



## GAN모델과 DNN모델을 활용한 Anomaly Detection에 관한 연구

A Study on the Anomaly Detection using GAN model and DNN model

---

저자 (Authors)	강경석, 이충현, 장건희, 조성래 Kang Kyeung Seok, Lee Chung Hyun, Jang Gun Hee, Cho Sung Rae
출처 (Source)	<a href="#">한국통신학회 학술대회논문집</a> , 2018.11, 260-261 (2 pages) <a href="#">Proceedings of Symposium of the Korean Institute of communications and Information Sciences</a> , 2018.11, 260-261 (2 pages)
발행처 (Publisher)	<a href="#">한국통신학회</a> Korea Institute Of Communication Sciences
URL	<a href="http://www.dbpia.co.kr/Article/NODE07565090">http://www.dbpia.co.kr/Article/NODE07565090</a>
APA Style	강경석, 이충현, 장건희, 조성래 (2018). GAN모델과 DNN모델을 활용한 Anomaly Detection에 관한 연구. 한국통신학회 학술대회논문집, 260-261.
이용정보 (Accessed)	중앙대학교 165.194.35.*** 2019/01/28 11:46 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독 계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# GAN모델과 DNN모델을 활용한 Anomaly Detection에 관한 연구

강정석, 이충현, 장건희, 조성래

중앙대학교

kkskang@uclab.re.kr

## A Study on the Anomaly Detection using GAN model and DNN model

Kang Kyeung Seok, Lee Chung Hyun, Jang Gun Hee, Cho Sung Rae

Chung-Ang Univ

### 요약

본 연구에서는 GAN모델과 DNN모델을 활용한 Anomaly Detection 기법에 관해 연구한다. 최근 급격하게 늘어나고 있는 Advanced Persistent Threat (APT)와 다양한 방법을 활용한 해킹방식이 나오고 있는 상황에서, 기존 detection 기법인 misuse, list 보안기법들이 한계를 나타내고 있다. 하여 이를 보완하기 위한 다양한 머신러닝 기법 중, GAN(적대적생성신경망)과 DNN(심층신경망) 모델을 활용한 새로운 Anomaly Detection기법을 제안하고 실험한다. 연구의 목표는 부족한 데이터셋으로 인한 모델 학습이 충분히 되지 못하였을 경우, 연구에서 제안한 모델을 이용해 데이터셋을 충분히 생성하고 모델 학습을 하여 최종 탐지율을 높이는 데에 있다.

### I. 서론

침입 탐지 시스템은 크게 두 가지 방식으로 나뉘어진다. 첫째로 기존에 사용되는 기법 대부분 Misuse Detection을 바탕으로 탐지하는 것이다. 다른 말로 Signature-Base 기반 탐지라고도 불린다. 이 기법은 미리 설정된 Rule을 기준으로 이상과 정상을 구분하는 방법이다. 하지만 이 기법은 정해진 Rule과 일치하는 경우에 Intrusion으로 판단하므로 물과 테이터에 대한 의존도가 매우 높다는 단점이 있다. 이러한 단점은 다양하고 복잡한 형태를 가진 최신 해킹기법에 대한 치명적인 단점으로 부각되고 있다. 하여 최근에 주목받는 방식은 Anomaly Detection 방식이다. 이 기법은 비정상적인 행위나 컴퓨터 자원을 사용하여 탐지하지만 구현비용이 크다는 단점이 있었다. 하지만 머신러닝 기술이 발전되면서 다양한 모델을 통해 구현비용을 줄이고 빠른 속도로 개선되었다.

연구에서는 이러한 정보를 바탕으로 다음과 같은 Anomaly detection 모델을 제안한다.

1. DNN을 이용해 정상&비정상을 구분하고 결과 중에 오탐지된 데이터셋을 따로 수집한다.
2. 수집된 데이터셋을 GAN모델의 real data에 넘겨주어 이를 바탕으로 새로운 데이터를 생성한다.
3. 생성된 데이터를 원본 DNN학습 데이터에 추가하고 이를 이용해 DNN을 학습시켜 탐지율을 높인다.

다음 목차에서는 연구에 사용된 데이터셋, 제안된 모델에 대해 설명을 하고 결과를 가지고 이를 비교 분석한다.

### II. 본론

연구에 사용된 데이터셋 NSL KDD는 42의 Feature로 이루어진 총 125,973개의 데이터셋이다. 기존의 KDD.09 데이터셋은 Redundant Data가 너무 많아 학습에 bias가 발생한다는 문제점이 있었다. 이러한 이유로 개선된 데이터셋인 NSL KDD를 사용하게 되었다. 하지만 NSL KDD 또한 완벽한 데이터셋은 아니다. 그러나 현재 최근 연구에서 기준으로 사용되는 데이터셋이라는 점과 특정 Anomaly에 대한 분류가 잘 되어있다는 점이 연구에서 보여주고자 하는 목표와 잘 부합되기에 이번 연구에서는 NSL KDD 데이터셋을 사용한다. NSL KDD 125,973개 중 임의로 샘플링된 120,000개의 데이터셋만을 사용한다. 100,000개는 학습을 위한 Training Dataset으로 20,000개는 실험을 위한 Test Dataset으로 사용한다.

<Table1> Anomal to Code

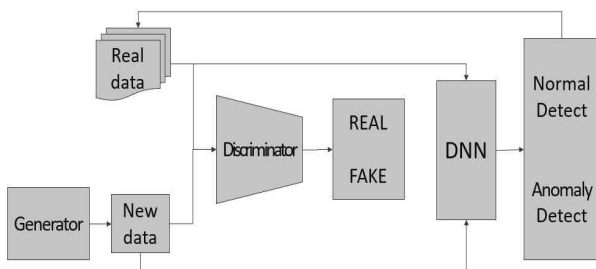
Anomal code	normal	apache2	back	buffer_overflow	ftp_write	guess_passwd
	0	1	2	3	4	5
Anomal code	httptunnel	imap	ipsweep	land	loadmodule	malibomb
	6	7	8	9	10	11
Anomal code	mscan	multihop	named	neptune	nmap	perl
	12	13	14	15	16	17
Anomal code	phf	pod	porstweep	processtalb	ps	rootkit
	18	19	20	21	22	23
Anomal code	saint	satan	sendmail	smurf	snmpgetattack	snmpguess
	24	25	26	27	28	29
Anomal code	sqlattack	teardrop	udpstorm	warezmaster	warezclient	worm
	30	31	32	33	34	35
Anomal code	xlock	xsnoop	xterm	spy		
	36	37	38	39		

NSL KDD의 41번째 Feature는 Label Data로써 normal부

터 spy까지 총 40개로 분류되어있다. 우리는 이 중 34번째 Anomaly인 Warezcclient Label<Table1>를 사용하여 연구를 한다. 또한, integer형 데이터는 각 Column의 가장 큰 수에 비례하여 소수 6자리까지 값을 가질 수 있는 0~1사이의 값으로 Normalization한다. 그 밖에 string형 데이터를 가진 service, protocol, flag는 각각의 코드변환표를 가지고 0부터 70, 3, 11까지로 Transformation 한다. <Table2>는 결과물이다.

<Table2>		Transformation														
Before	2	tcp	ftp_data	SF	12983	0	0.61	0.04	0.61	0.02	0	0	0	0	normal	21
	0	tcp	private	REJ	0	0	0	0.06	0	0	0	0	1	1	neptune	21
	0	tcp	telnet	SF	0	0	0.5	0.01	0	0	0	0.66	0.32	0	mscan	9
	0	tcp	ftp_data	SF	334	0	1	0	1	0.2	0	0	0	0	warezclic	15
	0	icmp	eco_j	SF	18	0	1	0	1	1	0	0	0	0	ipsweep	18
After	2	1	1	10	12983	0	0.61	0.04	0.61	0.02	0	0	0	0	0	0
	0	1	3	2	0	0	0	0.06	0	0	0	0	1	1	15	
	0	1	2	10	0	0	0.5	0.01	0	0	0	0	0.66	0.32	12	
	0	1	1	10	334	0	1	0	1	0.2	0	0	0	0	34	
	0	3	8	10	18	0	1	0	1	1	0	0	0	0	8	

연구에서는 <Fig1>과 같은 Flow Diagram을 따른다. <Fig1>은 GAN과 DNN 모델을 함께 이용하는 탐지모델이다. Flow는 다음과 같다. 처음 Training Data를 가지고 DNN을 학습시킨다. 학습된 DNN에 Test Data를 통과시켜 정탐(Normal Detect),오탐(Anomaly Detect)을 판별한다. 이 중 오탐된 경우의 데이터셋만 따로 모아 GAN의 Real Data로 보낸다. Real Data를 이용해 GAN 모델을 Training하여 New Data를 생성한다. 이때 생성된 데이터는 오탐된 데이터와 비슷한 값을 가지게 된다. 이렇게 생성된 New Data와 기존 Training Data를 합쳐 DNN를 다시 Training한다. 이후에 DNN를 통해 다시 Test Data를 판별한다.



<Fig1> GAN+DNN 모델 flow diagram

연구에서는 GAN 모델에게 Anomaly Detect 데이터를 보내고 이를 바탕으로 New data 20,000개 생성하였다. 하지만 GAN 모델은 생성된 데이터가 점차 정교해지는 특징을 가지고 있으므로, 20,000개 중 후에 만들어진 10,000개의 데이터셋을 기존 NSL-KDD데이터에 추가시켰다. 이로써 최종 실험에 사용된 Training 데이터셋은 110,000개의 데이터를 가지게 된다.

<Table3>는 실험데이터를 보여준다. 실험은 총 5회 반복하였다. DNN모델은 1000번 학습시키고 Test Data에 대한 정확도를 98.5% 이상을 모두 보여주었다. Test Data(20,000) 중 Warezcclient Label Data는 137개가 있다. Before와 After

Column은 오탐된 Warezcclient Data/Total Warezcclient Data를 나타낸 것이다. Before는 아직 GAN모델로 New data를 생성하기 전에 테스트한 결과이고 After는 생성하여 새롭게 Training한 후에 테스트한 결과이다. 표를 보아 알 수 있듯이 오탐율이 1/4가량 줄어들었으며 정탐율이 평균 61.6%에서 91.6%로 개선이 되었다.

<Table3> Test Data				
Step	Loss	ACC	FP+FN	
			Before	After
1	0.031	98.65%	63	11
2	0.024	98.85%	59	11
3	0.029	98.98%	58	14
4	0.031	98.73%	24	11
5	0.022	98.67%	62	11
FP+FN/Total average			38.40%	8.40%

\*FP=False Positive,FN=False Negative

### III. 결론

본 연구에서는 머신러닝 기법인 GAN과 DNN모델을 이용하여 기계학습 침입 탐지 모델을 실험하였다. 모든 Anomaly Label에 대해 실험하지는 않았지만 그 중 하나인 Warezcclient Label을 통해 기존 탐지율(61.6%)에서 향상된 탐지율(91.6%)을 보여주었다. 이번 연구를 통해 기존 모델에서 데이터셋이 부족하여 모델에 효과적인 학습되지 않았던 기존의 한계를 보완하고 개선할 수 있게 되었다.

### ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학 사업의 연구결과로 수행되었음(1711073399)

### 참 고 문 헌

- [1] 강재곤,조성훈,김진, "Neural Network 기반의 분류 알고리즘을 통한 침입탐지 방법연구",2018년도 한국인터넷정보학회 춘계학술발표대회 논문집 제 19권1호, 2018
- [2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. "Generative Adversarial Nets", NIPS2014, 2014
- [3] Houssam Zenati, Chuan-Sheng Foo, Bruno Lecouat, Gaurav Manek, Vijay Ramaseshan Chandrasekhar, "EFFICIENT GAN-BASED ANOMALY DETECTION", ICPR 2018, 2018
- [4] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, Georg Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery", IPMI2017 ,17 Mar 2017