지 않음을 확인 하였다.

이 논문에서는 다량의 경보들 중에서 오경보의 패턴을 찾아내어 보안 관리자로 하여금 오 경보의 수를 줄이기 위해 도움을 주고자 하는 데 목적이 있다. 하지만, 공격 탐지 율을 높이기 위해서는 많은 시그네처를 포함하여야 하기 때문에 그 만큼 오 경보 발생률이 높아질 수밖에 없다. 따라서 공격 탐지율과 오 경보 발생율의 관계는 보안관리자가 적절하게 수준을 고려를 해야 될 것이다.

## 5. 결 론

이 논문에서는 침입탐지 시스템에서 발생되는 모든 경보들을 지속적으로 확인하여 오 경보 패턴을 검출해 냄으로서 경보데이터의 수를 감소시키고, 침입탐지 시스템의 성능도 향상시킬 수 있는 오 경보 분석을 위한 프레임워크를 제안하였다. 제안된 오경보 분석 프레임워크는 GUI, DB Manager, Alert Preprocessor, False Alarm Analyzer로 구성 되어 있으며, 저장구조는 침입탐지 시스템에서 발생된 경보를 수집하여 저장하는 Raw Alert Data 테이블, 오경보 분석을 위해 노이즈 제거 와 유효한 속성들만을 추출하여 저장하는 Clear Alarm Data 테이블, 그리고 오경보로 분석된 규칙들을 관리하는 False Alarm Rule 테이블로·구성되어 있다.

우리는 점진적으로 갱신되는 데이터 분석을 위해 점진적 연관규칙 기법을 기반으로 확장하여 적용 하였다. 이 기법은 새로운 항목이 데이터베이스에 추가될 때 이전에 발견된 규칙을 유지하면서 갱신하는 기법으로서, 갱신이 일어날 경우 이전에 발견된 규칙을 이용하여 갱신된 데이터베이스의 크기를 줄이기 때문에 점진적으로 추가된 데이터 분석 시 I/O 비용과 계산 비용을 최소화할 수 있었다. 또한 비빈발항목집합을 추가함으로서 규칙 생성 시 발생되는 후보항목의 수도 줄였다.

실험을 통해 가장 빈발한 오 경보 패턴을 탐지 한 후 탐지된 오 경보를 제거하기 위해 관련 시그네처를 찾아 수정하였다. 실제 침입 탐지 시스템에 적용해 본 결과 수정된 시그네처를 적용 한 시점 이후부터 발생된 경보들을 확인한 결과 오 경보로서 탐지된 경보들이 더 이상 발생되지 않음을 확인 하였다.

제안된 프레임워크는 지속적인 오 경보 패턴 분석을 통해 보안관리자로 하여금 실제로 발생된 공격에 대한 대응 정책을 설정하는 데 도움을 주며, 또한 공격에 대한 시그네처를 관리하는 데 도움을 주고자 하는 데 목적이 있다. 하지만 이 기법은 아직까지는 침입탐지 시스템의 효율성을 증가시키는 측면에서 보조적인 도구 이다. 따라서 이러한 기능을 침입탐지 내에서 자동으로 필터링 해 줄 수 있는 연구가 계속 되어야 할 것이다.

## 참 고 문 헌


[1] F. Cuppens, Miege, A, "Alert correlation in a cooperative intrusion detection framework," IEEE Symposium on Security and Privacy, pp.202~215, May., 2002.

[2] D, O, Cunningham, R, "Fusing a heterogeneous alert stream into scenarios," ACM Workshop on Data Mining for Security Applications, pp.1~13, Nov., 2001.

[3] Debar, H., Wespi, A, "Aggregation and correlation of intrusion-detection alerts," Recent Advances in Intrusion Detection, pp.85~103, Oct., 2001.

[4] F. Cuppens, R. Ortalo, "LAMBDA: A language to model a database for detection of attacks," Recent Advances in Intrusion Detection, pp.197~216, Oct., 2000.

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

[6] M. Joshi, R. Agarwal, V. Kumar, PNrule, "Mining Needles in a Haystack: Classifying Rare Classes via Two-Phase Rule Induction," ACM SIGMOD Conference on Management of Data, pp.91~102, May., 2001.

[7] B. Morin, L. Me, H. Debar, and M. Ducasse. "M2D2: A formal data model for IDS alert correlation," International Symposium on Recent Advances in Intrusion Detection, pp.115~137, 2002.

[8] K. Julisch. "Mining alarm clusters to improve alarm handling efficiency," Annual Computer Security Applications Conference, pp.12~21, Dec., 2001.

[9] P. Ning, Y. Cui, and D. S Reeves. "Constructing attack scenarios through correlation of intrusion alerts," ACM Conference on Computer and Communications Security, pp.245~254, Nov., 2002.

[10] P. A. Porras, M. W. Fong, and A. Valdes. "A mission impact based approach to INFOSEC alarm correlation," Recent Advances in Intrusion Detection, pp.95~114, Oct., 2002.

[11] W. Lee. "A Data Mining Framework for Constructing Features and Models for Intrusion Detection System," PhD thesis, Computer Science Department, Columbia University, NY, 1999.

[12] S. Manganaris et al. "A Data Mining Analysis of RTID Alarms," Recent Advances in Intrusion Detection, pp.7~9, Sep., 1999.

[13] F. Provost and T. Fawcett, "Robust Classification for Imprecise Environments," Machine Learning, vol. 42/3, pp.203~231, 2001.

[14] S. Staniford, J. A. Hoagland, and J. M. McAlerney. "Practical Automated Detection of Stealthy Portscans," ACM Computer and Communications Security IDS Workshop, pp.105~136, 2000.
[15] Templeton, S., Levit, K, "A requires/provides model for computer attacks," New Security Paradigms Workshop, pp.31~38, 2000.
[16] A. Valdes, "Probabilistic Alert Correlation," Recent Advances in Intrusion Detection, pp.54~68, 2001.
[17] Moon Sun Shin, Eun Hee Kim, Keun Ho Ryu, Ki Young Kim. "Data Mining Methods for Alert Correlation Analysis," IJCIS, 2003.
[18] Moon Sun Shin, Eun Hee Kim, Keun Ho Ryu, "False Alarm Classification Model for Network Based Intrusion Detection System," Intelligent Data Engineering and Automated Learning, pp.259~265, May., 2004.
[19] 신문선, 김은희, 문호성, 류근호, 김기영, "데이터 마이닝 기법을 이용한 경보데이터 분석기 구현", 정보과학회논문지, 제31권, 1호, 2004.
[20] J. Han, Y. Cai, and N. Cercone, "Data driven discovery of quantitative rules in relational databases," IEEE Transactions on Knowledge and Data Engineering, pp.29~40, 1993.
[21] Snort. Open-source Network Intrusion Detection System. http://www.snort.org
[22] K. Julisch. "Dealing with False Positives in Intrusion Detection," In 3nd Workshop on Recent Advances in Intrusion Detection, 2000.

**김 은 희**

e-mail : ehkim@dblab.chungbuk.ac.kr
2001년 삼척대학교 정보통신공학과(공학사)
2003년 충북대학교 전자계산학전공 (이학석사)
2003년~현재 충북대학교 전자계산학과 박사수료
관심분야 : 접근제어, 네트워크 침입 탐지 시스템, 무선 센서네트워크 보안

**류 근 호**

e-mail : khryu@dblab.chungbuk.ac.kr
1976년 숭실대학교 전산학과(이학사)
1980년 연세대학교 전산전공(공학석사)
1988년 연세대학교 전산전공(공학박사)
1976년~1986년 육군군수 지원사 전산실 (ROTC 장교),한국전자통신연구원 연구원, 한국방송대학교 전산학과 조교수 근무
1989년~1991년 Univ. of Arizona Research Staff (TempIS 연구원(Temporal DB)
1986년~현재 충북대학교 전기전자컴퓨터공학부 교수
관심분야 : 시간데이터베이스, 시공간 데이터베이스, Temporal GIS 및 지식기반 정보검색 시스템, 데이터마이닝 및 데이터베이스 보안, 바이오인포매틱스, 유비쿼터스 컴퓨팅과 스트림 데이터 처리 기술