https://doi.org/10.33778/kcsa.2023.23.5.073

Syn Flooding 탐지를 위한 효과적인 알고리즘 기법 비교 분석★

김 종 민*, 김 홍 기**, 이 준 형***

요 약

사이버 위협은 기술의 발전에 따라 진화되고 정교해지고 있으며, DDoS 공격으로 인한 서비스 장애를 발생 이슈들이 증가하고 있다. 최근 DDoS 공격은 특정 서비스나 서버의 도메인 주소에 대량의 트래픽을 유입시켜 서비스 장애를 발생시키는 유형이 많아지고 있다. 본 논문에서는 대역폭 소진 공격의 대표적인 공격 유형인 Syn Flooding 공격의 데이터를 생성 후, 효과적인 공격 탐지를 위해 Random Forest, Decision Tree, Multi-Layer Perceptron, KNN 알고리즘을 사용하여 비교 분석하였고 최적의 알고리즘을 도출하였다. 이 결과를 토대로 Syn Flooding 공격 탐지 정책을 위한 기법으로 효과적인 활용이 가능할 것이다.

Comparative Analysis of Effective Algorithm Techniques for the Detection of Syn Flooding Attacks

Jong-Min Kim*, Hong-Ki Kim**, Joon-Hyung Lee***

ABSTRACT

Cyber threats are evolving and becoming more sophisticated with the development of new technologies, and consequently the number of service failures caused by DDoS attacks are continually increasing. Recently, DDoS attacks have numerous types of service failures by applying a large amount of traffic to the domain address of a specific service or server. In this paper, after generating the data of the Syn Flooding attack, which is the representative attack type of bandwidth exhaustion attack, the data were compared and analyzed using Random Forest, Decision Tree, Multi-Layer Perceptron, and KNN algorithms for the effective detection of attacks, and the optimal algorithm was derived. Based on this result, it will be useful to use as a technique for the detection policy of Syn Flooding attacks.

Key words : DDoS, Syn Flooding, Random Forest, Decision Tree, Multi-Layer Perceptron, KNN

접수일(2023년 05월 19일), 수정일(2023년 06월 20일), 게재확정일(2023년 06월 20일)

★ 본 과제(결과물)는 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(2021RIS-002)

* 동신대학교 정보보안학과 교수(주저자)

** 동신대학교 정보보안학과 교수(공동저자)

*** 동신대학교 신재생에너지학과 교수(교신저자)

1. 서 론

현대 사회에서는 디지털 기술의 급격한 발전으로 인해 네트워크 기반 서비스가 빠르게 확대되고 있다. 이러한 발전과 함께 네트워크의 광범위한 사용 증가로 데이터의 흐름이 급증하면서 개인 및 기업에 업무적 효율성 또한 향상되고 있다. 반면에 사이버 위협의 발생과 위협 역시 증가하고 있다[1].

분산 서비스 거부 공격인 디도스(DDoS)공격은 네트워크 기반 서비스에 대한 심각한 위협 중 하나다[2]. 디도스(DDoS)의 공격 유형 중 Syn Flooding 공격은 TCP의 3-Way-Handshake 과정에서 SYN 패킷을 전송한 후 돌아오는 SYN+ACK 패킷에 대한 응답을 악의적으로 보내지 않아 백로그 큐(Backlog Queue)에 할당된 메모리를 꽉차게 만들어 더 이상 연결을 받아들일 수 없게 하여 서비스 거부 상태로 만드는 악의적인 공격이다. Syn Flooding공격의 개념이 오래전부터 알려지면서 간단하게 실행할 수 있는 공격 툴이나 소스가 만들어졌고 현재까지 이 공격이 빈번하게 발생되고 있다[1][3].

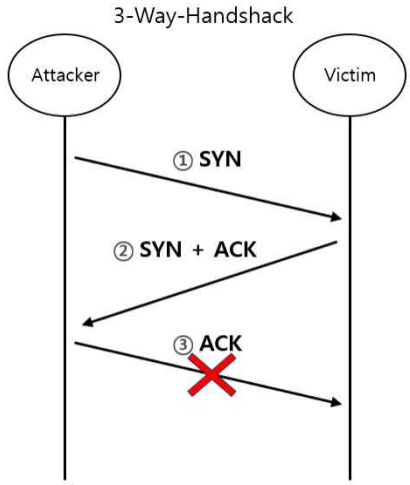
본 연구에서는 Syn flooding 공격 데이터를 생성 후, Random Forest, Decision Tree, Neural Network, KNN 알고리즘을 활용하였다. 진행 절차는 데이터셋을 불러와 전처리를 수행한 다음, 선택한 4가지 알고리즘을 이용하여 학습시킨다. 학습이 완료된 각 모델에서는 Syn Flooding 공격 탐지 정확도 및 속도가 도출되었고, 이 결과를 토대로 효과적인 탐지 기법에 대해 비교 분석하고자 한다.

2. 관련 연구

2.1 Syn Flooding

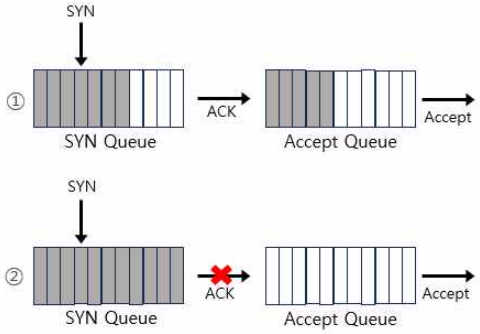
Syn Flooding은 네트워크 보안 공격인 분산 서비스 거부 공격인 디도스(DDOS)중 자원소모 공격이다. TCP(Transmission Control Protocol)의 연결 과정인 3-way handshake의 취약점을 악용한다. 3-way handshake는 TCP 연결을 설정하기 위한 과정으로, 클라이언트가 서버에게 SYN 패킷을 보내고, 서버는 이에 대한 응답으로 SYN+ACK 패킷을 보내며, 클라이언트는 다시 ACK 패킷을 보내어 연결을 확립한다.

(그림 1)과 같이 Syn Flooding은 여러 Attacker가 대량의 ①SYN 패킷을 서버에 보내어 서버의 자원을 고갈시키는 공격이다. Attacker은 실제로 연결을 설정하지 않고, 가짜 ①SYN 패킷을 전송하고 전달 받은 ②SYN+ACK 패킷에 대한 ③ACK 패킷을 전달하지 않는다.



(그림 1) Syn Flooding attack

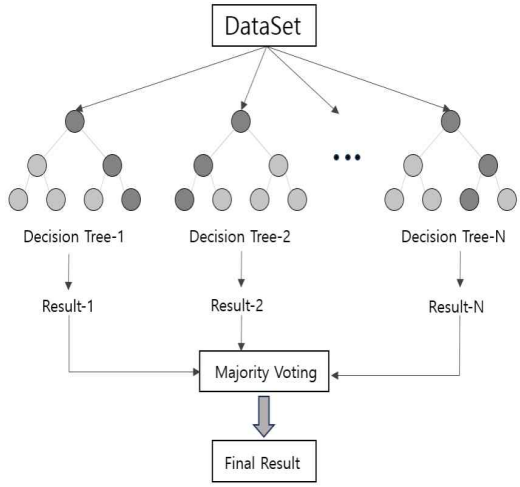
이 과정에서 Victim은 ACK 패킷을 전달 받기 전 까지 “Half Open” 상태가 되며 일정 시간 후 패킷이 오지 않을 경우 연결을 초기화 하게 된다. 하지만 연결을 초기화 하기 전까지 연결은 (그림 2)의 ①과 같이 메모리인 백 로그 큐(Backlog Queue)에 계속 쌓이게 된다. 만일 Attacker의 연결 요청이 초기화하는 속도보다 빠르게 이루어진다면 (그림 2)의 ②와 같이 백 로그 큐(Backlog Queue)의 메모리가 꽉 차게 되어 더 이상의 연결을 받아들일 수 없는 상태가 되게 만드는 자원 소진 고갈 공격이다.



(그림 2) Backlog Queue

2.2 Random Forest

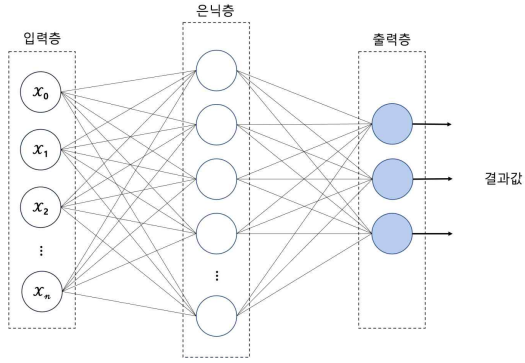
Random Forest는 머신러닝 분야에서 널리 사용되는 알고리즘 중 하나로, 여러 개의 의사결정 트리(Decision Tree)를 생성하고, 이들의 예측을 결합하여 최종 예측값을 도출하는 알고리즘이다[4]. 이는 앙상블 학습을 기반으로 하여 여러 개의 분류기를 생성하고 이를 각 예측 데이터에 결합하여 정확하고 빠른 결과를 얻는 것이 특징이다. 또한 데이터의 일부를 임의로 선택하여 각각의 트리를 훈련하는 부트스트랩 샘플링(Bootstrap Sampling)을 하며, 각 트리는 랜덤하게 선택된 특성들을 사용하여 노드를 분할한다. 이는 다양성을 증가시키고 모델의 일반화 능력을 향상하는 역할을 한다. Random Forest는 무작위로 선택된 특성들의 부분 집합을 사용하고 부트스트랩 샘플링(Bootstrap Sampling)을 통해 다양한 상황을 반영하여 네트워크에서 발생하는 다양한 공격을 효과적으로 탐지가 가능하다는 이점이 있다. 이에 따라 최근에는 네트워크 침입탐지 분야에서의 연구도 활발히 이루어지고 있다.



(그림 3) Random Forest Algorithm

2.3 MLP(Multi-Layer Perceptron)

(그림 4)와 같이 MLP는 입력 계층 (input layer), 은닉 계층(hidden layer), 출력 계층(output layer)으로 이루어져 있으며, 은닉 계층에서는 데이터의 입·출력 과정에서 직접적으로 보이지 않는 숨겨진 특징을 학습하는 역할을 한다. 입력 계층과 연결된 은닉 계층의 한 노드만 보면 하나의 단층 퍼셉트론과 같은 구조를 갖는 것을 확인할 수 있다. 즉, 은닉 계층에 있는 각각의 노드는 한 퍼셉트론의 활성화 함수의 역할을 한다고 볼 수 있다[5].

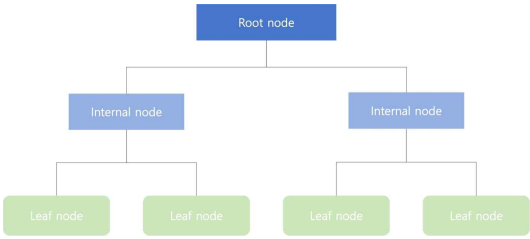


(그림 4) MLP 구조

2.4 Decision tree

Decision tree는 복잡한 의사결정 과정을 단순한 여러 개의 의사결정의 조합으로 구성하여 최종적인 종속변수 예측이 가능한 알고리즘으로 회귀 및 분류에 모두 사용되는 비모수 (Non-parametric)모델이다. 데이터 세트를 특정한 기준을 가지고 분기점을 만들어 나가며 결과를 매칭하는 방법으로 분기점에 의해 나뉘어진 전체적인 모델의 형상이 나무와 같아 의사결정 트리라 한다[6][7].

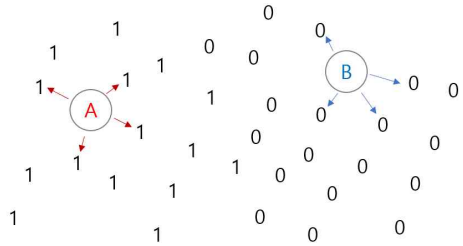
(그림 5)는 Decision tree 구조로서, Root node, Internal node, Leaf node로 계층적 구조를 가지고 있다.



(그림 5) Decision tree 구조

2.5 KNN

KNN(K-Nearest Neighbors)은 머신러닝 분야에서 사용되는 다변량 분석(Multivariate Analysis) 알고리즘이다. 입력 데이터를 특정 값으로 분류를 하는데 유클리드 거리(Euclidean Distance)와 맨해튼 거리(Manhattan distance)를 사용해 현재 데이터와 가장 가까운 K개의 데이터를 찾아리 이를 기반으로 예측을 수행한다[8]. (그림 6)와 같이 K가 4라고 가정 하였을 경우 A는 1, B는 2가 된다.

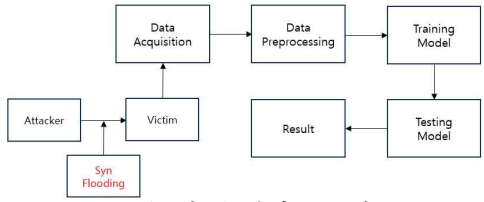


(그림 6) KNN Algorithm

KNN(K-Nearest Neighbors)은 학습 단계가 없으며, 모델 구축이 필요하지 않아 간단하게 사용할 수 있다는 장점이 있다. 또한 수치 기반 데이터 부류 작업에서 성능이 우수하다. 이에 따라 KNN(K-Nearest Neighbors)은 데이터 예측 분야에서 활발히 활용되고 있다.

3. 제안하는 방법

3.1 분석 프로세스



(그림 7) 분석 프로세스

(그림 7)은 분석 프로세스이다. 우선 Attack PC에서 Victim으로 Kali-Linux에서 사용 가능한 hping3을 이용해 Syn Flooding 공격을 수행한다. hping3는 command-line tool로 네트워크 보안 테스트 및 문제 해결에 사용된다[9]. 이 과정에서 WireShark를 사용하여 패킷을 캡처한다. 이후 Tshark를 사용하여 pcap 파일을 csv파일로 저장함으로써 데이터 수집을 한다. 그 후 데이터 전처리을 통해 Ethenet Frame, IP Packet, TCP Segment에서 25개의 Column을 확립하고 이를 기반으로 데이터 셋을 생성한다. 데이터 셋을 이용해 Random Forest, MLP, Decision tree, KNN 4가지 알고리즘을 통해 모델 학습을 진행한 후 모델 테스트을 거쳐 결과를 도출한다. 도출한 결과는 예측 정확도와 속도를 기준으로 하였다.

4. 분석 및 검증

4.1 분석 환경

<표 1>은 Syn Flooding 데이터 수집을 위한 데이터 수집 환경이다.

<표 1> 데이터 수집 환경

구분	Attack	Victim
OS	Kali-Linux x64	Ubuntu 22.04
IP	172.22.70.150	172.22.70.220

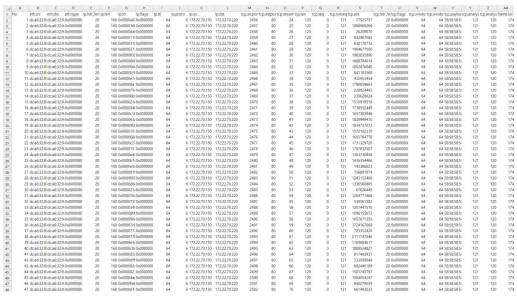
<표 2>는 데이터 분석 환경이다.

.

<표 2> 분석 환경

운영체제	macOS Sonoma 14.2
프로그래밍 언어	Python3.8
에디터	jupyter notebook
라이브러리	- pandas - scikit-learn - matplotlib - numpy 등

4.2 데이터 셋



(그림 8) Data Sample

(그림 8)은 분석에 사용된 데이터이며, 데이터 수집은 2023년 11월 23일 ~ 2023년 11월 27일까지 수집을 하였으며, 그림7은 데이터의 샘플을 나타낸다.

데이터 전처리를 위해 선택한 Columm들은 <표 3>, <표 4>, <표 5>와 같다. Ethernet Frame, IP Packet, TCP Segment에서 총 25개의 Column을 선택하였다.

<표 3> 수집 Ethernet Frame

Ethernet Frame	
eth.src	출발지 MAC Address
eth.dst	목적지 MAC Address
eth.type	Type

<표 4> 수집 IP Packet

IP Packet	
ip.hdr_len	Header의 Length
ip.len	길이
ip.id	식별자
ip.flags	Flag
ip.ttl	Time-to Live 값
ip.proto	상위 Protocol
ip.src	출발지 Address
ip.dst	목적지 Address

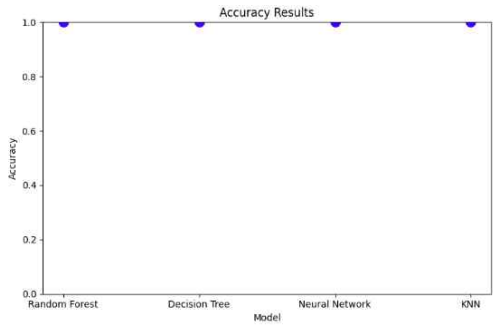
<표 5> 수집 TCP Segment

TCP Segment	
tcp.srcport	출발지 Port 번호
tcp.dstport	목적지 Port 번호
tcp.stream	Stream 번호
tcp.len	길이
tcp.seq	Sequence 번호
tcp.nxtseq	다음 Sequence 번호
tcp.ack	확인 응답 번호
tcp.hdr_len	Header의 길이
tcp.flags	Flag
tcp.window_size_value	Window 크기 값
tcp.window_size	Widow 크기
tcp.payload	Payload 데이터
tcp.analysis.bytes_in_flight	전송 중인 Byte 수
tcp.analysis.push_bytes_sent	전송된 Push Byte 수

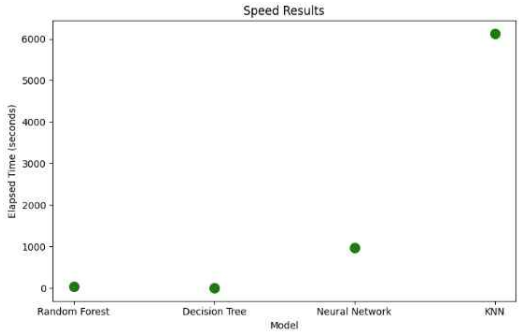
총 84만여개의 데이터를 추출했으며 이를 데이터를 학습용과 테스트용(8:2)으로 분리하여 모델 학습(Training Model)을 진행하였다. 분석 결과로는 예측 정확도와 속도를 이용하였다.

4.3 결과 분석

분석 결과 (그림 9)와 같이 예측 정확도는 Random Forest, MLP, Decision tree, KNN 모든 알고리즘에서 100%로 나왔다.



(그림 9) Accuracy Results



(그림 10) Speed Results

속도에서는 (그림 10)과 같이 Decision tree, Random Forest, MLP, KNN 순으로 나왔으며, <표 6>에서 보이시듯이 예측 정확도와 속도를 종합적으로 비교하였을 때 Decision tree 알고리즘이 Syn Flooding 공격을 탐지하기에 가장 적합한 알고리즘이라는 결과가 나왔다.

<표 6> 알고리즘 비교 분석

구분	Accuracy(%)	Speed(초)
Random Forest	100	45.3119
Decision tree	100	0.5636
MLP	100	970.2907
KNN	100	6129.9145

5. 결론

본 논문은 Syn Flooding 공격을 탐지하기 위한효과적인 머신러닝 알고리즘을 비교 연구하였다. 이를 위해 공격의 특징을 분석하고 데이터 세트를 만든 후 Random Forest, MLP, Decision tree, KNN 4가지의 머신러닝 알고리즘에 학습하여 예측 정확도와 속도를 비교하였다. 비교 결과 모든 알고리즘에서 정확도는 100%로 나타났고, 분석 속도에서 Decision tree가 차이가 많이 나타났다. 속도에서는 0.5636초로 가장 빨랐다. 정확도와 속도를 종합적으로 비교해봤을 때 Decision tree가 Syn Flooding 공격을 탐지하기 위한 최적의 머신러닝 알고리즘이라는 결과를 도출하였다. 이 연구를 토대로 Decision tree 기반 DDoS 탐지를 위한 도구로 유용하게 활용될 것으로 보이며, 향후 연구에서는 Syn Flooding 공격에 대한 탐지 방법 및 방어기법에 대해 연구하고자 한다.

참고문헌

[1] Mi-Ran Han, Keun-Heui Kim, Young-Mo Kang, and Jong-Bae Kim, "SYN Flood Attack Vulnerability & IPS Architecture Research", Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol.6, No.2, pp. 01-08, 2016.

[2] Hee-Sik Choi and Moon-Seog Jun, "DDoS TCP Syn Flooding Backscatter Analysis Algorithm", Journal of the Korea Society of Computer and Information, vol. 14, no. 9, pp. 55-66, 2009.

[3] D. Moore, G. Voelker, and S. Savage, "Inferring Internet Denial-of-Service Activity", ACM Transaction on Computer Systems, Vol. 24, No. 2, pp. 115-139, 2006.

[4] Sung-Min Jang, Jun-Hak Kim, Hee-Jung Kwon, Eun-Hee Oh, and Chang-Min Seo, "Development of Malicious Code Detection System Using RandomForest and XGBoost", 대한전기학회 학술대회논문집, pp. 2465-2466, 2023.

[5] 권홍필, 배대현, 하재철, "Multi-Layer Perceptron 기법을 이용한 전력 분석 공격 구현 및 분석", 정보보호학회논문지, vol. 29, no. 5, pp.997-1006, 2019.

[6] Kidong Lee, Sanghoon Kang, Minjung Kang, Sung Yi, Soongkeun Hyun, Cheolhee Kim, "Modeling of Laser Welds Using Machine Learning Algorithm Part I: Penetration Depth for Laser Overlap Al/Cu Dissimilar Metal Welds", Journal of Welding and Joining, vol. 39, No. 1, pp. 27-35, 2021.

[7] 차기욱, 홍원화, "Decision tree 기반 알고리즘을 활용한 해체폐기물 발생량 예측모델 개발", 대한건축학회논문집, vol. 39, no. 3, pp. 179-187, 2023.

[8] Sun, M, Yang, R., "An efficient secure k nearest neighbor classification protocol with high-dimensional features." Int J Intell Syst, vol.35, issue 11, pp. 1791-1813, 2020. https://doi.org/10.1002/int.22272

[9] Dsouza,D.,Bansal,R.,Krishnan,S.N.,"NETWORK SECURITY TESTING, ANONYMIZING TRAFFIC AND PASSWORD CRACKING : A REVIEW OF HPING3, PROXY CHAINS AND JOHN THE RIPPER TOOLS", International Research Journal of Modernization in Engineering Technology and Science,vol.05,issue05,2023. https://www.doi.org/10.56726/IRJMETS38746.

〔저자소개〕

김 종 민 (Jong-Min Kim)
2015년 산업보안학박사
현 재 동신대학교 정보보안학과 교수

email : dyuo1004@dsu.ac.kr

김 홍 기 (Hong-Ki Kim)
1996년 2월 전남대학교 전산통계학과 (이학박사)
현 재 동신대학교 정보보안학과 교수

email : hkkim@dsu.ac.kr

이 준 형 (Joon-Hyung Lee)
2002년 UC Berkely, 화학공학박사
현 재 동신대학교 신재생에너지학과 교수

email : jhlee@dsu.ac.kr