기계 학습을 적용한 이상 공격 탐지에 대한 연구

서지원, 안선우, 이영한, 방인영, 백윤흥*
*서울대학교 전기정보공학과
e-mail : jwseo@sor.snu.ac.kr

A Study on Anomaly Attack Detection with Machine Learning

Ji-Won Seo, Sun-Woo Ahn, Young-Han Lee, In-Young Bang, Yun-Heung Paek*
*Department of Electrical and Computer Engineering, Seoul National University.

요 약

기계 학습은 인간의 지능을 아직 일부만 모델링하여 활용하는 기술임에도 불구하고 다양한 기술 분야에서 새로운 가능성을 열어주는 미래 시장의 핵심이다. 상용 네트워크 보안 시스템은 특정 규칙들을 정해 놓고 규칙에 어긋난 정보에 대하여 보안 위험이 있을 수 있다고 판단을 한다. 하지만 규칙을 잘 정의해 놓은 시스템에서 보안 위험이라고 경보가 나는 경우의 80% 이상이 일반적으로 오탐이다. 상용 네트워크 보안 시스템에 기계 학습을 활용하면 사람이 규칙으로 정의하기 어려운 정보의 내재 의미를 스스로 학습하여 분류에 활용할 수 있다. 본 연구에서는 이처럼 네트워크 공격 중 이상 공격 탐지에 기계 학습을 활용한 연구들에 대해 살펴보도록 하겠다.

1. 서론

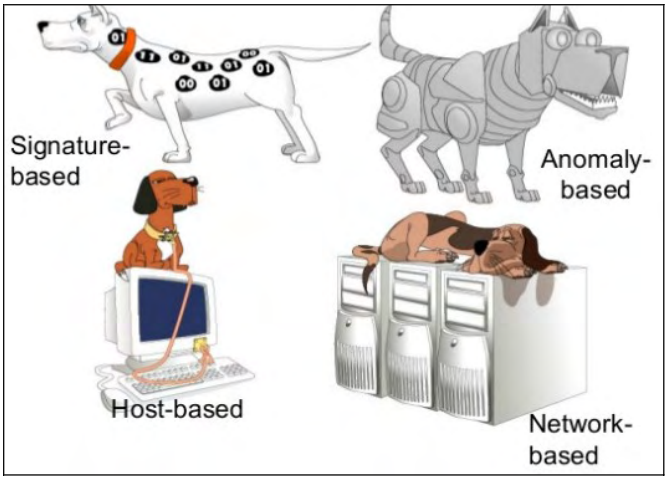
기계학습은 다양한 기술 분야에서 새로운 가능성을 열어 주는 미래 시장의 핵심 기술이다. 이미 알고 있는 정보로부터 그 안에 내재되어 있는 의미를 학습하여 새로운 정보에 대해 그 내재된 의미를 분별하는 사람의 지능을 모방한 기술로 정보를 분류하는 등의 판단에 굉장히 유용하게 활용될 수 있다. 보안의 영역에서 이와 같이 내재되어 있는 의미를 파악하여 분류하여 판단하는 행위는 백신, 침입탐지시스템 등 다양한 분야에서의 핵심적인 요소에 속하기에 보안 분야에 기계학습을 도입하려는 연구는 전세계적으로 행해지고 있다.

기존 침입탐지시스템은 정보를 분석하여 분류 및 판단을 하는 보안 분야에서는 전통적으로 각 정보에 내재되어 있는 의미에 따라 다르게 분류를 하도록 하는 특정 규칙들을 정해 놓고 이에 어긋나는 정보에 대해서 보안 위험이 있을 수 있다고 판단을 하게 된다. 하지만 현재 통용되는 시스템 및 프로그램들은 복잡도가 높기 때문에 규칙들이 완벽하게 상황을 파악하는 것이 불가능하다. 그러므로 보안 프로그램에서 규칙에 어긋난 정보에 대해 보안 위험을 알리게 되면 보안 전문가가 직접 그와 관련된 정보를 모아서 수작업으로 분석을 해야만 한다. 여기에서 문제는, 프로그램에 규칙을 잘 설정해 놓은 최신의 상용 보안 시스템에서도 보안 위험이라고 경보가 나는 경우의 80% 이상이 일반적으로 오탐이며 이는 전문가에게 시간 낭비가 된다. 본 연구에서는 이러한 규칙 기반의 보안 프로그램의 오탐율을 보정하기 위해 기계 학습 혹은 딥 러닝을 통한 정보 분류를 추가하는 연구들을 살펴볼 것이며 이를 위하여 먼저 침입탐지시스템 (Intrusion Detection System) 에 대해 소개하고, 침입탐지시스템 분석 기준 방법에 대해 서술한 후 기존 연구에서 제시한 방어들을 소개한 뒤 결론을 제시하도록 하겠다.

2. 침입탐지시스템 (Intrusion Detection System)

침입 탐지 시스템은 탐지 대상 시스템이나 네트워크를 감시하여 비인가 되거나 비정상적인 행동을 탐지하여 구별하는 시스템이다 이러한 침입 탐지 시스템은 자료 수집 위치에 따라 호스트 기반 침입 탐지 시스템 (Host-based IDS) 와 네트워크 기반 탐지 시스템 (Network-based IDS) 로 분류할 수 있다. 호스트 기반 탐지 시스템은 시스템 자원에 침입 탐지 시스템을 설치하는 것으로 호스트의 자원 사용 실태를 분석하여 탐지한다. 네트워크 기반 탐지 시스템은 네트워크 자원에 침입 탐지 시스템을 설치하여 네트워크상의 모든 트래픽에 대한 패킷을 분석하여 탐지한다. 침입탐지 방식에 따라 오용탐지 (Misuse Detection IDS) 와 이상탐지 (Anomaly Detection IDS) 로 분류할 수 있다[1]. 오용탐지는 공격에 해당하는 특정 규칙들을 정해 놓고 이에 해당하는 패턴을 탐지하였을 때 보고하는 방식이다. 이상탐지는 정상에서 벗어나는 행위를 탐지하는 것으로 정해진 모델을 벗어나는 경우를 침입으로 간주한다. 오용탐지는 탐지 오류율이 낮고 비교적 효율적이라는 장점이 있으나 제로 데이 공격과 같이 알

려지지 않은 공격에 대해서는 탐지가 어렵다는 단점이 있다. 이상탐지는 사전에 공격에 대한 특정 지식이 없어도 제로 데이 공격을 탐지 할 수 있다는 장점을 갖고 있으나 False positive 가 높다는 단점을 갖고 있다. 본 연구에서는 침입 탐지 시스템 중 이상 탐지에 기계 학습을 적용시킨 기존 논문들에 대해 소개할 것이다.



(그림 1) 호스트 기반 침입 탐지 시스템 및 네트워크 기반 침입 탐지 시스템

3. 비정상 공격 탐지의 학습 데이터에 대한 연구

3.1 비정상적인 페이로드 기반 네트워크 침입 탐지[2]

본 논문은 페이로드 기반 이상 탐지기인 PAYL 을 구현하여 연결 포트, 인 바운드/아웃 바운드 기준으로 분류하여 학습한다. 이상 탐지에 대한 학습을 하기 위한 로그는 침입 탐지 시스템 로그에 기록된 페이로드를 n-gram 방식으로 분할한다. 이때 실제 나타난 n-gram 의 상대적 빈도수를 feature 벡터로 사용한다. 각 feature 는 하나의 페이로드에서 256 개의 가능한 바이트 값 중 나타날 수 있는 발생 빈도를 나타낸다. Feature 의 평균 및 표준 편차를 계산하여 정상적인 통신의 모델을 구성할 수 있다. 따라서 정상적인 통신 모델과 테스트 중인 페이로드 간의 Mahalanobis distance 가 이전에 설정한 임계 값을 초과하면 해당하는 테스트 페이로드를 비정상적인 것으로 간주한다. 본 논문을 통해 학습을 위한 데이터 셋을 미리 추가적으로 분류하는 것과 2 장에서 소개한 이상탐지의 falser positive 가 높다는 점을 보완할 수 있는 방안으로 본 논문에서 제시한 데이터인 페이로드로 학습하기 적합하다고 생각한다.

3.2 Anagram: 모방 공격을 탐지 할 수 있는 비정상 탐지기[3]

본 논문은 Anagram 을 제시하여 페이로드 내 비 정상적인 바이트 순서와 그 위치를 감지 할 수 있다. Anagram 모델은 고효율의 Bloom 필터를 사용하는 것으로 구현되는데 필터 1 에는 정상적인 패킷으로부터 추출된 개별적인 n-grams 을 Bloom 필터 b1 에 저장한다. 필터 2 에는 기존에 알려진 공격으로부터 추출된 개별적인 n-grams 를 저장한다. 테스트 단계 동안 각 패킷에 대해 모든 개별적은 n-grams 이 페이로드에서 추출되고 필터 1 및 필터 2 와 비교한다. 이때 필터 1 에 존재하지 않거나 필터 2 에 존재하는 n-gram 이 너무 많은 페이로드는 비정상적인 페이로드로 분류하여 이상 공격에 대해 탐지한다.

4. 기계학습을 이용한 비정상 공격에 대한 연구

기계학습을 네트워크 보안에 접목시키면 앞장에서 제시한 문제인 기존 보안 솔루션에서 잡아내지 못하던 공격 유형에 대한 감지가 가능할 뿐 아니라 높은 false alarm rate 을 감소시킬 수 있다. 다음에서 이러한 기계학습을 접목시킨 기존 연구들에 대해 소개하도록 하겠다.

4.1 이상 탐지를 위한 비지도 이기종 로그 기반 프레임워크[4]

본 논문에서 제시하는 UHAD 모델은 다양한 형태의 로그 데이터 마다 비지도 클러스터링을 적용한다. 학습을 하는데 사용되는 입력 값으로는 각 연결마다 가지고 있는 feature 의 수가 다르기 때문에 아래 <표 1> 과 같이 방화벽 로그를 연결 특징 (TCP, UDP, ICMP) 에 따라 분리된 파일로 저장하여 이벤트들을 저장한 후 사용하였다. 본 논문에서는 다양한 형태의 로그 데이터를 통합적으로 활용하기 위해 클러스터링을 적용하였다. 이때 가장 적은 양을 차지하는 클러스터를 악의적인 그룹으로 가정하고 공통적인 feature 공간으로 매핑함으로서 입력 벡터를 생성한다. 이러한 공통적인 feature 공간을 다시 한번 클러스터링 한 후 그룹별로 규칙 기반 IP 와 포트 번호 분석을 적용하였다. 클러스터링 기법으로는 전통적인 k-means, EM, FF 를 적용하였으며 k-means 가 전체적인 성능이 가장 좋은 것으로 나타났다.

<표 1> 이상 탐지 학습에 선택된 Features

Log type	Selected features
Access	Date, Time, IPClient, ClientRequestLine, StatusCode, ObjectSize, Agent
Error	Date, Time, Severity, ClientIP, Msg
SSL_Error	Date, Time, Severity, Msg
TCP	Date, Time, Direction, PHYSIN, PHYSOUT, LEN, TOS, PREC, TTL, ID, DF, SPT, DPT, WINDOW, RES,STATUS, URGP, SRC, DST
UDP	Date, Time, Direction, PHYSIN, PHYSOUT, LEN, TOS, PREC, TTL, ID, DF, SPT, DPT, LEN, SRC, DST
ICMP	Date, Time, LEN, TOS, PREC, TTL, ID, DF, id, seq ,SRC, DST
Message	Date, Time, Daemon, PID, Operation, User, Tty, UID, EUID, Remotehost, Systemmessage
Mail	Date, Time, From, To, Daemon, Mailer, Stat, Priority, Protocol, Message_ID, Relay, Control_address, DSN, Queue_ID, Messages_queued, Messages_delivered, Bytes_queued, Bytes_delivered
Security	Date, Time, Daemon, PID, Operation, User, Source, Systemmessage
Snort IDS	Date, Time, Relenumber, Rule, Classification, Priority, Protocol, SourceIP, Sourceport, DestinationIP, Destinationport

4.2 시계열에서 비정상 탐지를 위한 장기 단기 메모리 네트워크[5]

본 논문은 시계 열 데이터 처리에 적합한 Recurrent 모델을 활용하여 정상상태를 모델링한 방법론이다. Recurrent sigmoid layers 를 적용한 Stacked LSTM Networks 는 미래의 시간 단계에서 발생할 수 있는 비정상적인 반응을 감지할 수 있다. 비정상적이지 않는 데이터를 사용하여 정상 상태일 때의 확률 분포로서 학습하며 이는 시간의 행동이 비정상적일 가능성을 계산하는데도 사용할 수 있다.

4.3 다중 센서 이상 탐지를 위한 LSTM 기반 인코더-디코더[6]

본 논문은 비정상 탐지를 위해 인코더-디 코더 모델을 적용하였다. LSTM 구조를 기반으로 하는 인코더-디코더 모델을 연속적인 시계 열 데이터에 적용하여 비정상 공격을 탐지하는데 사용하였다. 인코더-디코더 모델을 이용하여 정상적인 데이터들의 구조를 학습하였으며 재구성 오류를 기반으로 하는 비정상적인 점수를 계산하여 비정상적인 데이터를 판별하였다. 결과적으로 높은 정확도를 얻었으며 좋은 F-score 를 획득하였으나 recall 은 모두 실험에서 10 % 미만으로 나타났다. 본 연구를 통해 비정상적인 공격을 탐지하는 측면에서 LSTM 모델의 우수성과 시계 열 데이터에 대한 장기간 의존성을 보존하는 능력을 확인 할 수 있다.

5. 결론

본 논문에서는 기계 학습을 네트워크 보안 시스템에 접목시킨 기존 연구들에 대해서 살펴보았다. 기존 규칙 기반의 보안 시스템에서는 탐지하지 못하였던 비정상적인 공격을 기계 학습을 적용하면 식별이 가능하다. 이때 학습 모델에 대한 데이터로 알려진 공격 유형에 대한 데이터를 학습시킬 경우 알려진 공격 유형이나 유사한 공격에 대한 탐지에는 탁월한 성능을 가지나 여전히 새로운 공격에는 취약하다는 단점을 갖는다. 하지만 정상 데이터를 학습할 경우 제로데이 공격과 같은 새로운 공격에 대해 판별이 가능하나 정상 데이터를 확보하는 데 있어서 어렵거나 앞서 설명한 공격 유형에 대한 데이터를 학습한 경우에 비해 오탐율이 높다는 단점을 갖는다.

최근 공격들은 다양한 패턴을 갖고 있고 심지어 공격을 하는데 있어 기계 학습을 적용시킨 공격 연구들도 많이 이루어지고 있다. 기존 보안 솔루션에서는 이와 같은 공격을 탐지하는데 어려움이 있기 때문에 이에 따른 기계 학습을 접목시킨 탐지 방법에 대한 추가적인 연구가 이루어져야 할 것이다.

6. ACKNOWLEDGEMENT\

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었으며 (IITP-2017-2015-0-00403), 2017 년도 정부(미래창조과학부)의 재원으로 정보 통신 기술 진흥 센터의 지원을 받아 수행된 연구 (No.2016-0-00078, 맞춤형 보안 서비스제공을위한클라우드기반지능형보안기술개발) 및 2017 년도 두뇌 한국 21 플러스 사업에 의하여 지원되었음.

참고문헌

[1] Specification-based anomaly detection: a new approach for detecting network intrusions. In: Proceedings of the Ninth ACM Conference on Computer and Communications Security; 2002. p. 265–74.

[2] K. Wang, J. J. Parekh, and S. Stolfo. Anagram: A content anomaly detector resistant to mimicry attack. In *Proceedings of the International Symposium on Recent Advances in Intrusion Detection (RAID)*, 2006.

[3] Wang, K., Parekh, J.J., Stolfo, S.J.: Anagram: A content anomaly detector resistant to mimicry attack. In: Proceedings of the 9th International Symposium on Recent Advances in Intrusion Detection (2006)

[4] Hajamydeen AI, Udzir NI, Mahmod R, Ghani AAA. An unsupervised heterogeneous log-based framework for anomaly detection. Turkish Journal of Electrical Engineering and Computer Sciences 2014. doi:10.3906/elk-1302-19.

[5] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff and Puneet Agarwal, ”Long Short Term Memory Networks for Anomaly Detection in Time Series”,European Symposium on Artificial Neural Networks, vol 23, Computational Intelligence and Machine Learning, Belgium, 2015.

[6] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, “LSTM-based Encoder-Decoder for Multisensor Anomaly Detection,” in Presented at ICML 2016 Anomaly Detection Workshop, New York, NY, Jul. 2016. [Online]. Available: https://arxiv.org/abs/1607.00148