논문 2019-2-6 http://dx.doi.org/10.29056/jsav.2019.12.06

랜섬웨어 방지를 위한 딥러닝 기반의 사용자 비정상 행위 탐지 성능 평가

이예슬*, 최현재*, 신동명*, 이정재**†

Deep Learning based User Anomaly Detection Performance Evaluation to prevent Ransomware

Ye-Seul Lee*, Hyun-Jae Choi*, Dong-Myung Shin*, Jung-Jae Lee**†

요 약

IT 기술의 발달에 따라, 컴퓨터 관련 범죄가 빠르게 급증하고 있으며 특히 최근에는 국내외에서 랜섬웨어 감염에 대한 피해가 급격하게 늘어나고 있다. 기존의 보안 솔루션으로는 랜섬웨어 감염을 방지하기에는 역부족이며 나날이 발전하는 악성코드 및 랜섬웨어와 같은 위협을 방지하기 위해서는 딥러닝 기술을 결합하여 비정상 행위 및 이상 징후를 탐지하는 기법이 필요하다. 본 논문에서는 CNN-LSTM 모델 및 다양한 딥러닝 모델을 사용하여 사용자 비정상 행위를 탐지하는 기법을 제안했으며, 그중 제안하는 모델인 CNN-LSTM 모델의 경우 액 99%의 정확도로 사용자 비정상 행위를 탐지해내는 것을 확인할 수 있었다. 본 연구를 활용하여 사용자 비정상 행위의 랜섬웨어 특징점을 파악하여 랜섬웨어를 방지하는 시스템을 마련하는 데 도움을 줄 수 있을 것으로 기대한다.

Abstract

With the development of IT technology, computer-related crimes are rapidly increasing, and in recent years, the damage to ransomware infections is increasing rapidly at home and abroad. Conventional security solutions are not sufficient to prevent ransomware infections, and to prevent threats such as malware and ransomware that are evolving, a combination of deep learning technologies is needed to detect abnormal behavior and abnormal symptoms. In this paper, a method is proposed to detect user abnormal behavior using CNN-LSTM model and various deep learning models. Among the proposed models, CNN-LSTM model detects user abnormal behavior with 99% accuracy.

한글키워드 : 비정상 행위 탐지, 이상 징후 탐지, 랜섬웨어, 딥러닝, 성능 비교, CNN-LSTM

keywords : Abnomal behavior detection, Anomaly Detection, Ransomware, Deep Learning, Performance Comparison, CNN-LSTM

* 엘에스웨어(주)
** 숭실사이버대학교 교수
† 교신저자: 이정재(bobtree12@naver.com)
접수일자: 2019.11.29. 심사완료: 2019.12.12.
게재확정: 2019.12.20.

1. 서 론

IT 기술이 발달함에 따라, 컴퓨터 시스템과 연관된 침입탐지 및 범죄 등이 빠르게 급증하고 있

다. 특히, 최근 국내외에서 단순 개인 컴퓨터뿐만 아니라 특정 서버를 노리는 랜섬웨어가 기승을 부렸고, 이에 따른 피해가 급격하게 늘어나고 있다. 국내에서는 웹 호스팅 업체인 '나야 나'가 랜섬웨어 '에레보스'에 공격을 당하면서 153대의 서버가 랜섬웨어에 감염되고 웹 호스팅과 서버 호스팅을 맡긴 수천 개가 넘는 홈페이지들이 마비되고 각종 데이터가 암호화되는 등의 대규모의 피해가 발생했다. 해커에게 13억 원이라는 거액을 지불하고 피해 복구를 요구했지만, 해커들이 만든 복호화 프로그램에 오류가 많아 일부 서버와 특정 파일들은 복구가 불가능한 것으로 알려졌다[1]. 이러한 문제는 알려지지 않은 비정상 행위에 대해 대응하는 솔루션으로 랜섬웨어 감염을 방지할 수 있다. 과거부터 현재까지 이상 징후를 탐지하기 위한 다양한 연구들이 수행되었다. 과거에는 이미 알려진 위협의 패턴을 분석하여 관리자가 이를 직접 탐지해내는 지식 기반 침입탐지 시스템이 주를 이루었다. 하지만 기존의 방식으로는 나날이 발전하는 악성코드 및 랜섬웨어와 같은 위협을 해결하기에는 한계가 존재한다. 그러므로 최근에는 행위 기반 침입탐지 시스템에 인공지능 기술을 결합해 비정상 행위 및 이상 징후를 탐지하는 기법을 적용하고 있다. 국외에서는 활발하게 이러한 비정상 행위 및 이상 징후 탐지를 위한 시스템에 인공지능을 적용하는 시스템을 출시하고 있지만, 국내에서는 아직 국외만큼 활발한 시스템 출시가 이뤄지지 않고 연구개발 단계에 머물러있다. 따라서 본 논문에서는 인공지능 기술 중 딥러닝 모델을 적용하는 모델에 대하여 제안한다.

본 논문은 다음과 같이 구성된다. 2장에서는 현재 공개된 침입탐지를 위한 보안 데이터 세트의 특징에 대해 비교하고 3장에서는 딥러닝 기반의 사용자 비정상 행위 탐지 기법에 대한 실험을 설계하고 그 방법을 제시한다. 4장에서는 3장에서 설계한 실험에 대한 실험 결과를 제시한다. 마지막으로 5장에서는 결론 및 향후 연구를 언급한다.

2. 관련 연구

2.1 기존 공개 데이터 세트 분석

2.1.1 DARPA 2009

DARPA 2009[2] 데이터는 인터넷에 연결된 16개의 로컬 서브넷 네트워크 시뮬레이션으로 생성되었다. 데이터 세트는 HTTP, SMTP, DNS를 사용하는 엄청난 양의 네트워크 데이터로 구성된다. DARPA 2009 데이터 세트에는 10일 동안 발생하는 46개의 보안 이벤트가 포함되어 있다[3].

2.1.2 Kyoto 2006+

Kyoto 2006+[4] 데이터 세트는 2006년 11월부터 2009년 8월까지 2년간 실제 트래픽 데이터를 기반으로 작성되었다. KDDcup99에서 파생된 14개의 통계적 특징과 IDS 네트워크 분석 및 평가를 위한 10개의 추가 특징으로 구성되어 있다. Kyoto 2006+ 데이터 세트는 허니팟, 다크 넷 센서, 이메일 서버 및 웹 크롤러를 사용하여 캡처된다. 50,033,015개의 정상 트래픽과 43,043,255개의 공격 트래픽 및 알 수 없는 공격과 관련된 425,719개의 트래픽이 있다.

2.1.3 KDDcup99

KDDcup99[5] 데이터 세트는 네트워크 이상 탐지를 위한 실험에 가장 많이 사용되는 데이터 세트이다. KDDcup99 데이터 세트는 가상 환경에서 데이터를 수집한다. 여러 번의 공격 시뮬레이션을 통해 수집되었으며 9주 동안 TCP 덤프 데이터를 수집했다. 데이터 세트는 Massachusetts Institute of Technology (MIT)

Lincoln Laboratory에서 수집 및 배포되었다. KDDcup99 데이터 세트는 4,898,431개의 정상 또는 공격 데이터를 가지며 41개의 기능으로 구성된다.

2.1.4 NSL-KDD

NSL-KDD[6] 데이터 세트는 KDDcup99 데이터 세트에서 중복 레코드를 포함하거나 포함하지 않는 선택된 기능으로 구성된다. 트레이닝 세트의 일반 트래픽에는 67,24개의 인스턴스가 포함되어 총 126,620개의 인스턴스가 있다. 공격 타입은 Probe, DoS, U2R, R2L 카테고리로 나눠진다.

2.2 UNSW-NB15 데이터 세트

UNSW-NB15 데이터 세트는 호주 사이버 보안센터(Australian Center for Cyber, ACCS)의 Cyber Range 실험실에서 IXIA PerfectStorm 도구를 사용하여 업데이트하였다. Argus, BRO-IDS 도구를 사용하여 총 49개의 피처를 추출한다. 트레이닝 데이터 세트는 총 2,520,044개의 레코드를 가지며 4개의 CSV 파일로 저장되어 있다[7]. 총 9개의 공격 타입으로 분류되며 공격의 종류 및 레코드 수는 아래와 같다.

표 1. UNSW-NB15 공격 클래스 [8]
Table 1. UNSW-NB15 attack class

공격 타입	레코드 수
Normal	2,218,761
Fuzzers	24,246
Analysis	2,677
Backdoor	2,629
DoS	16,353
Exploit	44,525
Generic	215,481
Reconnaissance	13,987
Shellcode	1,511
Worm	174

3. 분석 및 설계

본 실험에서는 사용자 비정상 행위 탐지를 위한 모델을 제안하고, 모델의 성능을 비교하기 위해 다양한 딥러닝 모델을 적용하여 그 성능을 비교 분석하는 것에 대해 제안한다. 딥러닝 기반 사용자 비정상 행위 탐지 모델은 데이터 수집 단계, 보안툴을 사용한 패킷 데이터 처리 단계, 데이터 전처리 및 정규화 단계, 데이터 학습 및 딥러닝 모델 생성 단계, 딥러닝 모델 검증 단계로 구성된다. 그림 1은 딥러닝 기반의 사용자 비정상 행위 탐지를 위한 모델 설계도이다.

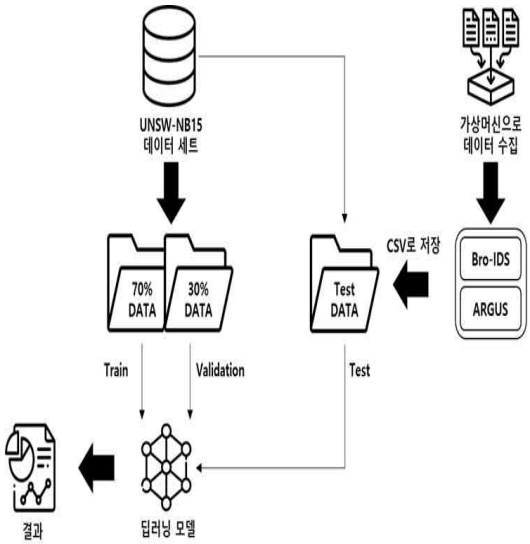


그림 1. 딥러닝 기반 사용자 비정상 행위 탐지 모델
Fig. 1. Deep Learning-based User Anomaly Detection Model

3.1 데이터 수집

UNSW-NB15에서 공개한 비정상 행위 CSV 파일을 사용하여 트레이닝 데이터 세트로 사용한다. 실제 데이터 수집을 위하여 20대의 가상머신을 사용하여 랜섬웨어에 감염된 PC에서 공통적

으로 발생하는 비정상 행위에 대한 데이터와 정상 행위에 대한 데이터를 수집한다. 이러한 데이터는 시계열적인 특성을 지닌다. 또한, 수집한 비정상 행위 데이터는 C&C 서버와 통신 트래픽, 악성 스크립트 파일 다운로드를 위한 DNS 질의, 내부 유포를 위한 특정 포트 스캐닝 등이 사전에 알려진 서버 공격 패턴을 가진다.

3.2 패킷 데이터 처리

20대 가상머신을 통해 수집한 Pcab 파일을 사용하여 트래픽 데이터를 분석하고 데이터를 추출하기 위해 오픈소스인 BRO-IDS, Argus 보안툴을 사용한다. 이를 사용하여 데이터를 추출하고 각 툴에서 분석된 피처를 매핑 시키고 학습 데이터 생성을 위해 CSV 파일로 변환한다.

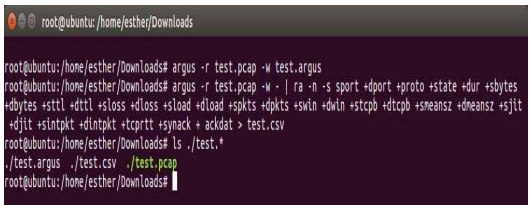
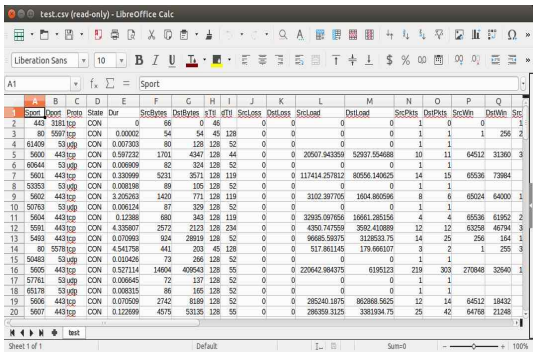


그림 2. ARGUS 툴을 통한 트래픽 데이터 추출
Fig. 2. Traffic data extraction through ARGUS tool



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Sport	Dport	Proto	State	Dur	SrcBytes	DstBytes	sTtl	dTtl	SrcLoss	DstLoss	SrcLoad	DstLoad	SrcPkts	DstPkts	SrcWin	DstWin
2	443	3181	tcp	CON	0	66	0	46		0	0	0	0	1	0	0	1
3	80	5597	tcp	CON	0.00002	54	54	45	128	0	0	0	0	1	1	1	256
4	61409	53	udp	CON	0.007303	80	128	128	52	0	0	0	0	1	1		
5	5600	443	tcp	CON	0.597232	1701	4347	128	44	0	0	20507.943359	52937.554688	10	11	64512	31360
6	60644	53	udp	CON	0.006909	82	324	128	52	0	0	0	0	1	1		
7	5601	443	tcp	CON	0.330999	5231	3571	128	119	0	0	117414.257812	80556.140625	14	15	65536	73984
8	53353	53	udp	CON	0.008198	89	105	128	52	0	0	0	0	1	1		
9	5602	443	tcp	CON	3.205263	1420	771	128	119	0	0	3102.397705	1604.860596	8	6	65024	64000
10	50763	53	udp	CON	0.006124	87	329	128	52	0	0	0	0	1	1		
11	5604	443	tcp	CON	0.12388	680	343	128	119	0	0	32935.097656	16661.285156	4	4	65536	61952
12	5591	443	tcp	CON	4.335807	2572	2123	128	234	0	0	4350.747559	3592.410889	12	12	63258	46794
13	5493	443	tcp	CON	0.070993	924	28919	128	52	0	0	96685.59375	3128533.75	14	25	256	164
14	80	5578	tcp	CON	4.541758	441	203	45	128	0	0	517.861145	179.666107	3	2	1	255
15	50483	53	udp	CON	0.010426	73	266	128	52	0	0	0	0	1	1		
16	5605	443	tcp	CON	0.527114	14604	409543	128	55	0	0	220642.984375	6195123	219	303	270848	32640
17	57761	53	udp	CON	0.006645	72	137	128	52	0	0	0	0	1	1		
18	65178	53	udp	CON	0.008315	86	165	128	52	0	0	0	0	1	1		
19	5606	443	tcp	CON	0.070509	2742	8189	128	52	0	0	285240.1875	862868.5625	12	14	64512	18432
20	5607	443	tcp	CON	0.122699	4575	53135	128	55	0	0	286359.3125	3381934.75	25	42	64768	21248

그림 3. CSV로 변환된 pcap 데이터
Fig. 3. Pcap data converted to CSV

3.3 데이터 전처리 및 정규화

이 단계에서는 문자열 데이터를 수치형 데이터로 치환하고 N/A 값에 대한 처리 과정을 가진다. 또한, 딥러닝 학습에 사용될 피처를 추출하기 위하여 트래픽 피처와 정상/비정상 행위 레이블 간의 상관관계를 알기 위하여 피어슨(Pearson) 상관 계수[9]를 구하였고, 그를 통하여 불필요한 피처를 제거하였다. 그 후 값 범위의 차이를 왜곡시키지 않고 데이터 세트를 공통 스케일로 변경하기 위하여 트래픽 데이터를 0과 1사이로 스케일을 조정하는 정규화 과정을 거친다. 사용된 정규화 공식은 다음과 같다.

$$x_{\neq w} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

3.4 딥러닝 학습 및 모델 생성

딥러닝 기반의 사용자 비정상 행위 탐지를 위하여 UBSW-NB15 csv로 저장된 데이터를 피어슨 상관 계수 분석을 통해 선택된 25개의 피처를 사용하여 정상, 비정상의 이진 분류를 하는 딥러닝 학습을 진행한다. 사용자 비정상 행위 탐지를 위하여 LSTM, GRU, CNN-LSTM 모델을 적용하여 딥러닝 모델을 생성한다. 그 후 20대의 가상머신을 통해 수집되고 전처리 과정을 거친 csv로 저장된 데이터를 생성된 모델에 테스팅 단계 입력 값으로 사용한다.

4. 실험 결과

4.1. 비정상 행위 탐지 결과

사용자 비정상 행위를 탐지하기 위해 다양한 딥러닝 모델을 적용했다. 정상, 비정상 행위의 레이블이 있는 데이터를 사용하여 이중 분류를 가

능하게 하는 모델들을 본 실험에 사용했다. 본 연구에서 제안하는 CNN-LSTM 모델[10]은 CNN과 LSTM을 결합하여 시계열 데이터 양상을 가지는 실험 데이터에 대해 최적의 학습효과를 내기 위해 설계된 사용자 비정상 행위 탐지를 위한 분류 알고리즘이다. CNN-LSTM 모델은 특징을 추출하고, 필터의 값을 비선형 값으로 바꿔주는 1개의 컨볼루션 층과 학습하는 매개변수를 줄여서 출력을 간소화하는 1개의 풀링층을 사용하고, 컨볼루션 층을 통과한 정보를 시계열 데이터 모델링에 적합한 LSTM 계층의 입력으로 사용되어 신경망 분류를 수행하고 시그모이드 함수를 사용하여 이진 분류 값을 출력하는 그림 4와 같은 CNN-LSTM 모델을 제안한다. 그 외에도 제안한 모델과 성능 비교를 위하여 LSTM, GRU 모델을 적용한다.

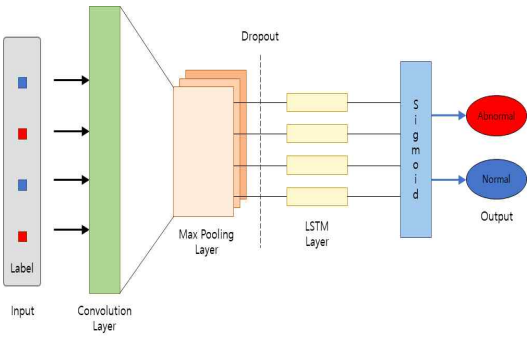


그림 4. CNN-LSTM 모델
Fig 4. CNN-LSTM model

실험을 위하여 UNSW_NB15 데이터의 70%는 학습 데이터로, 나머지 30%는 검증 데이터로 사용했다. 20대의 가상머신을 통해 수집된 데이터는 테스트 데이터로 사용되었다. 제안하는 CNN-LSTM 모델은 99.15%의 비정상 행위 탐지율을 보였다. 그림 5는 실제 데이터의 레이블과 검증 데이터를 통해 예측된 레이블의 결과를 비교하는 사용자 비정상 행위 탐지 Confusion matrix이다. 비정상 행위를 정상 행위로 탐지해내는 비율은 전체의 0.13%, 정상 행위를 비정상 행위로 탐지해내는 비율은 0.72%로 나타났다.

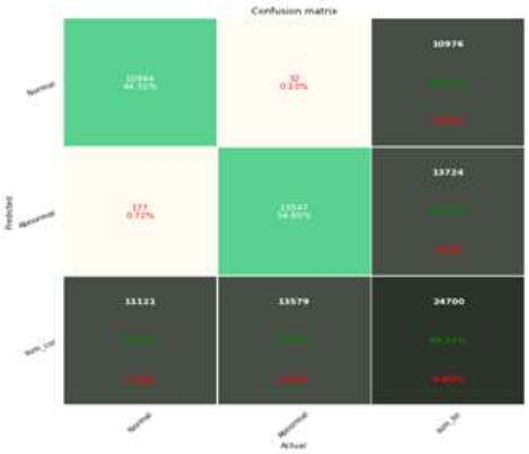


그림 5. 사용자 비정상 행위 탐지 Confusion matrix
Fig. 5. User Anomaly Detection Confusion matrix

4.2. 딥러닝 모델별 성능 비교

사용자 비정상 행위 탐지를 위한 각 딥러닝 알고리즘별 성능 비교는 표 2와 같다. 모델별 정확도 순서로는 CNN-LSTM, LSTM, GRU 순으로 나타났다. 제안하는 모델인 CNN-LSTM의 경우 가장 높은 정확도인 99%에 달하는 정확도를 보이고, 가장 정확도가 낮은 GRU 모델의 경우 89%의 정확도를 보였다. 가장 성능이 좋은 CNN-LSTM 모델과 GRU 모델의 정확도는 약 10%의 탐지 성능 차이를 나타냈다. 또한, 비정상 행위 탐지 성능을 평가하기 위해 평가 척도 중 f-score를 사용한다. f-score는 precision(정확도)과 recall(재현율)을 사용하여 계산한 값으로 데이터 분포를 고려한 지표이다. F-score는 의미상으로는 precision과 recall에 대한 평균인데, 그냥 평균을 계산하면 값의 왜곡 현상이 생기기 때문에, 가중치를 주어 계산한 값이다. F1-score를 비

교해봐도, CNN-LSTM 모델이 다른 모델에 비해 성능 뛰어난 것을 알 수 있다.

표 2. 딥러닝 알고리즘 사용자 비정상 행위 탐지 성능 비교
Table 2. Deep Learning Algorithm Performance Comparison of User Anomaly Detection

Algorithm	Precision	Recall	F1 Score
LSTM	0.93	0.99	0.96
GRU	0.89	0.92	0.90
CNN-LSTM	0.99	0.99	0.99

5. 결 론

본 논문에서는 딥러닝 알고리즘을 사용하여 사용자 비정상 행위를 탐지해내는 모델에 대해 제안하고, 그 외의 다양한 딥러닝 알고리즘을 사용하여 비정상 행위를 탐지해내는 것에 대해 비교 분석했다. 딥러닝 기반의 비정상 행위 탐지를 통하여, 알려지지 않은 비정상 행위에 대해 대응하는 솔루션으로 확장해 랜섬웨어 감염을 방지할 수 있을 것이다.

다양한 딥러닝 알고리즘으로 사용자 비정상 행위를 탐지해내는 모델을 생성한 결과 CNN-LSTM 모델이 최고 99%, 그다음으로 LSTM, GRU 모델의 순서로 나타났다.

하지만 이는 랜섬웨어에 감염된 PC에서 공통적으로 발생하는 비정상적인 행위에 대해 수집된 데이터에 대해 탐지해내는 지표로, 아직은 본 연구를 실제 랜섬웨어 방지를 위한 솔루션에 적용하기에는 무리가 있어 보인다. 추후 랜섬웨어 특징점을 가지는 비정상 행위에 대한 데이터를 더 수집하여 학습시키고, 비정상 행위로 탐지되는 IP를 자동으로 차단해 랜섬웨어를 감염 전 미리 방지할 수 있게 설계할 것이다. 또한, 수집된 데이터를 지속해서 학습할 수 있도록 하는 연구가 필요하며 다음 연구를 통해 진행될 예정이다.

Acknowledgement
본 연구는 중소벤처기업부의 2019년도 혁신기업기술개발 연구결과로 수행되었음

참 고 문 헌

[1] Byline Network, “인터넷나야나, ‘APT 공격’으로 랜섬웨어 감염…보안관리 허술” URL:https://byline.network/2017/06/1-792/
[2] Nour Moustafa and Jill Slay, “Creating novel features to anomaly network detection using darpa-2009 data set”, Proceedings of the 14th European Conference on Cyber Warfare and Security. Academic Conferences Limited, p.204, 2015. URL: https://www.researchgate.net/profile/Nour_Moustafa4/publication/279850205_Creating_Novel_Features_to_Anomaly_Network_Detection_Using_DARPA-2009_Data_set/links/559be12d08ae7f3eb4cedcca.pdf
[3] The ground truth of darpa-2009 dataset. May 2016. URL: http://www.darpa2009.netsec.colostate.edu/
[4] SONG, Jungsuk, et al., “Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation”, Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security. ACM, 2011. pp.29-36. DOI: https://doi.org/10.1145/1978672.1978676
[5] PROTIĆ, Danijela D. Review of KDD Cup'99, NSL-KDD and Kyoto 2006+ datasets. Vojnotehnički glasnik, 2018, 66.3: 580-596.9 DOI: https://doi.org/10.5937/vojtehg66-16670

[6] DHANABAL, L.; SHANTHARAJAH, S. P. “A study on NSL-KDD dataset for intrusion detection system based on classification algorithms”, International Journal of Advanced Research in Computer and Communication Engineering, 2015, 4.6: 446-452. URL: https://pdfs.semanticscholar.org/1b34/80021c4ab0f632efa99e01a9b073903c5554.pdf

[7] MOUSTAFA, Nour; SLAY, Jill. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective, 2016, 25.1-3: 18-31. DOI: https://doi.org/10.1080/19393555.2015.1125974

[8] MOUSTAFA, Nour, Nour. “Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic”, 2017. PhD Thesis. University of New South Wales, Canberra, Australia. URL: https://www.researchgate.net/profile/Nour_Moustafa4/publication/328784548_Designing_an_online_and_reliable_statistical_anomaly_detection_framework_for_dealing_with_large_high-speed_network_traffic/links/5be2e4164585150b2ba57c6a/Designing-an-online-and-reliable-statistical-anomaly-detection-framework-for-dealing-with-large-high-speed-network-traffic.pdf

[9] BENESTY, Jacob, et al., “Pearson correlation coefficient”, Noise reduction in speech processing. Springer, Berlin, Heidelberg, pp.1-4, 2009. DOI: https://doi.org/10.1109/tasl.2008.919072

[10] WANG, Jin, et al., “Dimensional sentiment analysis using a regional CNN-LSTM model”, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp.225-230, 2016. DOI: https://doi.org/10.18653/v1/p16-2037

저 자 소 개

이예슬 (Ye-Seul Lee)

2018년 숭실대학교
융합소프트웨어학과 석사
2018년-현재 엘에스웨어(주) 주임

<주관심분야> 딥 러닝, 빅 데이터, 블록체인

최현재 (Hyunl-Jae Choi)

2018년 성균관대학교
전자전기컴퓨터학과 석사
2018년-현재 엘에스웨어(주) 주임

<주관심분야> 네트워크/시스템 보안, 취약점분석, 블록체인

저 자 소 개

신동명 (Dong-Myung Shin)

2003년 대전대학교 컴퓨터공학과 박사
2001년-2006년 한국정보보호진흥원(KISA) 응용기술팀 선임연구원
2006년-2014년 한국저작권위원회 저작권기술팀 팀장
2014년-2016년 한국스마트그리드사업단 보안인증팀 팀장
2016년-현재 엘에스웨어(주) 연구소장/상무이사

<주관심분야> 오픈소스 라이선스, 시스템/네트워크보안, 스마트그리드 인증/보안, SW 취약점분석・감정, 블록체인

이정재 (Jung-Jae Lee)

2011년 숭실대학교 경영학박사
1995년-1996년 대일화학공업주식회사 전산실
2000년-2013년 한국저작권위원회 기술, 유통 팀장
2013년-현재 숭실사이버대학교 교수

<주관심분야> 콘텐츠유통, 플랫폼, 저작권 라이선스, 오픈소스, 블록체인, 딥 러닝