Journal of the Korea Convergence Society
Vol. 10. No. 11, pp. 1-7, 2019

ISSN 2233-4890
https://doi.org/10.15207/JKCS.2019.10.11.001

네트워크 공격 탐지 성능향상을 위한 딥러닝을 이용한 트래픽 데이터 생성 연구

이우호[1], 함재균[2], 정현미[2], 정기문[2*]
[1]전남대학교 정보보안협동과정 박사과정
[2]한국과학기술정보연구원 슈퍼컴퓨팅본부 연구원

Traffic Data Generation Technique for Improving Network Attack Detection Using Deep Learning

Wooho Lee[1], Jaegyoon Hahm[2], Hyun Mi Jung[2], Kimoon Jeong[2*]
[1]Ph.D Course, Interdisciplinary Program of Information Security, Chonnam National University
[2]Researcher, Div. of National Supercomputing, Korea Institute of Science and Technology Information

요 약 네트워크 공격을 탐지하기 위하여 기계학습을 이용한 다양한 연구가 최근 급격히 증가하고 있다. 이러한 기계학습 방법은 많은 데이터에 의존적이며 연구를 위해 다양한 실험 데이터가 공개되어 사용되고 있다. 하지만 실험 데이터 및 실제 환경에서 수집되는 데이터는 class간의 수량이 불균형하다는 문제점을 가지고 있다. 본 연구에서는 기계 학습을 이용한 침입탐지시스템의 한계점 중 학습데이터의 class간 불균형으로 인한 분류 성능 저하를 해결하기 위한 방법을 제안한다. 이를 위해 네트워크 트래픽 데이터를 처리하고 seqGAN를 이용하여 부족한 데이터를 생성하였다. 제안된 방법은 NSL-KDD, UNSW-NB15 데이터 셋을 대상으로 Text-CNN을 이용하여 분류하는 테스트를 실행한 결과 정밀도가 향상되는 것을 확인할 수 있었다.

주제어 : 네트워크 보안, 침입탐지, 네트워크 트래픽 데이터, 딥러닝, GAN

Abstract Recently, various approaches to detect network attacks using machine learning have been studied and are being applied to detect new attacks and to increase precision. However, the machine learning method is dependent on feature extraction and takes a long time and complexity. It also has limitation of performace due to learning data imbalance. In this study, we propose a method to solve the degradation of classification performance due to imbalance of learning data among the limit points of detection system. To do this, we generate data using Generative Adversarial Networks (GANs) and propose a classification method using Convolutional Neural Networks (CNNs). Through this approach, we can confirm that the accuracy is improved when applied to the NSL-KDD and UNSW-NB15 datasets.

Key Words : Network security, Intrusion detection, Network traffic data, Deep learning, GAN

1. 서론

침입탐지시스템(Intrusion Detection Systems. IDS)은 네트워크 공격으로부터 내부 자원을 보호하는데 사용되는 시스템으로 최근 기계학습 및 딥러닝 기술을 적용하는 연구가 발전하고 있다.

*This research was supported by Korea Institute of Science and Technology Information(KISTI)
*Corresponding Author : Kimoon Jeong(kmjeong@kisti.re.kr)
Received September 27, 2019 Revised November 4, 2019
Accepted November 20, 2019 Published November 28, 2019

특히 네트워크 트래픽의 payload를 분석하면 네트워크 공격을 효과적으로 탐지할 수 있기에 이를 이용하여 공격을 탐지하기 위한 연구가 활발하다.

Wang K 등[1]은 비정상적인 행위를 탐지하기 위하여 payload 특징(feature)의 분포를 이용하여 분류하는 방법을 제안했다. Nigel Williams 등[2]은 payload에서 22개의 트래픽 특징을 추출하고 bayesian 과 decision tree 알고리즘을 사용하여 데이터를 분류하는데 트래픽 특징의 통계적 수치를 이용하였다.

이러한 payload를 이용한 분석 연구를 위해서는 네트워크 트래픽 데이터가 필요하기에 Knowledge Discovery and Data Mining 1999 (KDD99)[3]와 같은 공개된 데이터셋이 주로 사용되고 있다.

기계학습과 딥러닝 기술은 데이터 중심으로 학습을 하여 모델을 생성하므로 양질의 데이터셋을 이용 하는것이 무엇보다 중요하지만 네트워크 침입탐지 분야에서 사용되는 데이터셋은 데이터가 불균형한 문제를 가지고 있다. Table 1과 같이 KDD99를 개선한 NSL-KDD[4] 데이터셋의 데이터 현황에서 학습데이터를 살펴보면 원격에서 공격을 의미하는 R2L 데이터는 정상(Normal) 데이터의 1.4%, Root 권한에 대한 공격을 의미하는 U2R 데이터의 경우는 정상데이터의 약 0.08% 밖에 되지 않는다.

Table 1. NLL-KDD Data set distribution

	Normal	DoS	Probe	R2L	U2R
Training	67,343	45,930	11,656	995	49
Test	9,711	7,458	2,421	2,754	200

위와 같이 데이터 class간 불균형한 상태에서 정상과 비정상을 학습할 경우 정확한 학습이 되기 어렵고 이것은 침입탐지스템의 성능저하로 연결될 수 밖에 없다.

데이터 불균형을 해결하기 위해서는 과포함된 데이터를 삭제하는 undersampling 방법과 부족한 데이터를 보완하는 oversampling방법 등이 있다[5,6].

본 논문에서는 딥러닝 기술중의 하나인 seqGAN[7]을 이용하여 침입탐지 연구를 위해 사용되는 데이터셋의 불균형 문제를 해결하고자 한다. 이를 위해서 pcap 형태로 되어있는 네트워크 원본 데이터셋에서 분석하고자 하는 데이터를 추출하여 seqGAN에 입력할 수 있는 형태로 변환하는 것이 과정이 필요하다.

네트워크 침입탐지 연구에 주로 사용되는 NSL-KDD 데이터셋과 UNSW-NB15[8] 데이터셋에 대해서 본 논문에서 제안한 방법을 적용하여 데이터를 생성한 후 딥러닝 분류 알고리즘에 적용하였을 때 기존 데이터셋에 비하여 정밀도가 약 7% 정도 향상되는 것을 확인할 수 있었다.

본 논문의 구성은 2장에서 최신 침입탐지 연구 및 데이터셋에 대해서 알아보고 3장에서 네트워크 트래픽 원본데이터를 가공하여 딥러닝 기술로 생성하는 방법을 설명한다. 4장에서 생성된 데이터를 이용한 분류 실험결과를 보여주고 5장에서 결론을 맺는다.

2. 관련연구

네트워크 침입탐지 분야에서 머신러닝 및 딥러닝 기술을 이용한 연구가 증가하고 있다.

Bo dong 등은 DBN모델을 사용하여 데이터를 부호화하고 feature engineering을 사용하지 않고 딥러닝 모델을 이용하여 공격을 탐지 하는 방법을 제안했다[9].

Lopez-Martin 등은 네트워크 흐름을 분류하고 하이퍼 파라미터 및 기능 집합의 측면에서 그 환경에서 최상의 설정을 발견하기 위해 CNN 아키텍처를 이용한 분류 방법을 제시했다[10].

Rahul 등의 연구자는 또한 네트워크 트래픽을 분류하기 위해 CNN을 사용하도록 제안했지만 한정된 양의 데이터로만 3 가지 애플리케이션을 분류하는 한계점이 있었다[11].

Auld 등은 다층 지각의 형태로 베이지안 신경망을 배치하고 이에 따라 그들의 데이터 세트를 분류하였다. 이 작업에서 가장 낮은 성능 수치는 minor class를 의미한다. 따라서 minor class의 data가 적을수록 성능에 부정적인 영향을 미치는 것을 볼 수 있다[12].

WANG는[13]에서 세션 트래픽 선두 수백 바이트만을 사용하여 CNN 모델을 적용하는 방법을 제안하였다. 이는 초기 단계의 멀웨어 트래픽을 탐지하는 기능을 가진다. CNN알고리즘을 이용하여 이미지 분류 방법을 사용하기 때문에 프로토콜에 독립적이라는 특징이 있다.

다음으로 네트워크 침입 탐지 연구분야에서 많이 사용되고 있는 NSL-KDD[4], UNSW-NB15[8]에대해서 알아본다.

NSL-KDD 데이터셋은 KDD99 데이터 셋에서 학습데이터와 테스트 데이터 간의 중복되는 문제점을 해결하여 개선한 데이터 셋으로 그 데이터셋은 Table 1에서 확인할 수 있다.

UNSW-NB15 데이터 셋은 오스트레일아의 사이버 보안 센터(ACCS)에서 생성한 것으로 원본 데이터를 비롯하여 다양한 형태의 데이터를 제공하고 있다. 그 구성은 Table 2와 같다. 학습 데이터에서 shellcode, worms 등의 데이터 수가 정상(Normal) 데이터에 비해서 현저히 적은 것을 확인할 수 있다.

Table 2. UNSW-NB15 Data set distribution

	Training	Test
Normal	56,000	37,000
Analysis	2,000	677
Backdoor	1,746	583
DoS	12,264	4089
Exploits	33,393	11,132
Fuzzers	18,184	6,062
Generic	40,000	18,871
Reconnaissance	10,491	3,496
Shellcode	1,133	378
Worms	130	44
Total	175,341	82,332

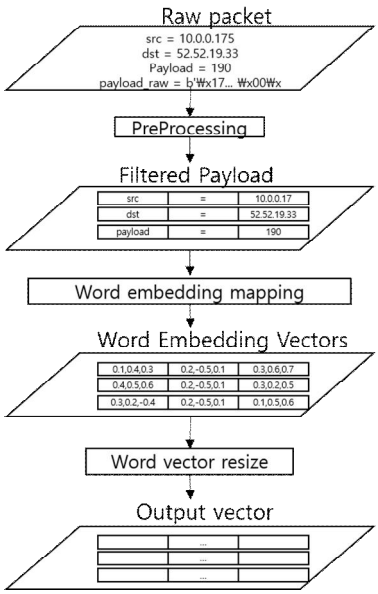

Fig. 1. Packet Data Preprocessing Process

3. 네트워크 침입탐지 데이터 생성방법

본 논문에서는 네트워크 트래픽 데이터를 이용하여 minor class에 대한 유사 데이터를 생성하는 oversampling방법을 제안한다. 먼저 네트워크 트래픽 원천 데이터인 pcap 파일에 대하여 DPI 기반 분석을 위하여 IP, Serivice protocol, payload 등을 추출하여 딥러닝 알고리즘에 잘 적용될 수 있도록 Vector화 과정을 거친다. 이후 도출된 트래픽 Vector 데이터는 seqGAN 알고리즘을 이용하여 원본 데이터와 유사한 데이터를 생성하게 된다.

우선 데이터를 전처리하는 과정은 Fig. 1과 같다.

먼저 Raw packet 데이터는 scapy 프로그램을 이용하여 패킷의 양방향 IP, service protocol, payload 데이터를 추출한 후 vector형태의 인코딩 단계를 거친다.

추출된 payload는 word 특성을 같기에 word embedding 방법을 사용하여 인코딩한다[14]. 이와 같은 방법은 자연언어 처리 분야에서는 전반적인 성능 향상을 위해 사용한다[15].

따라서 tanh를 이용하여 vector로 변환하고 Gaussian Mixture Model(GMM)[16]을 이용하여 데이터를 군집화 한다. GMM은 Gaussian 분포가 여러 개 혼합된 clustering 알고리즘으로 Feature의 데이터 정규화 과정을 거친 값을 바탕으로 공변량과 최빈값(mode)의 평균값을 이용하여 multivatirate gaussian 분포를 구한다.

먼저 특징(feature)중 숫자 유형을 가진 데이터의 경우 tanh를 이용하여 -0.99에서 0.99의 범위 값으로 변환하게 된다. 이는 연산의 결과가 0이 나올 경우 최적화 과정에서 over fitting 문제가 발생할 수 있기 때문이며, 0의 값이 나오지 않도록 clip을 지정한다.

두 번째 service protocol 특징의 경우 범주 유형의 데이터 특성을 가지고 있다. 범주형 데이터의 경우 일반적인 자연어에 비해 크기가 작으므로 soft max를 이용하여 처리한다. 그리고 one hot encoding을 이용하여 vector를 생성할 경우 0이 발생할 수 있으므로 noise 값을 넣어 정규화를 한다.

세 번째 payload와 같은 긴 문자열 데이터의 경우 word-gram을 이용한 후, Gumbel-softmax을 이용하여 분포를 정리한다. encoding 단계에서 먼저 payload를 word로 분리한 다음 각 word embedding을 통해 vector로 변환한다. payload에 많이 포함된 단어(user name, paswword, control) 등은 word embedding 과정에 노이즈가 유입되는 경우를 고려했다. 또한 잡음의 영향을 없애기 위해 word embedding mapping을 수행하기 전에 payload를 필터링 한다. 먼저 학습 데이터 셋에서 다양한 word frequency를 계산한 다음 high frequency word를 기반으로 사전을 생성하여 payload를 필터링한다. 이러한 전처리를 단계를 거친 데이터는 각 트래픽 별로 vector화 된다.

이렇게 도출된 데이터는 seqGAN[7] 알고리즘을 통해

새로 생성될 수 있다. seqGAN은 기존 GAN알고리즘의 discrete한 text data에 적용하기 위해 기존 Generator에 강화학습을 접목한 모델이다. 텍스트 Generation을 실행하기 위해 생성하기 위한 구조와 알고리즘은 각각 Fig. 2와 Fig. 3에서 확인할 수 있다.

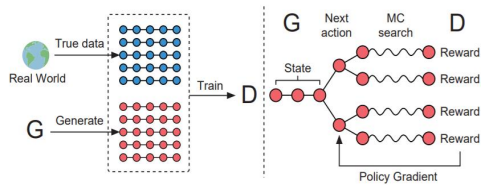

Fig. 2. Architecture of seqGAN

Algorithm 1 Sequence Generative Adversarial Nets

Require: generator policy G_θ; roll-out policy G_β; discriminator D_ϕ; a sequence dataset $S = X_{1:T}$

Initialize G_θ, D_ϕ, with random weights θ, ϕ.
Pre-train G_θ using MLE on S
$\beta \leftarrow \theta$
Generate negative samples using G_θ for training D_ϕ
Pre-train D_ϕ via minimizing the cross entropy
repeat
  for g-steps **do**
    Generate a sequence $Y_{1:T} = (y_1, \cdots, y_T) \sim G_\theta$
    for t in 1 : T **do**
      Compute $Q(a = y_t; s = Y_{1:t-1})$
    end for
    Update generator parameters via policy gradient
  end for
  for d-steps **do**
    Use current G_θ to generate negative examples and combine with given positive examples S
    Train discriminator D_ϕ for k epochs
  end for
  $\beta \leftarrow \theta$
until SeqGAN converges

Fig. 3. Algorithm of seqGAN

seqGAN 알고리즘을 이용하여 생성되는 데이터는 생성기의 loss 값은 0.98이상 1.0이하로 설정하고, 분류기 loss 값은 0 이상 0.1이하 수치의 트래픽만을 구분하여 생성하도록 하였다.

4. 분류 실험

본 장에서는 3장에서 제안한 방법으로 NSL-KDD 데이터셋과 UNSW-NB15 데이터셋의 정상 데이터 대비 양적으로 불균형이 심한 데이터를 일부 생성하여 딥러닝 분류 알고리즘을 적용하였을 때 정확성 및 정밀성 등이 얼마나 향상되는지 알아본다.

제안한 방법으로 NSL-KDD의 학습데이터에서 양이 충분하지 않은 miniority class인 Probe, R2L, U2R class의 데이터를 생성시켰다. UNSW-NB15 학습데이터셋는 Analysis, Backdoor, Shellcode, Worm class의 데이터를 생성시켰다. 그 결과는 Table 3과 Table 4에서 각각 확인할 수 있다.

다음으로 생성된 데이터가 포함된 데이터셋의 성능을 확인하기 위하여 딥러닝 분류 알고리즘을 적용하여 똑같은 환경에서 데이터 생성 이전과 이후의 정확성 등 성능을 확인하는 실험을 진행하였다.

본 논문에서는 sequential한 형태의 데이터에 적합한 딥러닝 분류 알고리즘으로 Text-CNN[17] 알고리즘을 선택하여 적용한다.

이 알고리즘은 Fig 4.에 도식된 바와 같이 3개의 Convolution 층, 2개의 max-Pooling 층 및 1개의 soft max층으로 구성된다. Convolution 층은 multi conv-olution kernels를 포함하며 single perspective 관점에서 특징을 추출한다.

Table 3. NSL-KDD Data set distribution(Actual data and Gan applied data)

	Category	Normal	Dos	Probe	R2L	U2R	Total
NSL-KDD	Actual	67,343	45,930	11,656	995	49	125972
	Gan applied	67,343	45,930	15,152	1,293	500	129,718

Table 4. UNSW-NB15 Data set distribution(Actual data and Gan applied data)

	Category	Normal	Generic	Exploits	Fu-zzers	DoS	Reconnaisance	Analysis	Backdoor	shell code	worm
UNSW-NB15	Actual	56,000	40,000	33,393	18,184	12,264	10,491	2,000	1,746	1,133	130
	Gan applied	56,000	40,000	33,393	18,184	12,264	10,491	2,600	2,269	2,266	1,000

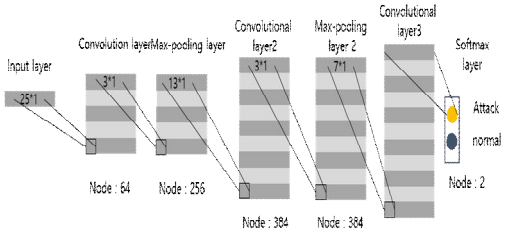

Fig. 4. Architecture of Text-CNN

좀 더 상세히 설명하면 첫 번째 convolution층은 low level 특징을 추출하고 convolution kernel size는 word embedding vector의 차원보다 커야한다.

두 번째 convolution층은 low level 특징 보다 높은 수준의 특징을 추출하고 convolution 크기를 점차 감소시킨다.

max pooling층은 특징의 중 가장 높은 점수를 가진 특징은 보존하고 낮은 점수를 얻은 특징은 삭제한다. 이 층은 알고리즘의 noise의 영향을 줄이고 일반화하는 능력을 향상시킨다.

학습 단계에서 soft max 계층은 예측된 값과 실제 값 사이의 loss를 계산하고 네트워크의 가중치를 loss값을 이용하여 조정한다. test 단계에서 soft max 계층의 예측된 class의 확률을 출력한다.

실험 환경은 windows10 64비트 OS가 설치된 서버이며, 서버의 사양은 8코어 및 16GB 메모리, I7-6770 CPU@3.40GHz이다. GeForce mini GTX 1060 6GB GPU가 가속기로 사용되었다. 원본 데이터의 경우 데이터 셋의 16%를 테스트 데이터로 활용하였고 나머지는 학습 데이터로 실험하였다. 생성된 데이터를 추가하여 유사 트래픽을 생성한 데이터셋의 경우 학습 데이터와 테스트 데이터의 비율을 각각 70:30으로 분류하였다.

준비된 두 개의 데이터셋의 분류 성능을 평가하기 위하여 실제 데이터와 제안된 방법으로 생성된 데이터 각각에 대하여 총 10번을 반복하여 실험하였다. 반복 시행한 실험 결과에 대해서 각각 Precision, Recall, F1-score, Accuracy 값을 산출하여 산술 평균한 값은 Table 5에서 확인할 수 있다.

제안된 방법으로 생성된 데이터가 포함된 데이터셋을 이용할 경우 NSL-KDD 데이터 셋에서 Precision은 약 8%, Recall은 약 5%, F1-score는 약 7%, Accuracy 약 8%의 성능이 향상됨을 확인할 수 있었다. UNSW-NB15 데이터셋은 약 2% ~ 3% 정도의 성능 향상을 보였다.

다음으로 각 데이터셋의 영역별로 자세히 비교를 실시했다. 침입탐지에서 중요한 부분은 오탐율을 낮추는 것이기에 실험 결과값중 Precision 값만을 적용하였다.

Fig. 5는 KDD-NSL 데이터셋의 실험 결과인데 Dos와 Benign에 대한 Precision 값은 비슷했지만 생성된 데이터가 적용된 minor class(U2R, R2L, Probe)의 경우 Precision 값이 개선되는 것을 볼 수 있다.

Fig. 6은 UNSW-NB15에 데이터 셋에 대한 분류 실험이다. NSL-KDD 데이터 셋과 유사하게 생성된 데이터가 적용된 minor class(Analysis, Backdoor, sheelcode, worm)의 Precision 값이 향상되는 것을 볼 수 있다.

또한 제안한 방법으로 생성된 데이터를 적용할 경우, 생성된 데이터가 적용되지 않은 major class의 성능도 미미하지만 향상된 것을확인할 수 있다.

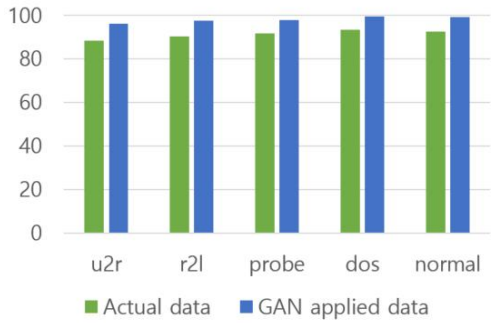

Fig. 5. NSL-KDD Precision Comparison

Table 5. Experimental result of Two datasets

	Dataset	Precision	Recall	F1-Score	Accuracy
Actural	NSL-KDD	93.926	92.39	93.151	92.22
	UNSW-NB15	92.408	94.30	93.344	93.04
Gan applied	NSL-KDD	98.12	98.81	98.46	98.38
	UNSW-NB15	96.128	96.41	96.268	96.07

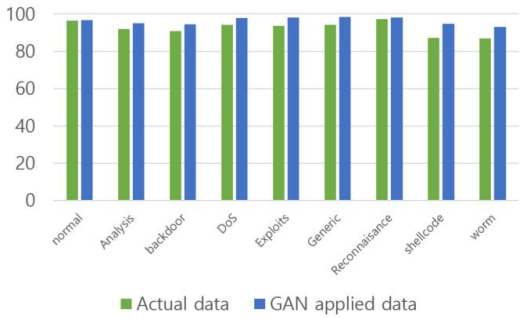

Fig. 6. UNSW-NB15 Precision Comparison

5. 결론

네트워크 침입탐지의 성능을 향상시키기 위해 최근 머신러닝 및 딥러닝 등 데이터를 중심으로 하는 연구가 증가하고 있다. 이러한 연구에 있어서 양질의 데이터의 중요성은 증가하고 있지만 연구를 위해 사용되는 많은 데이터는 항목별로 데이터가 불균한 문제를 안고 있으며 이는 연구의 질을 향상시키는데 장애가 되고 있다.

불균형한 데이터셋의 한계를 극복하기 위하여 본 연구에서는 딥러닝 기법중의 하나인 SeqGAN을 이용하여 유사 데이터를 생성하는 방법을 제안하였다.

제안하는 방법은 네트워크 트래픽 pcap 파일에 대하여 DPI 기반 분석을 위해 IP, Service protocol, payload를 기준으로 하여 특성을 추출하여 분류를 하였다. 또한 데이터 생성에 대한 효율성을 유지하기 위해 충분한 정보를 유지하는 데이터 전처리 방법을 설계했다. 제안하는 방법을 통해 기존의 데이터 셋의 유사트래픽을 추가하여 데이터 불균형 문제를 감소함으로써 정밀성을 향상시키기 위한 실험을 진행하였다.

실험결과 공격 트래픽의 희소성이 높을수록 정밀성이 평균 7% 정도 개선되는 것을 확인 할 수 있었다. 본 접근 방법을 통해 기존의 딥러닝을 이용한 침입탐지시스템의 한계점인 데이터 불균형으로 인한 오탐율 개선에 기여할 수 있을 것으로 보인다.

마지막으로 SeqGAN을 이용했을 경우 트래픽의 특성에 대한 독립성을 고려하지 않고 생성함에 따라 각 특성의 상관관계에 대한 학습이 어렵다. 이를 해결하기 위한 GAN 방법에 대한 추가적인 연구가 필요하다.

REFERENCES

[1] K. Wang & S.J. Stolfo. (2004, September). Anomalous payload-based network intrusion detection. *RAID.* (pp. 203-222). Berlin : Springer.

[2] N. Williams, S. Zander & G. Armitage. (2006). A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. ACM SIGCOMM Compute Commun, *Rev, 36(5),* 5-16.

[3] UCI KDD Archive. (2005) *kdd aRCHIVE. KDDcup99 dataset.* KDD [Online]. https://kdd.ics.uci.edu/databases/kddcup99/task.html

[4] L. Dhanabal & S. P. Shantharajah. (2015). A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms. *International Journal of Advanced Research in Computer and Engineering, 4(6),* 446-452.

[5] N. V. Chawla et al. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research. 16,* 321-357.

[6] S. Hu et al. (2009). MSMOTE: Improving classification performance when training data is imbalanced. *2009 Second international workshop on computer science and engineering,* (2, pp.13-17). IEEE.

[7] L. Yu et al. (2017). Seqgan: Sequence generative adversarial nets with policy gradient. *Thirty-First AAAI Conference on Artificial Intelligence.*

[8] N. Moustafa & J. Slay. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *Military communications and information systems conference(MilCIS),* IEEE.

[9] B. Dong & X. Wang. (2016). Comparison deep learning method to traditional methods using for network intrusion detection. *2016 8th IEEE International Conference on Communication Software and Networks(ICCSN),* (pp.581-585). IEEE.

[10] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas & J. Lloret. (2017). Network traffic classifier with convolutional and recurrent neural networks for internet of things. *IEEE Access, 5,* 18042-18050.

[11] R. K. Rahul et al. (2017). Deep learning for network flow analysis and malware classification. *International Symposium on Security in Computing and Communication.* Singapore : Springer.

[12] T. Auld, A. W. Moore & S. F. Gull. (2007). Bayesian neural networks for internet traffic classification. *IEEE Transactions on Neural Networks, 18(1),* 223-239.

[13] W. WANG et al. (2017). Malware traffic classification using convolutional neural network for representation learning. *2017 International Conference on Information Networking(ICOIN),* (pp. 712-717). IEEE.

[14] T. Mikolov, K. Chen, G. Corrado & J. Dean. (2013). *Efficient estimation of word representations in vector space.* arXiv preprint.

[15] V. Nair & G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning(ICML-10),* (pp. 807-814).

[16] Z. Zivkovic. (2004, August). Improved adaptive Gaussian mixture model for background subtraction. *ICPR,* (2, pp. 28-31), IEEE.

[17] X. Zhang, J. Zhao & Y. LeCun. (2015). Character-level convolutional networks for text classification. *Advances in neural information processing systems.* (pp. 649-657).

이 우 호(Wooho LEE) [정회원]

· 2016년 2월 : 순천대학교 정보통신학과 공학사
· 2018년 2월 : 전남대학교 정보보안협동과정 이학석사
· 2018년 3월 ~ 현재 : 전남대학교 정보보안협동과정 박사과정
· 관심분야 : 정보보호, 딥러닝, 네트워크 보안
· E-Mail : leeouho@naver.com

함 재 균(Jaegyoon Hahm) [정회원]

· 1999년 2월 : 전남대학교 전산학과 학사
· 2002년 2월 : 한국과학기술원 전산학 석사
· 2014년 2월 : 서울대학교 계산과학전공 박사 수료
· 2002년 3월 ~ 현재 : 한국과학기술정보연구원 책임연구원
· 관심분야 : 클라우드, HPC, 딥러닝, 클라우드보안
· E-Mail : jaehahm@kisti.re.kr

정 현 미(Hyun Mi Jung) [정회원]

· 2014년 2월 : 한남대학교 컴퓨터공학전공(공학박사)
· 2012년 10월 ~ 현재 : 한국과학기술정보연구원 선임연구원
· 관심분야 : HPC, HPC 보안, 서버HW 보안, 클라우드 컴퓨팅
· E-Mail : hmjung@kisti.re.kr

정 기 문(Kimoon Jeong) [정회원]

· 1999년 2월 : 전남대학교 전산학과 학사
· 2001년 8월 : 전남대학교 전산통계학과 석사
· 2009년 8월 : 전남대학교 정보보안협동과정 이학박사
· 2001년 7월 ~ 2004년 12월 : 한국정보보호진흥원
· 2004년 12월 ~ 2005년 6월 : 국가사이버안전센터
· 2005년 6월 ~ 현재 : 한국과학기술정보연구원 선임연구원
· 관심분야 : 딥러닝, 네트워크보안, 클라우드보안
· E-Mail : kmjeong@kisti.re.kr