시계열 그래프를 이용한 내부 데이터 유출 탐지 시스템

서민지*,신희진*,김명호*,박진호**
*숭실대학교 융합소프트웨어학과
**숭실대학교 소프트웨어학부
e-mail: porito2@ssu.ac.kr,shinhj0213@ssu.ac.kr,
kmh@ssu.ac.kr,j.park@ssu.ac.kr

Internal Information Leakage Detection System using Time Series Graph

Min Ji Seo*, Hee Jin Shin*, Myung Ho Kim*, Jin Ho Park**
*Dept. of Software Convergence, Soongsil University
**Dept. of Software, Soongsil University

요 약

최근 데이터 기술의 발달에 따라, 기업에서는 중요 데이터를 서버와 같은 데이터 저장 장치에 보관하고 있다. 하지만 기업 내부 직원에 의해 기업의 기밀 데이터가 유출될 수 있는 위험성이 있기 때문에, 내부 직원에 의한 데이터 유출을 탐지 및 방지해야 할 필요성이 있다. 따라서 본 논문에서는 각 보안 솔루션에서 수집한 보안 로그를 데이터 유출 시나리오를 바탕으로 시계열 그래프로 작성하여, 이미지 인식에 뛰어난 성능을 보이는 합성곱 신경망을 통해 데이터 유출을 탐지하는 시스템을 제안한다. 실험 결과 유출된 데이터의 크기에 상관없이 95% 이상의 정확도를 보였으며, 복합적인 행동을 통해 데이터 유출을 시도한 경우에도 97% 이상의 정확도를 보였다.

1. 서론

최근 데이터 처리 기술이 발달하게 되면서, 기업에서 는 개인 정보와 같은 중요 데이터를 서버와 같은 데이터 저장 장치에 보관하고 있다. 데이터 저장 방식의 변화는 대용량의 정보를 쉽게 관리할 수 있는 장점을 가지나, 기업의 내부 직원에 의해 쉽게 외부로 유출될 수 있는 단점을 가진다. InfoWatch Analytical Center에 따르면, 기업의 정보가 유출되는 사고가 급증하고 있으며, 정보 유출 사고의 60% 이상이 재임 중인 직원에 의해 발생하고 있다고 밝혔다[1].

따라서 본 논문에서는 각 보안 솔루션에서 수집한 보안 로그를 데이터 유출 시에 나타날 수 있는 시나리오별로 시계열 그래프를 작성하여, 데이터 처리에 뛰어난 성능을 보이는 딥 러닝(Deep Learning)[2]에서 이미지 인식에 뛰어난 성능을 보이는 합성곱 신경망(Convolutional Neural Network)[3]을 통해 데이터를 유출한 내부 직원의 시계열 그래프와 유사한 패턴을 보이는지 판별해주는 시스템을 제안한다.

제안하는 시스템은 기업 내부 직원의 행동 이력을 그래프로 이미지화하여, 이미지 인식에 뛰어난 성능을 보이는 합성곱 신경망을 통해 데이터 유출 여부를 판별하였기 때문에, 대용량의 그래프 이미지를 빠른 시간 내에 기업 내부 직원의 데이터 유출 여부를 정확하게 판별할 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 기존에 연구된 데이터 유출 탐지 방법을 소개하고, 3장에서는 본 논문에서 제안하는 시스템을 설명한다. 4장에서는 제안하는 시스템의 실험을 통해 성능을 검증하고, 5장에서 결론을 내린다.

2. 관련연구

2.1 빅 데이터를 이용한 보안정책 개선에 관한 연구

빅데이터를 이용한 데이터 유출 탐지[4]는 다음과 같은 과정으로 이루어진다.

먼저, 기업 내부 직원의 업무 활동과 관련된 보안 로그를 군집화 알고리즘을 통해 정상 직원의 보안 로그 그룹과 데이터 유출 직원의 보안 로그 그룹으로 분류시킨다. 이후 시스템에 입력된 직원의 보안 로그 패턴이 데이터 유출 직원의 보안 로그 그룹과 높은 유사도를 보일 경우, 데이터 유출이 의심되는 직원으로 인지하여, 해당 직원의 행동 패턴에 따라 적절한 대응을 할 수 있도록 한다.

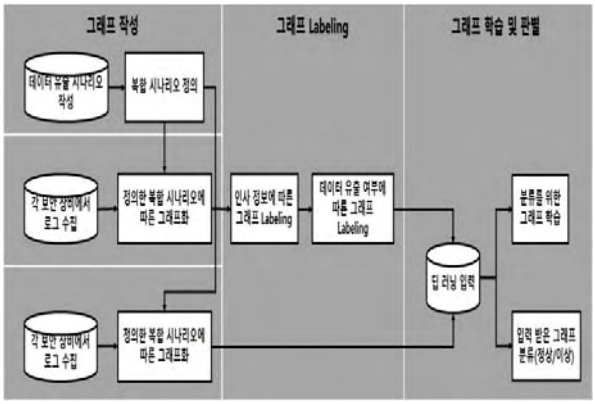
빅 데이터를 이용하여 데이터 유출을 탐지하는 경우, 데이터 유출 직원이 오랜 시간에 걸쳐 일정량의 데이터를 유출했을 때 정상적인 업무활동을 해 온 직원과 같은 군집으로 분류되어 데이터 유출 판별의 정확성이 떨어질 수 있다는 위험성을 가진다. 따라서 제안하는 시스템에서는 직원의 업무 활동을 데이터 유출 시나리오에 따라 그래프

로 작성하여, 합성곱 신경망을 통해 기존의 업무 활동 그래프와 비교하기 때문에 정확하게 데이터 유출을 판별할 수 있도록 한다.

3. 합성곱 신경망 기반의 데이터 유출 탐지 시스템

기업 내부 직원에 의한 데이터 유출 사고가 빈번하게 일어나면서, 기업 내부 직원을 감시해야 할 필요성이 증가하고 있다. 따라서 본 논문에서는 각 보안 솔루션에서 수집한 보안 로그를 정의한 데이터 유출 시나리오를 기반으로 시계열 그래프로 작성하여, 합성곱 신경망으로 입력한 기업 내부 직원의 데이터 유출 시나리오 그래프가 데이터 유출을 시도한 내부 직원의 데이터 유출 시나리오 그래프와 높은 유사도를 보이는지 판별해주는 시스템을 제안하며, (그림 1)과 같은 흐름으로 진행된다.

제안하는 시스템에서는, 그래프를 이용하여 이미지 인식에 높은 성능을 보이는 합성곱 신경망으로 데이터 유출 여부를 판별하여, 빠른 시간 내에 데이터 유출 여부를 파악할 수 있도록 하였다. 또한, 기업 내부 직원의 데이터 유출 시도를 정의한 데이터 유출 시나리오를 기반으로 보안 로그를 시계열 그래프로 표현하였기 때문에, 데이터 유출이 의심될 경우에 데이터 유출 시나리오 그래프를 이용하여 데이터를 유출한 시간대와 데이터를 유출하게 된 경위를 파악할 수 있도록 하였다.



(그림 1) 시계열 그래프를 이용한 유출 탐지

4. 성능평가

제안하는 시스템은 Docker 기반 Tensorflow 프레임워크에서 작업하였으며, 제안하는 시스템의 성능을 평가하기 위해 직원 1000명에 대하여 각각의 데이터 유출 시나리오 그래프를 생성하였다. 생성한 그래프 중에서 800명의 데이터 유출 시나리오 그래프는 학습용으로 사용하였고, 200명의 데이터 유출 시나리오 그래프는 테스트용으로 사용하였다.

제안하는 시스템에서는, 데이터 유출이 발생하는 시간과 상황에 따른 데이터 유출을 판별하는 실험을 통해 정확성을 검증하였다. 정확도는 합성곱 신경망에서 분류한 그래프의 데이터 유출 시나리오 클래스 번호와 실제 그래프가 나타내는 데이터 유출 시나리오 클래스 번호가 일치한 횟수를 총 실험한 그래프 수로 나누어 계산하도록 한다.

먼저, 여러 시간에 걸쳐 일정한 양의 데이터를 유출한 경우 95%의 정확도를 나타내었고, 특정 시간대에 집중적으로 데이터를 유출한 경우엔 99% 정확도를 나타내었다.

또한, 작성한 데이터 유출 시나리오 그래프를 데이터를 유출하기 위한 행동의 횟수를 변화시켜가며 정확성을 실험한 결과에서는 데이터를 유출하기 위한 행동의 개수와 상관없이 97% 이상의 데이터 판별 정확도를 보였다.

5. 결론

본 논문에서는 기업 내부 직원에 의한 데이터 유출을 빠르게 탐지하고 대응하기 위하여, 각 보안 솔루션 별로 수집한 기업 내부 직원의 보안 로그를 데이터 유출 시나리오를 통해 작성한 그래프를 딥 러닝인 합성곱 신경망에 입력시켜 데이터 유출 여부를 판별하는 시스템을 제안한다.

제안하는 시스템은 딥 러닝 알고리즘을 이용하여, 기업 내부 직원의 데이터 유출 여부를 정확하게 판별할 수 있고, 그래프를 통해 데이터를 유출한 시간대에 데이터를 유출하기 위해 시도한 행동을 쉽게 파악할 수 있어 데이터 유출에 빠르게 대처할 수 있게 되었다. 향후 연구로는 그래프를 좀 더 효율적으로 분석할 수 있도록 다양한 그래프를 개발할 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] Y. Lecun, Y. Bengio and G. Hinton, “Deep Learning,” Nature, vol. 521, no. 7554, pp. 436-444, May. 2015.

[3] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella and J. Schmidhuber, “Flexible, High Performance Convolutional Neural Networks for Image Classification,” Proc. of the Twenty-Second International Joint Conference on Artificial Intelligence, pp. 1237-1242, 2011.

[4] S. Y. Kim, J. Kim, J. I. Lim and K.H. Lee, ″A study on the security policy improvement using the big data,″ Journal of The Korea Institute of Information Security & Cryptology, vol. 23, no. 5, pp. 969-976, Oct. 2013.