점진적 연관 규칙을 이용한 침입탐지 시스템의 오 경보 패턴 분석 프레임워크 설계

전원용[0] 김은희 신문선 류근호
충북대학교 데이터베이스 연구실
{ chonwy2000[0], ehkim, msshin, khryu }@dblab.chungbuk.ac.kr

A design of framework for false alarm pattern analysis of intrusion detection system using incremental association rule mining

Won Yong Chon[0], Eun Hee Kim, Moon Sun Shin, Keun Ho Ryu
Database Laboratory, Chungbuk National University

요 약

침입탐지시스템에서 발생되는 오 경보는 false positive 와 false negative 로 구분된다. false positive 는 실제적인 공격은 아니지만 공격이라고 오인하여 경보를 발생시켜 시스템의 효율성을 떨어뜨리기 때문에 false positive 패턴에 대한 분석이 필요하다. 오 경보 데이터는 시간이 지남에 따라 데이터의 양뿐만 아니라 데이터 패턴의 특성 또한 변하게 된다. 따라서 새로운 데이터가 추가될 때 마다 오 경보 데이터의 패턴을 분석할 수 있는 도구가 필요하다. 이 논문에서는 오 경보 데이터로부터 false positive 의 패턴을 분석할 수 있는 프레임워크에 대해서 기술한다. 우리의 프레임워크는 시간이 지남에 따라 변하는 데이터의 패턴 특성을 분석할 수 있도록 하기 위해 점진적 연관규칙 기법을 적용한다. 이 프레임워크를 통해서 false positive 패턴 특성의 변화를 효율적으로 관리 할 수 있다.

1. 서 론

네트워크 기반 침입 탐지 시스템에서 여전히 해결되지 않은 문제는 False alarm을 어떻게 줄일 것인가 하는 문제이다. False alarm 은 침입 탐지 모델에서정상 행위를 침입으로 잘못 탐지하여 발생하는 경우를 의미한다. 이러한 다량의 False alarm 발생은 네트워크 전반에 걸친 보안 서비스의 질을 하락시키는 원인이 되며, 침입탐지 모델의 정확한 침입 판별 능력을 저하시켜 공격자들의 공격으로부터 시스템 또는 네트워크를 보호할 수 없게 된다.

따라서 침입탐지 시스템에서 발생되는 모든 경보들을 수시로 확인하여 False Alarm 패턴을 검출해 냄으로서 경보 데이터의 수가 줄어들고 침입탐지 시스템의 성능도 향상시킬 수 있는 툴이 필요하다. 경보데이터는 지속적으로 추가되며, 데이터의 변화에 따라 경보데이터의 특성도 지속적으로 변화한다. 이 문제를 해결하기 위하여 점진적인 마이닝 기법의 오 경보 패턴 분석 프레임 워크를 제안한다.

이 논문의 효율적인 전개를 위해 다음과 같이 구성 한다. 2장에서는 경보데이터 분석기법과 점진적 기법에 대하여 기술하고, 3장에서는 점진적 연관 규칙에 대하여 자세히 기술한다. 4장에서는 실험평가를 통해 점진적 연관규칙 모델이 기존의 연관규칙보다 빠른 규칙 생성의 결과를 보여준다. 마지막으로 결론으로 끝을 맺는다.

2. 관련연구

이 연구는 한국전자통신연구원의 정보보호연구원의 연구비 지원으로 수행되었음

침입탐지시스템의 성능 향상을 위해서 오 경보 발생을 최소화하여 성능을 향상하기 위한 연구들이 많이 진행되어 왔다[3]. 대표적인 연구로는, 시스템 로그 데이터 상관관계 분석[4], 경보 상관관계[5][6]분석 연구, 마이닝 기법을 이용하여 경보 최소화를 위한 연구[1][2]들이 진행되었다. 이들 연구의 특징은, 경보데이터 상관관계 분석은 선택된 속성에 의존적이며, 경보 데이터 간의 인과 관계를 완벽하게 탐사하기에 적당하지 못하다는 단점을 가진다. 데이터 마이닝 기법은 많은 양의 데이터로부터 알려지지 않은 유용한 지식을 추출해내는 기술이며 데이터 필터링의 효과도 있으므로 알려지지 않은 공격에 대한 시퀀스추출에 활용 가능하며 또한 많은 양의 경보데이타를 감소시키는 역할을 할 수 있다. 이 논문에서는 점진적 마이닝 기법을 이용하여 오 경보 분석 프레임 워크를 제안 한다.

점진적 연관규칙은 지속적인 데이터의 변화에서 이전에 마이닝 되어있는 규칙을 유지 시키는 기술이다. 추가된 데이터에의 발견된 규칙에 의해서 기존의 규칙이 유용하지 않을 수 있으며, 유용하지 않았던 이전의 규칙이 새로운 데이터의 추가로 인하여 유용한 규칙이 될 수 있다. 점진적 연관 규칙이란 이렇게 변하는 규칙에 대하여 기존 데이터에 대한 최소한의 연산과 변환하는 데이터의 연산을 통하여 규칙을 유지시키는 기술이다[6].

3. 오 경보 분석을 위한 프레임 워크

이 논문에서 설명한 오 경보 패턴 분석 프레임워크는 점진적 연관 규칙 기법을 적용하였다. 오 경보 데이터는 시간이 지날수록 데이터의 양 뿐만 아니라 데이터 패턴 특성 또한 변하게 된다. 새로운 data가 추가 되어 새로

운 패턴특성을 찾기 위해서는 기존의 기법은 대량의 데이터에 대하여 지속적인 마이닝을 하였다. 하지만 점진적 마이닝 기법은 추가된 데이터와 마이닝 된 데이터의 결합으로 그 규칙을 생성한다. 아래 그림1 은 오 경보 패턴 분석 프레임워크에 대한 전체 프레임 워크를 나타낸 것이다.

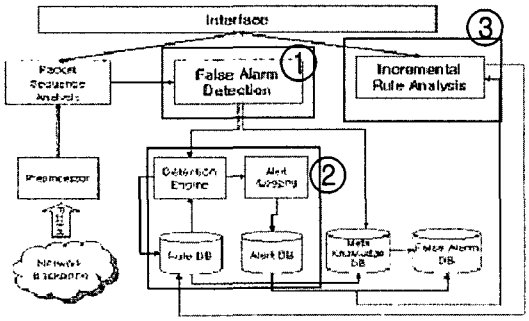

그림 1. False Positive 분석 프레임워크

전체적인 프레임 워크는 침입탐지 시스템(2), 오 경보 데이터에 대한 False Positive 추출(1), 오 경보 데이터의 분석단계(3)로 나누어진다. 오 경보 데이터 분석단계에서는 입력 데이터 중에서 분류하고자하는 false positive 데이터를 분류하여 분석단계로 넘겨주며, 침입탐지 시스템에서(2)는 모든 데이터들에 대한 경보를 발생하며 공격을 탐지하는 과정을 지속적으로 진행한다. 분류된 데이터에 대하여 점진적인 분석을 단계 (3) 에서 진행한다.

3.1 점진적 규칙 갱신 모듈 : 1- large item

표 1은 이 논문에서 사용하는 용어의 정의 이다.

표 1. 용어 정의

표기	설명	표기	설명
DB	초기의 데이터베이스	db	추가된 데이터베이스
C(DB)	DB 의 후보항목	C(db)	db 의 후보항목
L(DB)	DB 의 빈발항목	L(db)	db 의 빈발항목
U:DB U db	갱신된 데이터베이스	C': DB U db	갱신된 데이터베이스 후보항목
L':DB U db	갱신된 데이터베이스 빈발항목	Support(s)	지지도

점진적 연관 규칙은 발견된 규칙을 유지 시키는 기법으로 다음과 같은 특성을 가진다. 처리하는 빈발 항목을 줄일 수 있으며, DB 의 빈발항목도 U 의 빈발 후보항목이 될 수 있으며, db 의 후보항목도 U 의 빈발 후보 항목이 될 수 있다. 또한, 새로운 L' 를 생성하기 위해서 모든 데이터 후보항목을 검색해야한다.

점진적 연관 규칙 분석 과정은 추가 전 데이터에 대한 마이닝 규칙과 최소한의 추가 전 데이터에 대한 검색 을 시도하여 데이터 추가 후의 마이닝 결과를 결합하는 방식이다. DB와 db 그리고, U 에 대한 연관 규칙은 DB 와 db 의 연관 규칙의 결합이다. 그러므로 , DB에 대한 마이닝 규칙을 유지하여 마이닝 하는 것이 U에 대하여 모두 마이닝 하는 것보다 효율적이고 빠른 결과를 얻을 수 있다. 하지만 이전 데이터의 마이닝된 빈발 항목을 이용하므로 이전항목과 같은 지지도를 가져야 하는 단점과 점진적 마이닝을 위한 이전 마이닝 결과를 유지하는 단점을 가지고 있다.

연관규칙을 찾는 방법은 최소지지도 이상을 만족하는 빈발 아이템을 찾는 과정과, 빈발 아이템에서 최소 신뢰도를 만족하는 연관 규칙을 찾는 두 과정으로 나누어 생각할 수 있다. 두 번째 단계는 빈발항목을 검색해서 빠르게 찾을 수 있으므로 많은 연구가 빈발항목을 빠르게 검색하는 문제를 해결하고 있다[7]. 점진적 마이닝 기법도 많은 후보 항목이 생성되고, 많은 시간이 소비되는 빈발항목을 찾는 과정을 효율적으로 재구성 한다. 점진적 연관 규칙의 빈발항목 생성 과정은 1- large item 생성 과정과 n-large items 을 이용한 n+1 large item을 생성하는 과정으로 나눌 수 있다.

아래 그림과 같이 DB는 마이닝 되어 연관 규칙과 각 아이템 크기에 따른 빈발항목 그리고 후보항목을 유지하고 있다.

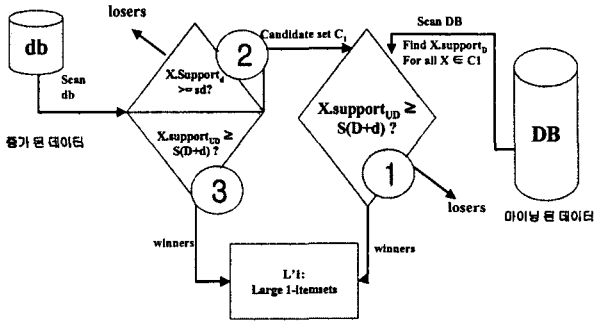

그림 2. 점진적 규칙 갱신 모듈 : 1- large item

U의 빈발 항목은 D와 d에 대한 지지도를 만족하는 항목이다(S(D+d)). 먼저 DB에 대한 빈발항목은 U에서 빈발하거나 빈발하지 않을 수 있다. 그러므로 DB 의 빈발항목은 db 의 빈발, 후보항목과 결합하여 U에서의 빈발 여부를 검사한다(1). 다음으로 DB에 대한 후보 항목을 검사한다(2,3). DB의 후보항목은 U에서 빈발항목이 되거나 후보 항목이 될 수 있다. 그러나 DB 에서의 후보 항목이었으므로 db 의 후보 항목과 결합해서는 빈발 항목이 될 수 없고, DB 의 빈발 항목과의 지지도를 결합하여 S(D+d)를 만족해야만 빈발 항목이 될 수 있다. DB 의 후보 항목에 대해서는 db 에서의 빈발, 후보 항목 여부를 조사한 후 (2) 빈발항목에 대해서만 DB 항목과 결합하여 S(D+d)를 만족하는 항목에 대해서 빈발항목으로 한다(3). 위 그림과 같이 1-item 의 빈발항목은 DB 의 마이닝 된 빈발, 후보 항목과 db의 각 항목의 count 수를 비교하여 찾는다.

3.2 점진적 규칙 갱신 모듈 : n- large item

n- large itme 생성 과정은 이전 항목의 빈발 항목으로 후보 항목을 생성하여 그 후보 항목이 전체 데이터에 대하여

빈발 항목인지 후보 항목인지를 검사하여 전체 데이터에 대한 빈발 항목을 검사 하는 단계이다. C_n 은 U 에서의 빈발 도를 만족해야 U의 L'_n이 될 수 있다.

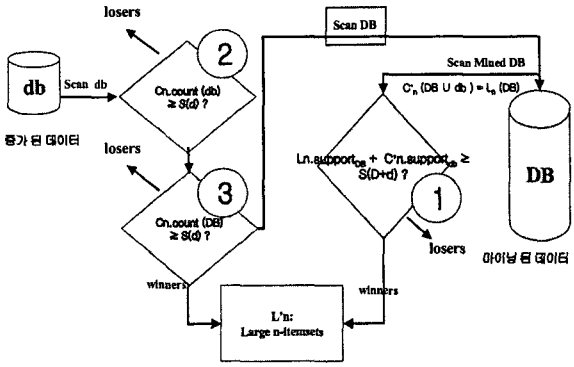

그림 3. 점진적 규칙 갱신 모듈 : n- large item

이전 항목의 빈발항목으로 생성한 C_n 은 U 에서의 빈발 도를 만족해해야 U의 L'_n이 될 수 있다. C_n 이 DB 의 빈발항목인 경우에는 DB 의 빈발도와 db 의 빈발도 (L_n.Support + C'_n.Support)를 S(D+d) 와 비교하여 L_n 여부를 결정한다(1). C_n 이 DB 의 빈발항목인 아닌 경우에는 db 에서의 C_n 의 빈발도를 조사하여 S(d) 이상인 경우만을(2) db 의 빈발도와 DB 의 빈발도를 결합하여 S (D+d) 와 비교하여 L_n 여부를 결정한다(3). 이전 항목에서의 빈발하지 않은 항목이므로 두 항목 모두를 조사하여 빈발 도를 비교 하여야 한다. 이 과정을 반복하여 더 이상의 빈발항목이 생성되지 않을 때까지의 빈발항목을 생성한다. 마이닝 되어 있는 DB 에 대한 새로운 연산(Scan DB)을 최소화 하는 것이 효율적인 방법이 된다. U에 대한 연관 규칙 생성 과정이 아닌, db 의 마이닝된 규칙과 규칙생성과정의 데이터, 그리고 DB에 대한 최소한의 연산을 이용하므로 효율적으로 규칙을 유지할 수 있다.

4. 실험평가

기존의 연관 규칙 마이닝과 점진적 기반의 연관 규칙 마이닝의 속도를 비교하는 실험을 진행 하였다. 생성된 데이터는 3장에서 설명한 분류기를 통하여 분류된 false positive 데이터로서 점진적인 패턴 분석 데이터 이다.

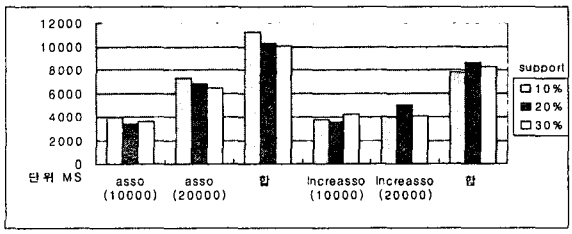

그림 4. 점진적 연관규칙 수행속도 비교

그림에서 보여주는 바와 같이 점진적 연관 기법과 이전의 연관규칙 기법을 비교하여 속도와 정확도를 비교한 결과 정확도는 유사한 반면 마이닝 속도에서 제안된 기법이 빠른 것을 나타냈다. 점진적 연관 규칙은 전체 데이터 크기에 상관없이 추가된 양에 따른 빠른 마이닝을 수행한다. 그 이유는 전체 데이터가 아닌 증가된 데이터에 대한 마이닝과 이전 마이닝 결과를 결합하는 마이닝 기법이기 때문이다. 증가 전후의 데이터의 크기가 같지만 속도가 다른 것은, 이전데이터(DB)에 대한 검색 정도의 차이로 나타난다.

5. 결론

침입탐지 시스템은 지속적으로 많은 경보를 발생 시키며, 발생되는 경보데이터는 시간이 지남에 따라 데이터의 양이 증가하며 그 패턴 또한 지속적으로 변하게 된다. 오 경보의 효율적인 관리를 위해서 시간에 따라 변하는 패턴을 분석할 도구가 필요하다. 이 논문에서는 침입탐지 시스템에서 발생되는 오 경보 패턴을 분석하기 위한 프레임 워크를 구현하기 위하여 기존의 마이닝 기법을 확장하여 점진적 마이닝 기법을 적용하였다. 기존의 마이닝 기법을 적용한다면 데이터의 변화에 따라 매번 데이터베이스를 스캔해야 하기 때문에 많은 비용이 소모되고, 대량의 데이터를 지속적으로 검색하는 많은 중복 작업이 진행된다. 하지만 점진적 데이터 마이닝 기법은 증가분에 대한 마이닝을 이전 마이닝 결과와 결합하므로 패턴들을 효율적으로 관리 할 수 있다. 구현된 프레임워크의 실험을 통해 침입탐지 시스템에서 발생되는 경보 데이터 중 오 경보 패턴 분석을 수행하여, 그 결과로서 침입탐지 시스템의 효율성을 향상 시킬 수 있음을 보였다.

참고문헌

[1]W.Lee, S.J.Stolfo,K. W. Mok "A Data Mining Framework for Building Intrusion Detection Models," 2001.
[2]W.Lee, S.J.Stolfo. "Data Mining Approaches for Intrusion Detection," Columbia University, Computer Science Department, 1998.
[3]K. Julisch. "Dealing with False Positives in Intrusion Detection". In 3nd Workshop on Recent Advances in Intrusion Detection,2000
[4]Cuppens, F., Miege, A. "Alert correlation in a cooperative intrusion detection framework". In Proceedings of the IEEE Symposium on Security and Privacy, 2002
[5]H. Debar, A.Wespi, "Aggregation and Correlation of Intrusion-Detection Alerts", In Recent Advances in Intrusion Detection, number 2212 in Lecture Notes in Computer Science, p 85-103, 2001
[6]M. Lin, S. Lee,"Incremental Update on sequential Patterns in large databases ", Proceedings of the Tools for Artificial Intelligence Conference, 1998
[7]R. Agrawal , R. Srikand, "Fast Algorithms for Mining Association Rules in Large Databases" In Proc, 20th Int. Conf. On Very Large Databases, 1994