CNN과 Kibana를 활용한 호스트 기반 침입 탐지 연구

박대경*, 신동규*, 신동일*
*세종대학교 컴퓨터공학과
dkpark@sju.ac.kr, shindk@sejong.ac.kr, dshin@sejong.ac.kr

Host-based intrusion detection research using CNN and Kibana

DaeKyeong Park*, Dongkyoo Shin*, Dongil Shin*
*Dept. of Computer Engineering, Sejong University

요 약

사이버 공격이 더욱 지능화됨에 따라 기존의 침입 탐지 시스템(Intrusion Detection System)은 기존의 저장된 패턴에서 벗어난 지능형 공격을 탐지하기에 적절하지 않다. 딥러닝(Deep Learning) 기반 침입 탐지는 새로운 탐지 규칙을 생성하는데 적절하다. 그 이유는 딥러닝은 데이터 학습을 통해 새로운 침입 규칙을 자체적으로 생성하기 때문이다. 침입 탐지 시스템 데이터 세트는 가장 널리 사용되는 KDD99 데이터와 LID-DS(Leipzig Intrusion Detection-Data Set)를 사용했다. 본 논문에서는 1차원 벡터를 이미지로 변환하고 CNN(Convolutional Neural Network)을 적용하여 두 데이터 세트에 대한 성능을 실험했다. 평가를 위해 Accuracy, Precision, Recall 및 F1-Score 지표를 측정했다. 그 결과 LID-DS 데이터 세트의 Accuracy가 KDD99 데이터 세트의 Accuracy 보다 약 8% 높은 것을 확인했다. 또한, 1차원 벡터에 대한 데이터를 Kibana를 사용하여 데이터를 시각화하여 대용량 데이터를 한눈에 보기 어려운 단점을 해결하는 방법을 제안한다.

1. 서론

사이버 공격이 진화함에 따라 공격자들은 알려지지 않은 취약점을 악용하고 지능적으로 다양해지고 있다. 점점 다양해지는 공격을 방어하는 것은 매우 중요한 문제인데, 대표적인 솔루션 중 하나는 침입 탐지 시스템(Intrusion Detection System)이다. 침입 탐지 시스템은 네트워크 기반인 NIDS(Network-based Intrusion Detection System), 호스트 기반인 HIDS(Host-based Intrusion Detection System) 두 가지 방식으로 나눌 수 있다. 네트워크 기반 침입 탐지 시스템과 달리 호스트 기반 침입 탐지 시스템은 시스템 내부와 외부를 전체적으로 관찰해야 하는 어려움 때문에 연구가 많이 부족하다. 또한, 침입 탐지 시스템에는 두 가지 유형이 있다.[1] 하나는 알려진 시그니처를 기반으로 공격을 탐지하는 오용탐지이고, 이상 탐지는 오용탐지 방법과 반대로 정상적인 사용 패턴을 기반으로 비정상 행위를 탐지하는 이상 탐지이다. 오용탐지는 알려지지 않은 공격을 탐지하기 어렵지만, 이상 탐지는 알려지지 않은 공격을 탐지 할 수 있다는 장점이 있다. 그러나 이상 탐지는 다양한 정상적인 사용 패턴을 정의하기가 어려우므로 오경보율이 증가한다는 문제점이 있다.[2] 최근 딥러닝 기술의 발전과 많은 연구로 인해 다양한 ICT(Information & Communication Technology) 분야 및 IoT(Internet of Things) 분야에 적용되어 지능형 서비스들이 제공되고 있다. 이러한 기술 발전과 적용에 따라 보안 분야에서는 침입 탐지 시스템에 적용되는 사례가 있다. 딥러닝은 심층 신경망을 통해 자체 기능을 학습하여 앞서 말한 약점을 보완하는 기술이다. 침입 탐지 시스템에 딥러닝을 사용하면 침입 탐지 시스템의 단점을 보완 할 수 있다. 즉, 기계 학습(Machine Learning)과 딥러닝은 자체적으로 이상 행위를 학습하고 정상적인 사용 패턴을 결정하여 오경보를 줄일 수 있다. 현재 다양한 연구들이 비정상 행위를 탐지하기 위해 딥러닝을 침입 탐지 시스템 연구에 사용한다. DARPA(Defense Advanced Research Projects Agency)에서 개발한 KDD99 데이터 세트는 침입 탐지 시스템 평가에 가장 많이 사용되는 데

이터 세트이다. KDD99는 공격을 DoS, U2R(User to Root), R2L(Remote to Local) 및 Probe와 같은 네 가지 범주로 분류한다. 기계학습이 침입 탐지 시스템 연구에 많이 사용되고 있어서 수많은 침입 탐지 시스템 연구에서 KDD99 데이터 세트를 사용하고 있다. 하지만 KDD99 데이터는 너무 오래된 컴퓨터 시스템의 특징과 공격 패턴으로 이루어져 있어 현재 사용하기에는 적합하지 않다.[3, 4] 2018년에 공개된 LID-DS 데이터 세트는 기존 공개되었던 데이터들과 다르게 현재 공개된 데이터 세트들보다 최신 컴퓨터 시스템의 다양한 특징들과 공격 방법 및 시나리오로 구성되어 있다.[5]

본 논문에서는 두 데이터 세트에서 비정상 행위에 대한 딥러닝 기반 탐지 모델을 제안한다. 제안하는 모델은 CNN을 기반으로 하며 1차원 벡터에 대한 데이터들을 이미지로 변환하여 다중 클래스 분류하는 방법과 CSV 파일을 ElasticSearch에 업로드하여 Kibana를 사용하여 데이터를 보기 좋게 시각화하는 방법을 제안한다.

2. 관련 연구

침입 탐지 시스템은 공격 패턴에 대한 매칭을 이용하여 위협을 탐지하고 차단하는 시스템이다.[6, 7]

KDD 및 UNM 데이터 세트는 공개적으로 사용이 가능한 데이터이며 침입 탐지 시스템의 검증 기초가 되고 성능 테스트의 기준이 되어 많은 연구가 진행되고 있다. 하지만 일부 네트워크 정보 중에서 시스템 호출을 통해 프로세스와 커널 간에 전달되는 데이터 형식으로 호스트에서 수집된 추적을 제공하는데 기존의 데이터들은 더는 현대적인 특징을 가지고 있지 않기 때문에 최신 컴퓨터 시스템의 다양한 특징들과 사이버 공격 특징들이 반영되지 않아 새로운 데이터가 필요하다.[8, 9, 10]

박정민 등[10]은 데이터를 테이블 형태뿐만 아니라 그래프를 통해 데이터를 한곳에 모으고 알아보기 쉽게 Kibana를 사용하여 시각화했다.

Laskov 등[11]은 Decision Tree, K-NN(K-Nearest Neighbor), MLP(Multi-Layer Perceptron), K-means, SVM 등 여러 기계학습 알고리즘을 침입 탐지에 적용했고, 각 알고리즘을 ROC(Receiver Operator Characteristic) 곡선을 통해 비교했다.

Kim 등[12]은 침입 탐지 시스템에서 SVM과 K-NN 같은 기계학습 알고리즘이 높은 오경보율을 보이는 문제점을 해결하기 위하여 딥러닝 알고리즘을 사용한 연구를 했다.

Kim 등[13]은 비정상 행위 기반의 호스트 침입 탐지 시스템을 설계하는 데 있어, LSTM-Based System-Call Language Modeling 방법을 제안하였다. 기존 방법들에서 자주 발생하는 높은 오탐율(False-Alarm Rate) 문제를 해결하기 위해서, 저자는 새로운 앙상블(Ensemble) 방법을 사용했다.

Ravipati 등[14]은 LID-DS 데이터 세트의 특징과 가장 유사한 KDD99 데이터 세트를 이용하여 8가지의 기계학습 알고리즘을 실험한 결과로 성능 평가 및 오탐율 수치를 보여주었다.

CNN을 기반으로 바이너리 및 여러 범주의 공격을 탐지하는 수많은 연구도 진행되고 있다.[15, 16]

Khan 등[17]은 침입 탐지 모델을 얻기 위해 기계학습 알고리즘을 사용할 때의 단점을 지적했다. 또한, CNN 기반 네트워크 침입 탐지 모델과 소프트맥스 알고리즘을 결합하는 방법을 제안했다. KDD를 사용하여 제안된 모델을 평가하고 실험 결과는 모델이 SVM 및 DBN (Deep Belief Network) 알고리즘보다 침입 탐지에 더 효율적임을 보여주었다.

Upadhyay 등[18]은 41개의 KDD 기능 중에서 무작위로 선택된 36개의 기능과 함께 KDD를 사용했다. 데이터 세트를 1 × 6 크기의 이미지로 변환한 다음 나머지 기능을 다른 변수에 저장하여 CNN 모델을 학습시켰으며, 실험 결과 제안된 모델의 침입 탐지 오류가 2% 미만인 것으로 나타났다.

3. CNN 기반 침입 탐지 시스템 설계

3.1 이상 행위 이미지 생성

KDD99는 41개의 트래픽 기능과 각 데이터가 속한 위치를 결정하는 1개의 기능으로 구성된다. 41개의 트래픽 특성 중 38개는 숫자 특성으로 표시되고 3개는 기호 특성으로 구성되어 있다. LID-DS는 8개의 속성 중 5개는 문자 특성 및 기호 특성으로 표시되고 3개는 숫자 특성으로 구성되어 있다. 또한, 모든 숫자 특성을 0에서 255 사이로 조절하여 64x64픽셀의 이미지로 변환했다. 이미지의 각 색상 채널은 0에서 255 사이의 값으로 표시되어야 하며 생성된 이미지를 CNN 모델에 사용한다. 데이터를 이미지로 변환하는 이유는 CNN 모델이 이미지 훈련을 위한 딥러닝 모델이기 때문이다. 본 논문에서는 3개의 색상 채널(빨강, 녹색, 파랑)이 있는 RGB 유형의 이미지 데이터 세트를 생성했다. RGB 이미지는 세 가지

유형의 컬러 이미지가 중첩된 구조고 최종적으로, M x N x 3픽셀 배열로 변환한다. M과 N은 각각의 열과 행의 수를 나타낸다.

3.2 CNN 모델

CNN 모델은 Sequential로 생성하였고 3x3 크기의 Conv2D 레이어를 32개의 필터 수를 처음에 생성했다. 마지막 층을 제외한 활성화 함수는 'relu'를 사용했으며 input_shape는 64x64 크기와 3개의 컬러 채널을 가지고 있으므로 (64, 64, 3)의 튜플 값을 가진다. 그리고 사소한 변화를 무시하기 위해서 Maxpooling2D를 통해 주요 값만 추출하여 작은 출력값을 만들어 사용했다. 또한, Conv2D와 Maxpooling은 2차원을 주로 다루기 때문에 전 결합층에 전달하기 위해서는 1차원으로 전달하여야 한다. Flatten 함수를 사용하여 1차원으로 변환하였고 loss 함수로 categorical_crossentropy와 활성화 함수로 'softmax'를 사용했다. 훈련 데이터와 테스트 데이터의 비율은 8:2, 반복실험은 100회 반복하여 실험을 진행했다.

4. 실험 결과

64x64의 크기로 이미지를 변환한 LID-DS 데이터와 KDD99 데이터를 이미지 처리를 위한 CNN 알고리즘을 사용하여 실험했다. 학습된 모델의 성능 평가는 Precision, Recall, F1 Score를 사용했다. 그 이유는 데이터에 대해서 Accuracy만을 가지고 평가하는 것은 부적합하기 때문이다.

학습된 모들의 성능 평가 결과는 표 1과 같으며 Accuracy는 표 2와 같다.

<표 1> 각각의 데이터 세트에 대한 성능 결과

데이터 세트	Precision	Recall	F1-Score
LID-DS	87%	87%	87%
KDD99	78%	79%	74%

<표 2> 각각의 데이터 세트에 대한 정확도 결과

데이터 세트	Accuracy
LID-DS	87%
KDD99	79%

실험 결과, KDD99 데이터에 대한 Accuracy는 약 79%, LID-DS 데이터의 Accuracy는 약 87%로 LID-DS 데이터의 정확성이 약 8% 높은 것을 볼 수 있다. 또한, 다양한 현대 컴퓨터의 공격 특징들을 반영하고 있는 LID-DS 데이터는 현재의 컴퓨터를 검증하고 연구하는 데 중요하다고 할 수 있다.

5. 결론 및 추후 연구

본 논문에서 소개하는 LID-DS 데이터 세트는 시스템의 최신 보안 취약점을 최신 상태로 유지했으며 기본 스레드 정보가 사라지지 않고 새로운 유형의 HIDS를 평가하는 데 사용할 수 있는 방식으로 데이터를 기록한다. KDK99는 앞서 언급했듯이 너무 오래된 컴퓨터 시스템의 특징과 공격 패턴으로 이루어져 있다. 그 결과 LID-DS 데이터의 Accuracy가 KDD99 데이터의 Accuracy 보다 약 8% 높은 것을 확인했다.

또한, ElasticSearch에 데이터를 업로드하여 Kibana를 통해 시각화했다. 그림 1과 같이 시각화된 자료를 확인하면서 대용량 데이터를 한눈에 보기 어려운 단점을 해결하는 효과도 있다.

(그림 1) Kibana를 통한 데이터 시각화

추후 연구로 변환한 이미지 샘플을 늘리기 위해 DCGAN(Deep Convolutional Generative Adversarial Network)를 이용하여 이미지 샘플을 만들어 모델의 성능과 보안에서 중요시하게 생각하는 오탐율을 개선하는 방법과 새로운 공격이나 내부 공격자에 대한 탐지 정확도를 올리기 위한 연구를 계획하고 있다.

Acknowledgement

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다. (UD200014ED)

참고문헌

[1] 최윤정, and 박승수. "이상탐지 (Anomaly

Detection) 및 오용탐지 (Misuse Detection) 분석의 정확도 향상을 위한 개선된 데이터마이닝 방법 연구." 한국정보과학회 학술발표논문집 (2006): 238-240.
[2] 최승오, and 김우년. "제어시스템 침입 탐지 시스템 기술 연구 동향." 정보보호학회지 24.5 (2014): 7-14.
[3] Mouttaqi, Tarik, Tajjeeddine Rachidi, and Nasser Assem. "Re-evaluation of combined Markov-Bayes models for host intrusion detection on the ADFA dataset." 2017 Intelligent Systems Conference (IntelliSys). IEEE, 2017.
[4] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 2186-2193.
[5] Röhling, Martin Max, et al. "Standardized container virtualization approach for collecting host intrusion detection data." 2019 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2019.
[6] Tidjon, Lionel N., Marc Frappier, and Amel Mammar. "Intrusion detection systems: A cross-domain overview." IEEE Communications Surveys & Tutorials 21.4 (2019): 3639-3681.
[7] Kwon, Hyun, et al. "Optimal cluster expansion-based intrusion tolerant system to prevent denial of service attacks." Applied Sciences 7.11 (2017): 1186.
[8] Mouttaqi, Tarik, Tajjeeddine Rachidi, and Nasser Assem. "Re-evaluation of combined Markov-Bayes models for host intrusion detection on the ADFA dataset." 2017 Intelligent Systems Conference (IntelliSys). IEEE, 2017.
[9] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 2186-2193.
[10] 박정민, and 현주연. "ElasticSearch 와 Kibana 를 이용한 웹 아티팩트 시각화." 대한전자공학회 학술대회 (2019): 1350-1353.
[11] Laskov, Pavel, et al. "Learning intrusion detection: supervised or unsupervised?." International Conference on Image Analysis and Processing. Springer, Berlin, Heidelberg, 2005.
[12] Kim, Jihyun, and Howon Kim. "An effective intrusion detection classifier using long short-term memory with gradient descent optimization." 2017 International Conference on Platform Technology and Service (PlatCon). IEEE, 2017.
[13] Kim, Gyuwan, et al. "LSTM-based system-call language modeling and robust ensemble method for designing host-based intrusion detection systems." arXiv preprint arXiv:1611.01726 (2016).
[14] RD Ravipati, M Abualkibash. "Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets-A Review Paper." International Journal of Computer Science & Information Technology (IJCSIT) Vol 11 (2019).
[15] Aydin, M.A.; Zaim, A.H.; Ceylan, K.G. A hybrid intrusion detection system design for computer network security. Comput. Electr. Eng. 2009, 35, 517 - 526.
[16] Al-Jarrah, O.; Arafat, A. Network Intrusion Detection System using attack behavior classification. In Proceedings of the 2014 5th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 1 - 3 April 2014; pp. 1 - 6.
[17] Khan, R.U.; Zhang, X.; Alazab, M.; Kumar, R. An Improved Convolutional Neural Network Model for Intrusion Detection in Networks. In Proceedings of the 2019 Cybersecurity and Cyberforensics Conference (CCC), Melbourne, Australia, 8 - 9 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 74 - 77.
[18] Upadhyay, R.; Pantiukhin, D. Application of convolutional neural network to intrusion type recognition. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics, Udupi, India, 13 - 16 September 2017.