업무 활동간 연관 관계를 이용한 데이터 유출 시나리오 정의 방법

서민지*, 김명호*
*숭실대학교 융합소프트웨어학과
e-mail: porito2@ssu.ac.kr, kmh@ssu.ac.kr

How to Define a Data Leakage Scenario Based on Relationships Between Work Activities

Min Ji Seo*, Myung Ho Kim*
*Dept. of Software Convergence, Soongsil University

요 약

기업에서 보유하는 기밀 정보가 내부 직원에 의해 유출되는 사고가 빈번하게 발생하고 있다. 기업에서 데이터를 유출하려는 내부 직원을 탐지하기 위하여 보안 로그를 분석해주는 보안 관제 시스템을 사용하고 있으나, 보안 관제 시스템은 관리자가 지정하는 기준에 대해 보안 로그를 분석하기 때문에 새로운 유형의 데이터 유출 사고가 발생하였을 때 제대로 데이터 유출을 탐지할 수 없는 문제점을 가진다. 따라서 본 논문에서는 내부 직원의 업무활동에서 발생하는 보안 로그 리스트에 연관 분석을 적용하여 새롭게 데이터 유출 탐지 시나리오를 작성하여 기존의 시스템이 가진 문제점을 해결할 수 있는 방법을 소개한다. 연관 분석을 활용하여 정의한 데이터 유출 탐지 시나리오를 활용한 결과, 결과적으로 데이터 유출 탐지 성능이 향상되었다.

1. 서론

기업에서 고객 정보와 같은 중요한 기밀 정보를 데이터베이스나 서버에 저장하게 되면서, 데이터 유출 방지의 중요성이 대두되었다. 기업에서는 데이터 유출을 방지하거나 탐지하기 위하여 내부 직원의 업무 활동에서 발생하는 보안 로그를 분석하는 시스템을 활용하고 있으나, 새로운 유출 패턴으로 데이터를 유출하는 내부 직원의 경우 시스템이 데이터 유출을 제대로 탐지하지 못하고 있는 실정이다[1].

따라서 본 논문에서는 과거 데이터 유출을 일으킨 직원에게서 수집한 보안 로그 집합에 연관 분석 알고리즘인 Apriori 알고리즘[2]을 적용시켜 새롭게 데이터 유출 시나리오를 작성하는 방법을 제안한다. 기존에 관리자에 의해 정의된 데이터 유출 시나리오와 더불어 보안 로그 간의 연관 분석을 통해 새롭게 데이터 유출 시나리오를 작성하여 내부 직원의 업무 활동을 감시하기 때문에, 데이터 유출 여부를 더 정확하게 판별해 줄 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 기존에 연구 된 데이터 유출 탐지 방법을 소개하고, 3장에서는 본 논문 에서 제안하는 데이터 유출 시나리오 정의 방법을 설명한다. 4장에서는 제안하는 시스템의 실험을 통해 성능을 검증하고, 5장에서 결론을 내린다.

2. 관련 연구

2.1 패킷 간 연관 관계를 이용한 네트워크 비정상 행위 탐지

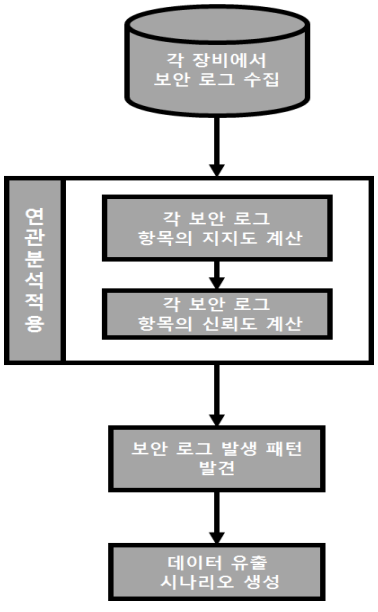
네트워크를 통한 침입 여부 및 비정상 행위를 탐지하기 위하여 네트워크 패킷 분석에 연관 분석 알고리즘을 활용하는 방법이 연구되었다[3].

패킷 간 연관 관계를 이용한 네트워크 비정상 행위 탐지 방법은 정상적으로 업무 활동을 한 내부 직원의 네트워크 패킷을 수집하여, 각 행위별로 네트워크 패킷을 분류시킨다. 네트워크 패킷의 분류가 완료되면 각 그룹별로 연관 분석을 적용하여 자주 발생 하는 네트워크 패킷의 집합 중에서 상호 연관성이 높은 네트워크 패킷 그룹을 추출하여 정상 행위에서 나타나는 네트워크 패킷 그룹으로 정의시킨다. 네트워크 침입 여부를 분석하기 위한 네트워크 패킷 집합이 입력되면, 시스템에서는 정상 행위 패턴과의 유사도를 비교하여 비정상행위 여부를 판별한다.

기존의 방법은 정상적인 네트워크 패킷 그룹과의 유사도를 통해 침입 여부를 탐지하였다. 하지만 내부 직원이 업무 중에 할 수 있는 행위의 종류가 다양하지 때문에, 정상적인 행위도 비정상 행위로 탐지할 수 있는 문제점을 가진다.

3. 업무 활동간 연관 관계에 따른 데이터 유출 시나리오 정의 방법

본 논문에서는 과거의 데이터 유출 사고에서 수집한 보안 로그에 Apriori 알고리즘의 적용을 통해 각 보안 로그 간의 연관 관계를 생성한다. Apriori 알고리즘이란, 데이터들에 대한 발생 빈도를 기반으로 데이터 간의 연관 관계를 분석하기 위한 알고리즘을 의미한다. 제안하는 시스템에서는 먼저 보안 로그 간의 연관 관계를 분석하기 위해 과거 데이터 유출 사고를 일으킨 내부 직원에게서 발생한 보안 로그 항목의 지지도(support)를 계산한다. 지지도는 총 생성된 보안 로그 그룹 중에 연관 관계를 분석할 보안 로그 항목이 발생하는 빈도를 나타내며, 지지도의 계산을 통해 데이터를 유출할 때 주로 나타나는 행동을 파악할 수 있다. 지지도는 최소 지지도를 만족하지 못할 경우 지지도 계산 후보에서 제거되며, 최소 지지도는 관리자에 의해 지정될 수 있다. 지지도의 계산이 완료되면, 신뢰도(Confidence)의 계산을 통해 각 보안 로그 집합이 서로 연관 되는 정도를 분석한다. 신뢰도는 데이터를 유출을 위해 어떤 행동을 하였을 때 또 다른 행동을 했을 확률을 나타내며, 지지도와 마찬가지로 최소 신뢰도를 만족하지 않는 보안 로그 집합은 제거시킨다. 지지도 및 신뢰도의 계산은 더 이상 계산할 수 있는 빈발 항목 보안 로그 집합이 없을 때까지 반복 된다.

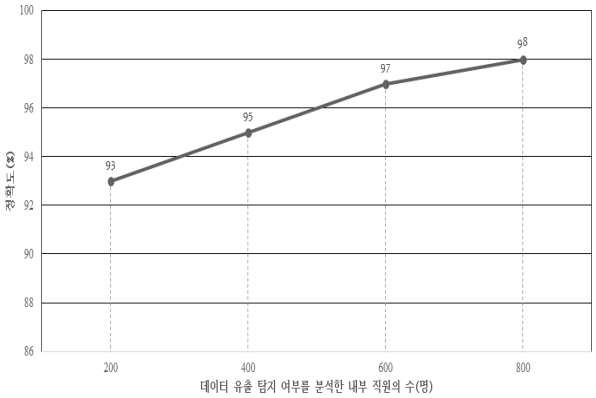


(그림 1) 데이터 유출 시나리오 생성 방법

4. 성능평가

4장에서는 데이터 유출이 발생하는 시간과 상황에 따른 데이터 유출을 판별하는 실험을 통해 정확성을 검증하였다. 정확도는 전체 데이터 유출 직원 중에서 제안하는 방법을 통해 작성한 데이터 유출 시나리오를 통하여 판별된 유출 직원의 수를 의미한다.

제안하는 방법으로 데이터 유출 시나리오를 작성하여, 데이터 유출 여부를 분석할 직원의 수를 늘려가면서 정확도를 측정하였다. 정확도 측정 결과는 그림 2와 같은 결과를 보인다.



(그림 2) 데이터 유출 시나리오를 활용한 데이터 유출 탐지 정확도 측정 결과

5. 결론

본 논문에서는 기업 내부 직원에 의한 데이터 유출을 정확하게 탐지기 위해, 내부 직원의 업무 활동에서 발생한 보안 로그를 분석하기 위한 데이터 유출 시나리오를 정의하는 방법을 제안한다.

본 논문에서는 연관 분석 알고리즘을 이용하여 연관성이 높은 보안 로그 항목의 집합을 데이터 유출 시나리오로 정의하였다. 따라서 제안하는 방법을 이용한 데이터 유출 시나리오를 활용할 경우 기존에 관리자에 의해 정의되지 않은 데이터 유출 패턴을 탐지할 수 있어 데이터 유출 탐지를 더 정확하게 탐지할 수 있다. 향후 연구로는 다양한 방면으로 연관 분석을 활용하여 좀 더 효율적으로 보안 로그를 분석할 수 있도록 연구할 계획이다.

Acknowledgements

이 논문은 2016년도 중소기업청 첫걸음 기술개발 사업(C0394819)에서 지원받았음

참고문헌

[1] “Global Data Leakage Report, H1 2016”, InfoWatch Analytical Center, 2016. Available: https://infowatch.com/report2016_half

[2] C. Borgelt and R. Kruse, “Induction of Association Rules: Apriori Implementation,” in Proceedings of the 15th Conference on Computational Statistics, Berlin, Germany, 2002, pp. 395-400.
[3] S. H. Oh and W. S. Lee, “Network Anomaly Detection based on Association among Packets,” Journal of The Korea Institute of Information Security and Cryptology, vol. 12, no. 5, pp. 63-73, Oct. 2002.