GAN모델과 DNN모델을 활용한 Anomaly Detection에 관한 연구

강경석, 이충현, 장건희, 조성래

중앙대학교

kskang@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

요 약

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. 서 론

침입 탐지 시스템은 크게 두 가지 방식으로 나누어진다. 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 룰과 데이터에 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.

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

<표1> Anomal to Code table

NSL KDD의 41번째 Feature는 Label Data로써 normal부터 spy까지 총 40개로 분류되어있다. 우리는 이 중 34번째 Anomaly인 Warezclient Label<표1>를 사용하여 연구를 한

다. 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.

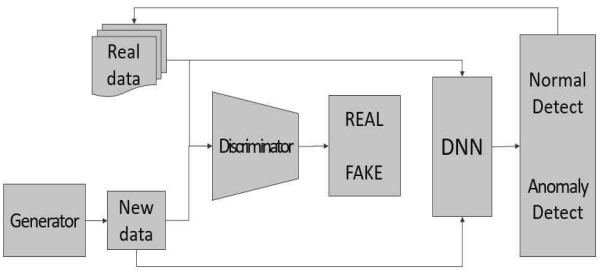
0	0	1	1	0	0	neptune
0	0	0	0	1	1	neptune
0	0	1	1	0	0	neptune
0.12	0.03	0	0	0	0	normal
1	0.2	0	0	0	0	warezclient

<표2> Before Normalization

0	1	1	0	0	15
1	0	0	0.98	1	8
0.04	0	0	0	0	0
0	0	0	0.02	0.02	2
0	1	1	0	0	15
0	0	0	0.02	0.02	2

<표3> After Normalization

연구에서는 <그림1>과 같은 Flow Diagram을 따른다. <그림1>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)를 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.



<그림1> 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 데이터셋를 기존 NSL-KDD데이터에 추가시켰다. 이로써 최종 실험에 사용된 Training 데이터셋은 110,000개의 데이터를 가지게 된다.

STEP	LOSS	ACC	Warez_client/Total(137)	
			Before	After
1000	0.031	98.65%	45%(63/137)	8%(11/137)
1000	0.024	98.85%	43%(59/137)	8%(11/137)
1000	0.029	98.98%	42%(58/137)	10%(14/137)
1000	0.031	98.73%	17%(24/137)	8%(11/137)
1000	0.022	98.67%	45%(62/137)	8%(11/137)

<표4> Test Data

<표4>는 실험데이터를 보여준다. 실험은 총 5회 반복하였다. DNN모델은 1000번 학습시키고 Test Data에 대한 정확도를 98.5% 이상을 모두 보여주었다. Test Data(20,000) 중 Warezclient Label Data는 137개가 있다. Before와 After Column은 오탐된 Warezclient Data/전체 Warezclient 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.

Ⅲ. 결론

본 연구에서는 머신러닝 기법인 GAN과 DNN모델을 이용하여 기계학습 침입 탐지 모델을 실험하였다. 모든 Anomaly Label에 대해 실험하지는 않았지만 그 중 하나인 Warezclient Label을 통 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,조성홍,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