

머신러닝과 딥러닝을 활용한 악성 패킷 탐지 기술 연구

안 병 옥*, 이 중 찬**, 최 재 성**, 박 원 형***

요 약

현재, 5G 및 IoT 기술의 발달함에 따라 실생활에 사용하는 사물들에 네트워크로 연결되어 사용되고 있다. 하지만, 네트워크로 연결된 컴퓨터를 악의적인 목적으로 사용하려는 시도가 증가하고 있으며, 사용자 정보의 기밀성 및 무결성을 침해하는 악성코드를 이용한 공격은 더욱 지능화되고 있다. 이에 대응하기 위한 방안으로 보안관계 시스템과 AI 기술인 지도 학습을 이용한 악성 패킷 탐지 방법에 대한 연구가 진행되고 있다. 사이버보안 관계 시스템 운영상 인력 및 비용 측면에서 비효율적으로 운영되고 있다. 또한, 코로나19 팬데믹 시대에 원격 근무가 증가하여 즉각적인 대응에 어려움이 있다. 그리고 기존 AI 기술인 지도 학습을 이용한 악성코드 탐지에는 변종 악성코드를 탐지하지 못하고 데이터의 양과 질에 따라 부정확한 악성코드 탐지율을 가진다. 따라서, 본 연구에서는 다양한 머신러닝과 딥러닝 모델을 통해 악성 패킷 탐지 기술을 융합하여 악성 패킷 탐지 정확도를 높이고 오탐률과 미탐률을 감소시키며 새로운 유형의 악성 패킷이 침입시 이를 효율적으로 탐지 할 수 있는 악성 패킷 탐지 기술을 제안 한다.

Malicious Packet Detection Technology Using Machine Learning and Deep Learning

Byounguk An*, JongChan Lee*, JeSung Chi*, Wonhyung Park**

ABSTRACT

Currently, with the development of 5G and IoT technology, it is being used in connection with the things used in real life through a network. However, attempts to use networked computers for malicious purposes are increasing, and attacks using malicious codes that infringe the confidentiality and integrity of user information are becoming more intelligent. As a countermeasure to this, research is being conducted on a method of detecting malicious packets using a security control system and AI technology, supervised learning. The cyber security control system is being operated inefficiently in terms of manpower and cost. In addition, in the era of the COVID-19 pandemic, remote work has increased, making it difficult to respond immediately. In addition, malicious code detection using the existing AI technology, supervised learning, does not detect variant malicious code, and has an inaccurate malicious code detection rate depending on the quantity and quality of data. Therefore, in this study, by converging malicious packet detection technologies through various machine learning and deep learning models, the accuracy of malicious packet detection is increased, the false positive rate and the false positive rate are reduced, and a new type of malicious packet can be efficiently detected when intrusion. We propose a malicious packet detection technology.

Keywords : Machine Learning, Deep Learning, Reinforcement learning, Guidance Learning, Malicious Code Detection System

접수일(2021년 08월 29일), 수정일(2021년 10월 18일),
게재확정일(2021년 10월 30일)

* 주 저 자 : 상명대학교 정보보안공학과 학부생
** 공동저자 : 상명대학교 정보보안공학과 학부생
*** 교신저자 : 상명대학교 정보보안공학과 부교수

1. 서 론

IT 기술의 발달로 인한 인터넷 사용 증가와 4차 산업 핵심기술인 IoT, 5G, 인공지능(AI) 등의 기술 발전에 따른 대중화로 인하여 사이버 공간과 현실 공간이 빠르게 융합하고 있다[1].

미국의 통신장비 기업 Cisco 연례 인터넷 보고서에 따르면 전 세계 인터넷 사용자 수는 2018년 51%에서 2023년 66%까지 증가할 것으로 전망하고 있다[2]. 인터넷 사용자 수가 증가함에 따라 사이버보안 피해도 증가할 것으로 전망하고 있다.

사이버 시큐리티 벤처스 자료에 따르면, 2021년 기준 6조 9,390억 달러에 달하는 사이버 피해 규모가 4년 후인 2025년, 10조 5,000억 달러까지 크게 증가할 것으로 예상된다[3]. 특히 사이버 공격 중, 사용자 정보의 기밀성 및 무결성을 침해하는 악성 패킷을 이용한 공격은 더욱 지능화되고 다양한 방식으로 동작하기 때문에 정보보호의 주요한 이슈이다. 악성 패킷으로 인한 보안 위협을 해결하기 위해 24시간 악성 패킷 침입 탐지 관제 시스템이 도입되어 있지만, 이는 관제사의 지속적인 모니터링이 필요하다. 따라서 관제 시스템 운영상 인력 및 비용 측면에서 비효율적이며, 원격 근무 시 즉각적인 대응이 어렵다는 한계가 존재한다. 이와 같은 문제를 해결하기 위해 관제사의 업무 부담을 30~50% 감소시키고, 수동 분석 및 대응 시 발생할 수 있는 누락 문제를 해결하기 위한 AI 기반 악성 패킷 탐지 연구가 활발히 진행되고 있다. 하지만, 기존 AI 기술인 지도 학습을 이용한 악성 패킷 탐지 방법은 데이터의 양과 질이 결과 값에 큰 영향을 미치는 오류가 존재한다. 예를 들어 부족한 데이터 학습으로 부정확한 탐지를 보이는 성향적 오류와 과도한 학습 데이터 사용으로 사소한 값에 민감하게 변화하는 결과 값을 보이는 오버피팅 오류가 존재한다[4]. 또한 알려진 네트워크 패킷을 학습하여 정상/악성을 분류하는 방법으로 탐지하기 때문에, 신종 및 변종 네트워크 패킷이 빠른 속도로 등장하고 있는 현실 세계에서는 적합하지 않다. 이러한 문제를 효과적으로 개선하기 위하여 머시러닝 알고리즘 중 강화 학습을 통한

악성 패킷 탐지 방안을 제시한다. 논문의 구성은 다음과 같다. 2장은 연구 배경을 소개한다. 3장은 구현 방법, 4장은 실험 과정, 5장은 결론을 요약하고 향후 연구 방향을 제시 한다.

2. 관련 연구

2.1 기술 발전과 보안 위협 증가

5G 및 IoT 기술의 발달에 따라 실생활에 사용하는 사물들에 크고 작은 컴퓨터가 탑재되며, 이러한 사물들은 네트워크로 연결되어 있다. 각 컴퓨터들이 네트워크로 연결되어 있어 컴퓨터를 악의적인 목적으로 사용하려는 시도가 증가하고 있다. 대표적인 사례로 금융감독원에 따르면, 국제 해커 집단인 '아르마다 콜렉티브'가 우리나라 은행에 비트코인을 요구하며 협박 서한을 보냈지만 반응하지 않자 디도스 공격을 수행했다[5]. 또한 미국 네트워크 보안 업체인 Sonicwall에 따르면, 전 세계 IoT 기기에 대한 공격 시도가 2020년 1월~6월 2020만 건에서 2021년 1월~6월 3220만 건으로 약 59% 증가했다[6]. 이러한 사이버 보안 위협으로부터 시스템을 보호하기 위해서 AI 기술인 지도 학습을 이용한 악성 패킷 탐지 방법에 대한 연구가 진행되고 있다. 또한, 악성 패킷을 탐지하기 위한 보안 관제 시스템이 운영중이나 사람이 24시간 모니터링 해야 하는 부분에 있어 비용적, 운영적 한계가 발생하고 있다. 이에 따라서 강화 학습을 활용한 네트워크 트래픽 기반 악성 패킷 탐지 방안 마련이 요구되고 있다.

2.2 지도학습을 이용한 악성 패킷 탐지

지도 학습은 훈련한 데이터로부터 하나의 함수를 유추하기 위한 머신러닝의 한 방법이다. 훈련 데이터에 레이블 즉 정답지가 포함되어 이것을 활용하여 값을 유추하는 것이다. 지도 학습의 모델에는 분류(Classification)와 회귀(Regression)가 있다. 분류는 미리 정의되거나 가능성 있는 여러 개의 클래스 레이블(Class Label) 중에서 하나를 예측한다. 회귀는 유추된 함수 중 연속적인 값을 출

력한다. 이러한 방법은 사람이 목표 값에 개입하여 정확도가 매우 높다. 그러나 학습 데이터양이 많고, 레이블을 달기 위해 많은 시간이 걸린다는 단점이 존재한다. 또한, 지도학습 방식은 모델을 만드는 것보다 훈련데이터를 만드는 일이 더 어렵다[7]. 잘못된 입력과 타깃을 훈련데이터에 포함시키면 잘못된 모델이 만들어질 수 있고, 데이터의 개수가 너무 적으면 모델을 충분히 훈련시킬 수 없다[7].

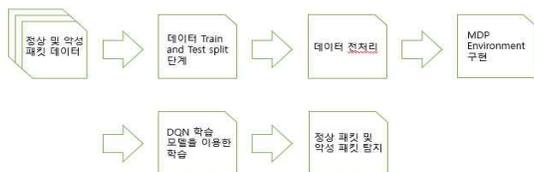
2.3 비지도학습을 이용한 악성 패킷 탐지

비지도 학습은 주어진 입력에 대응하는 출력 정보 없이 학습한다. 즉, 데이터를 분류할 수 있는 정보가 전혀 없이 패턴을 찾거나 데이터를 분류하려고 할 때 사용하는 방법이다. 이러한 방법은 목표값을 정해주지 않으므로 속도가 매우 빠르다는 장점이 존재한다. 그러나, 라벨링 되어있지 않은 데이터로부터 패턴이나 형태를 찾아야 하므로 지도학습보다는 난이도가 있다[8]. 또한, 레이블이 없기 때문에 모델의 훈련 결과를 평가하는데 어려움이 있다[7]. 따라서, 본 연구에서는 지도학습 및 비지도 학습의 문제점과 한계점을 보완하고자 머신러닝의 또 다른 방식인 강화학습을 활용한 네트워크 트래픽을 기반으로 악성 패킷 탐지 방안을 제안 한다.

3. 구현

3.1 구현 환경

본 논문에서는 (그림 1)과 같은 시스템 구성도를 따른다.



(그림 1) 시스템 구성도

기 위해 <표 1>과 같이 구성하였다. 많은 데이터를 분석하기 위해 NVIDIA GPU를 활용하여 고속으로 데이터 전처리를 수행하며, 인공지능 모델을 위해서 Jupyter Notebook의 Python 3.9를 활용하고 데이터 시각화를 위해서 구글 코랩을 사용하였다.

<표 1> 분석환경

분류	소분류	내용
수집 대상	-	raw 네트워크 패킷정보
데이터 처리	전처리	NVIDIA GPU 고속 전처리
데이터 분석	분석 도구	Jupyter Notebook
데이터 분석	분석 도구	python 3.9
시각화	-	구글 코랩 그래프

3.2 Train and Test split 과정

학습 데이터를 데이터 셋의 특징들을 칼럼으로 하여 X 리스트로, Label은 Y 리스트로 나누어주었다. 그 후 오버 피팅(Overfitting)이라는 머신러닝 모델에 Train 데이터로만 학습을 시키고 test 데이터에 모델을 적용했을 때 예상했던 것보다 성능이 나오지 않는 현상을 방지하기 위해 (그림 2)와같이 기존 데이터 셋을 Train data와 validation data를 일정 비율(7:3) 나눈 다음, 학습 시 Train data 셋으로 학습 후 중간중간 validation data 셋을 이용하여 학습한 모델을 검증하는 과정을 거친다.

Feature -> X													Label -> Y	
no	Source	S_start	Destination IP_Port	packets	bytes	packets_per_second	bytes_per_second	packet_size	bytes_per_packet	Duration	bits	bytes	label	
1	172.30.6.255	57041	192.168.24.7600	3	198	3	198	0	0	0.1596394	9.00777	175.2462	benign	
2	172.30.6.255	56688	192.168.24.80	86	67900	34	3483	52	64347	36.46818	22.02818	23166	benign	
3	172.30.15.2348	51682	192.168.12.443	95	82109	34	3246	61	79263	5313.787	409206	3954.895	154958	benign
4	192.30.15.1355	50719	192.24.67.80	11	1382	6	705	5	877	252.611	0.982625	4263.043	7626.873	malicious
5	172.30.15.1355	30864	54.199.137.80	10	1088	6	198	4	890	355.231	1.83015	4362.329	3890.25	benign
6	192.168.43.224	51209	172.217.16.443	23	2915	13	1522	10	809	284.132	1897.788	64.1038	41.60664	benign
7	192.168.43.224	45036	172.82.254.443	43	17305	23	10108	20	7397	3267.27	187.3583	431.6088	313.844	benign
8	192.30.15.1355	34891	172.24.67.80	11	1379	6	705	5	874	1289.378	4.30633	1353.978	1689.534	malicious
9	192.30.15.1355	52843	172.24.67.80	12	1379	6	705	5	874	1027.459	4420.396	1600.992	1576.577	malicious
10	172.30.6.255	56981	192.168.12.443	12	1379	6	705	5	874	1386.696	17.40344	811.0837	282.1943	benign
11	172.30.15.1094	50360	172.240.16.443	7	14001	3	5225	15	174	34483	8.549787	218.4749	1667.3098	benign
12	172.30.15.1355	51504	46.27.226.443	8	1068	3	198	0	902	184.4441	2.04	246.16	563.672	benign
13	172.30.15.1355	56452	78.152.24.5204	3	198	3	198	0	833.247	6.001848	175.0441		benign	
14	192.168.137.97	40430	194.17.161.443	22	8767	13	1323	9	3444	3521.182	2.231135	4743.348	17387.718	benign
15	192.168.43.224	5095	192.168.12.443	289	172945	149	18541	193	15766	223.6427	71.39189	150.9431	1706.106	benign
16	172.30.6.255	58391	214.58.196.80	19	16515	11	9396	28	7267	5526.257	14.96187	4943.144	3885.61	benign
17	192.30.15.1355	5095	192.168.12.443	289	172945	149	18541	193	15766	223.6427	71.39189	150.9431	1706.106	benign
18	192.168.43.224	57270	192.9.882.443	8	144	5	338	3	206	2571.252	14.86807	183.8932	112.1373	benign
19	172.30.6.255	5095	192.168.12.443	2	198	2	198	0	3.348495	6.091474			benign	
20	172.30.6.255	58569	192.217.16.443	15	6281	8	5439	7	626	5879.859	64.93768	102.0508	11387.82	benign
21	192.168.215.49	43093	192.240.16.443	34	2996	18	2953	14	3763	138.7456	46.29373	300.0543	495.6056	benign
12748	192.30.15.1355	60923	172.24.67.80	11	1582	6	705	5	877	1819.947	6.95502	5934.844	7382.778	malicious
12749	172.30.6.255	57675	192.240.16.443	19	1945	10	1317	9	808	3595.007	631.976	1493.071	1053.936	benign
12750	192.30.15.1355	53966	172.24.67.80	11	1379	6	705	5	874	7818.699	2.04675	2755.588	3416.148	malicious
12751	172.30.15.1355	5095	192.168.12.443	11	1382	6	705	5	786	3464.528	7.93254	282.1088	184.579	benign
12752	172.30.15.1355	50363	172.217.16.443	10	1088	6	198	4	522	651.186	44.09518	124.349	94.87979	benign
12753	192.168.137.97	39718	93.117.171.443	16	3003	8	512	8	4463	3623.642	60.85899	276.9495	586.8789	benign
12754	192.168.137.97	44892	168.17.471.443	16	3003	8	512	8	468	262.7199	115.6211	239.956	245.387	benign
12755	172.30.15.1377	50685	194.244.42.443	6	348	4	228	2	120	3849.615	2.534441	718.2683	378.0398	benign
12756	192.168.43.224	51287	238.14636.80	27	12080	15	11120	12	960	2781.992	125.3262	7098.935	61.272938	benign

(그림 2) Train and Test split 결과

실험 환경으로는 네트워크 패킷 데이터를 분석하

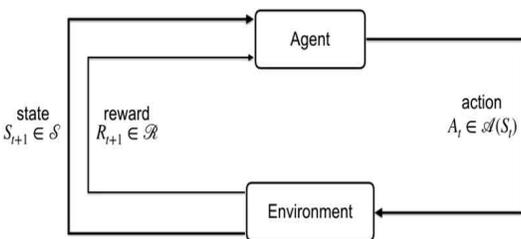
만약 검증 과정에서 예측률이나 오차율이 떨어지는 현상이 발생하면 이것은 모델이 과적합 되었다는 의미이기 때문에 그 즉시 학습을 종료한다. 그 후 과적합 되지 않는 방법으로 학습을 진행한다.

3.3 데이터 전처리

앞에 과정을 거친 데이터 셋을 모델이 이해할 수 있는 형태로 변환하고 데이터의 품질을 올리기 위해 데이터의 실수화와 불완전 데이터, 데이터 노이즈 등을 제거하고 데이터 불균형(과소 포집, 과대 포집)을 해결하는 과정인 데이터 전처리 과정을 수행하였다. 그리고 전처리를 수행한 데이터를 Pandas의 Data Frame을 사용하여 잘못된 정렬을 사전에 방지하고 머신러닝 모델이 데이터값에 손쉽게 접근할 수 있도록 하였다.

3.4 환경 - MDP(마르코프 결정 프로세스)

(그림 3)과 같이 마르코프 결정 프로세스에서는 에이전트(악성 패킷 탐지 시스템)와 환경은 상호 작용한다. 먼저 악성 패킷 탐지 시스템은 환경으로부터 상태를 받고, 이 상태와 경험에 기반한 행동을 선택한다. 이렇게 에이전트가 선택한 행동을 기반으로, 환경은 새로운 상태를 주고 또한 이 행동을 기반으로 보상을 준다. 이와 같이 에이전트가 현재 상태에서 다음 상태로 가는 것을 상태 전환이라고 한다[9].



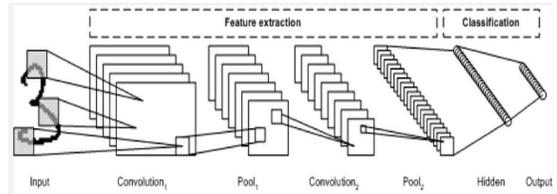
(그림 3) 마르코프 결정 프로세스

(그림 3)과 같은 과정을 반복하여 에이전트와 환경은 상태와 행동의 튜플을 만들게 되고 어떠한 상태는 이전의 모든 상태에 대한 정보를 포함한다고 가정하는데, 이를 마르코프 속성이라고 한다.

이러한 마르코프 속성에 의해 MDP에서 모든 상태를 예측할 수 있는 상태가 되는 것이다.

3.5 DQN(Deep Q-Network) 학습모델

Deep Q-Network(DQN)는 심층 큐 네트워크로 다층 신경망과 큐 러닝을 조합한 기술이다. DQN은 입력 데이터로 정보 추출을 다루는 입력층, 데이터의 분류를 다루는 은닉층, 데이터를 출력하는 출력층으로 나뉜다. Q-network에서 악성 패킷을 분류하기 위해서 CNN(Convolution Neural Network)을 사용했다.



(그림 4) CNN 구성

CNN은 (그림 4)와 같이 구성되어 있으며 입력층과 은닉층 사이에서 CNN의 입력으로 전처리 과정을 거치며 얻은 8개의 데이터로 특징을 추출하는 컨볼루션 과정을 거친다[10]. 컨볼루션(convolution)은 값들의 합성 곱을 이용해서 특징을 추출한다. 특징이 추출되면 활성화 함수인 ReLU(Rectified Linear Unit)를 지나며 뉴런을 활성화하고 은닉층으로 가중치를 조절하는 역전파를 통해 악성 패킷의 특징이 있으면 1, 없으면 0이 나온다. 특징 데이터에서 모든 특징이 필요하지 않으므로 적당량의 데이터로 만들어 주기 위해 크기를 줄여주는 풀링 과정(polling)을 거친다. ReLU 함수에서 받은 값 0, 1로 은닉층에서 소프트 맥스 함수로 악성 패킷을 탐지한다. 출력층은 소프트 맥스 함수(soft max)에서 나온 확률값으로 악성 패킷 탐지율(보상)로 강화 학습한다.

추가로 Q-network에서 DQN으로 강화 학습을 할 때 층이 점점 깊어질수록 정확한 값을 도출해 내기 위해 최적화(optimizer)를 해야 한다. 최적화를 통해 분리했던 train 데이터와 test 데이터가

일치하는지 평가해주는 손실 함수를 사용한다[11]. 악성 패킷 탐지 예측 확률과 실제로 나온 결과 확률의 오차를 줄이기 위해서 평균 제곱 오차(MSE) 공식을 사용한다.

3.6 Q-learning 학습 모델

Deep Q-Learning으로 강화 학습을 반복적으로 train 하며 악성 패킷 탐지를 진행하고 악성 패킷을 탐지를 다 하는 최종 스텝에 도달했을 때 끝낸다. 시간적 상관관계를 없애기 위해서 업데이트를 위한 기록 변수로 리플레이 메모리(Replay Memory)를 생성하고 리플레이 메모리에서 데이터를 무작위로 추출해 학습하는 미니 배치 과정(Mini Batch)으로 강화 학습을 진행한다. 업데이트 전에 입실론 탐욕 정책으로 선택된 행동으로 마르코프 결정 과정 환경과 악성 패킷 탐지율을 보상으로 주고 상호작용하며 얻은 강화 학습한 악성 패킷 탐지율, 강화학습을 위한 행동 등의 값들을 수치화시켜 메모리에 추가한다.

그리고 train과 update를 반복으로 샘플이 메모리에 쌓이기 때문에 메모리 간 상관관계를 깨주기 위해 메모리를 섞어주고(shuffle) 미니 배치 과정에서는 메모리를 활용해 DQN을 개선한다. 입실론 탐욕 정책으로 처음에는 다양한 탐사를 많이 하지만 시간이 지날수록 이용이 많아지게 된다.

업데이트하면서 최적의 행동을 위한 방법은 찾았지만, 보상으로 악성 패킷 탐지율을 받을 때 지금 받는 보상(악성 패킷 탐지율), 1년 후 받는 보상, 5년 후 받는 보상이 각각 다르다. 시간이 지날수록 보상이 달라지며 현재 보상보다 작을 수밖에 없기 때문에 Q 값에 미래보상에 대한 것을 다뤄야 한다.

$$\hat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \hat{Q}(s,a')$$

(수식 1) 벨만 방정식

이때 (수식 1)의 벨만 방정식을 이용한다. 현재 보상(r)과 미래 보상이 낮을 때 사용하는 r(gamma), 그리고 최적의 행동을 할 때 Q 값이 제일 높은 MaxQ로 Q행동 가치 함수를 도출해낸다. 벨만

방정식은 가치 함수와 행동 함수의 관계를 나타낸다. 그러므로 벨만 방정식을 사용하면 상태에 대한 보상(R)과 합쳐서 가치함수로 상태의 가치를 확률적으로 예측한 최적의 정책과 상태를 기반으로 행동을 계산해 최적의 행동을 도출하는 Q 행동 가치함수를 사용했다.

4. 실험 및 평가

4.1 실험 결과

본 연구의 강화학습을 활용한 악성 패킷 탐지 모델과 기존 악성 패킷 탐지 모델의 성능을 비교하기 위해서 D-Tree 분류, Logistic Regression, 인공신경망, SVM, Random Forest의 5개 알고리즘을 사용하여 3가지 방법의 성능 평가 실험을 진행하였다. 첫 번째 방법으로는 기존의 네트워크 악성 패킷에 대한 탐지율을 측정하기 위해 비슷한 패킷 유형의 데이터 셋을 7대 3의 비율로 학습 데이터와 성능 평가 데이터로 나누어 악성 패킷 탐지율을 측정하였다. 실험 결과는 <표 2>를 따른다.

<표 2> 실험 1 결과

	Category	Accuracy	성능 순위	학습 시간
실험 1	D-Tree Classifier	98%	2	1분
	Logistic Regression	84%	6	1분
	Neural Network	95%	4	1분
	SVM	95%	4	1분
	Random Forest	99%	1	1분
	Reinforcement Learning	97%	3	10분

두 번째 방법으로는 적은 양의 데이터로 인한 학습으로 얼마만큼의 악성 패킷을 탐지하는지 평가하기 위해 1,000개의 패킷 데이터로 학습을 시키고 19,000개에 대한 패킷 데이터를 성능 평가 데이터로 활용하여 악성 패킷 탐지율을 측정하였다.

실험 결과는 <표 3>를 따른다.

<표 3> 실험 2

	Category	Accuracy	성능 순위	학습 시간
실험 2	D-Tree Classifier	70%	6	1분
	Logistic Regression	77%	4	1분
	Neural Network	78%	2	1분
	SVM	78%	2	1분
	Random Forest	86%	5	1분
	Reinforcement Learning	93%	1	1시간 30분

세 번째 방법으로는 새로운 유형의 네트워크 악성 패킷에 대한 탐지율을 측정하기 위해 기존 데이터 셋으로 학습을 시키고 새로운 유형에 네트워크 패킷인 데이터 셋으로 성능평가를 진행하였다. 실험 결과는 <표 4>를 따른다.

<표 4> 실험 3

	Category	Accuracy	성능 순위	학습 시간
실험 3	D-Tree Classifier	34%	3	1분
	Logistic Regression	67%	2	1분
	Neural Network	24%	5	1분
	SVM	24%	5	1분
	Random Forest	34%	3	1분
	Reinforcement Learning	80%	1	1시간 30분

4.2 실험 결과 분석

기존의 패킷 유형을 탐지하는 실험 1에서는 지도 학습과 강화학습 두 가지 방법 모두에서 높은 탐지율을 보였다. 하지만 학습 시간 측면에서는 기존 악성 패킷 모델이 강화학습을 활용한 악성 패킷 모델에 비해서 빠른 속도로 학습을 끝마쳤다. 적은 양의 데이터로 학습한 실험 2와 새로운 유

형의 네트워크 악성 패킷에 대한 탐지율을 측정하는 실험 3에서는 강화 학습을 활용한 악성 패킷 탐지 모델이 학습 시간은 많이 걸렸지만, 기존 악성 패킷 탐지 모델에 비해서 높은 탐지율을 보여 주었다.

5. 결론

본 논문에서 강화 학습을 활용하여 효율적으로 악성 패킷 탐지를 하는 방안을 제안하였다. 연구 결과에 대한 시사점은 다음과 같다.

첫 번째, 데이터의 개수가 너무 작으면 모델을 충분히 훈련시킬 수 없다는 지도학습의 단점을 보완하여 훈련데이터를 만드는 시간을 단축할 수 있을 것으로 판단된다. 두 번째, 본래의 지도학습을 이용한 악성 패킷 탐지 시스템의 문제점과 한계점을 극복하여, 새로운 악성 패킷이나 변종 악성 패킷이 나타났을 때도 높은 탐지율로 성능을 검증했으며 앞으로 강화 학습이 보안 분야에 효율적으로 활용할 수 있을 것으로 기대된다. 향후 추가 연구로는 DQN 방식을 사용하는 강화학습은 Q-learning 값을 update하는 식에서 max 연산자를 사용하기 때문에 Q-value를 실제보다 높게 평가하고 그 결과 학습이 느려지는 경향이 있다. 이를 보완하기 위해서 진화된 인공신경망인 Double DQN이라는 알고리즘을 활용하여 학습 모델의 속도를 높이는 연구가 필요하다.

참고문헌

- [1] Daesung Lee. "차세대 사이버 보안 동향." 한국정보통신학회논문지23.11(2019):1478-1481.
- [2] CISCO, Cisco Annual Internet Report (2018 - 2023) White Paper, March 2020.
- [3] 박성규 기자, 디지털화·코로나 먹고 자란 사이버 테러, ···2025년 글로벌 피해액 1경 넘는다., <https://www.sedaily.com/NewsView/22L80BKPS0>
- [4] GugGyeong-Wan, ByungCheol-Gong, "Trends in Security Technology Development Using Artificial

Intelligence”.

- [5] 원병철 기자, 금융권 디도스? 아르마다 콜렉티브의 새로운 ‘공격 예고’일 뿐!, <https://www.boanews.com/media/view.asp?idx=55458>, 2017.
- [6] Sonicwall, 2021 Sonicwall Cyber Threat Report Mid-Year Update, July 2021.
- [7] 머신러닝 학습방식 3가지, “지도학습/비지도학습 단점” <https://hyjykelly.tistory.com/31> (21.10.12).
- [8] 비지도학습(Unsupervised Learning), “비지도학습 이란” <http://blog.skby.net/%EB%B9%84%EC%A7%80%EB%8F%84-%ED%95%99%EC%8A%B5-unsupervised-learning/> (21.10.12).
- [9] Xi-Lang Huang, Seon Han Choi, “A Simulation Sample Accumulation Method for Efficient Simulation-based Policy Improvement in Markov Decision Proces” Journal of Korea Multimedia Society Vol. 23, No. 7, July 2020(pp. 830-839).
- [10] Jihyeon Park, Taeok Kim, Yulim Shin, Jiyeon Kim, Eunjung Choi, “Design and Implementation of a Pre-processing Method for Image-based Deep Learning of Malware” Journal of Korea Multimedia Society Vol. 23, No. 5, May 2020 (pp. 650-657).
- [11] ZHIYANG FANG, JUNFENG WANG, JIAXUAN GENG and XUAN KAN “Feature Selection for Malware Detection based on Reinforcement Learning“ IEEE Access, vol. 7, 2019, pp.176177-176187.

[저자 소개]



안 병 옥 (Byoung-Uk An)

2017년 ~ 상명대학교 정보보안공학과 학부생

email : dbz1159753@naver.com



이 중 찬 (Joong-Chan Lee)

2017년 ~ 상명대학교 정보보안공학과 학부생

email : dlwndcks150@naver.com



최 재 성 (Jai-Sung Choi)

2017년 ~ 상명대학교 정보보안공학과 학부생

email : choi_jasung@naver.com



박 원 형 (Won-Hyung Park)

2002년 서울과학기술대 산업정보시스템 학사

2005년 서울과학기술대 정보산업공학과 석사

2009년 경기대학교 정보보호학 학사

2015년 성균관대학교 컴퓨터교육학 박사수료

2012년~2020년 극동대학교 사이버보안학과 부교수/학과장

현재 상명대학교 정보보안공학과 부교수

email : whpark@smu.ac.kr