사이버 공격의 분류와 최신 방어기법에 대한 연구: DDoS 탐지 및 Deep Learning의 활용

이영한, 백세현, 서지원, 방인영, 백윤흥*
*서울대학교 전기정보공학부
e-mail: yhlee@sor.snu.ac.kr

A Study of Cyber Attacks and Recent Defense System: DDoS Detection and Applying Deep Learning

Younghan Lee, Se-Hyun Baek, Jiwon Seo, In-young Bang, Yunheung Paek*
*Dept of Electrical and Computer Engineering, Seoul National University

요 약

사이버 공격은 점차 다양해지고, 그 위험성은 날로 심각해지고 있다. 가장 강력한 공격 중 하나는 DDoS (Distributed Denial of Service) 공격이다. 본 논문에서는 다양한 사이버 공격을 분류하고 이에 따른 방법 기법을 서술하겠다. 특히, 최신 DDoS 공격 탐지 방법을 소개하고 딥러닝 (Deep Learning) 을 활용한 최신 방어 기법 연구에 대해 살펴보도록 하겠다.

1. 서론

사이버 공격은 모든 것들이 서버에 연결되어가는 사회에 더욱 치명적으로 다가온다. 공격의 종류는 다양하고 그 방식은 조금씩 하지만 섬세하게 다르다. 가장 많이 알려지고 강력한 공격은 DoS(Denial of Service) 공격이다. 공격의 종류는 여러 가지가 있고 탐지 방법이 발전함에 따라 공격 또한 변형되고 진화한다. 탐지 기술은 알려진 공격만 막는 것에서 멈추지 않고 비정상 행동을 할 경우 공격으로 간주하는 방법으로 발전했다. 뿐만 아니라, 최근에 모르는 공격도 효과적으로 방어하기 위해서 딥러닝을 활용할 시도를 하고 있다. 2장에서 통상적으로 알려진 사이버 공격을 분류하고, 3장에서는 이를 탐지하는 방어 기법들의 종류를 간략히 알아본다. 본 논문은 DDoS공격과 딥러닝의 활용을 중점적으로 다루고 있으며 최신 사용되고 있는 방어기법에 대해 서술하고자 한다.

2. 사이버 공격의 분류

사이버 공격의 종류는 정말 다양하고, 새로운 공격 방법이 꾸준히 개발되고 있다. 본 장에서는 최신 사용되고 있는 사이버 공격들을 분류하고, 그 공격들의 간단한 설명을 예시를 통해서 소개하도록 하겠다. <표 1>은 주로 사용되는 공격 유형과 간략한 설명을 보여주고 있다.

2.1 DoS 공격

이 공격은 가장 흔히 대중들에게도 알려져 있는 공격임과 동시에 강력한 공격들 중 하나이다. 정의하자면, 공격자가 컴퓨터의 리소스를 최대치 이상으로 사용하게 만듦으로써 권한이 있는 정상 사용자가 컴퓨터 또는 서비스에 접속할 수 없도록 하는 것이다.[1] 공격자가 하나 이상의 VM (Virtual Machine)에서 동시다발적으로 DoS 공격을 시도할 경우 DDoS 공격으로 분류된다. DDoS 공격은 크게 세 가지로 분류 되는데, Volumetric 공격, Application Layer 공격, 그리고 State-exhausting 공격이다.[5] 이 특정 공격에 대해서는 4장에서 자세히 다루도록 하겠다.

2.2 Remote to Local (R2L) 공격

이 공격은 공격자가 로컬로 패킷들을 보내서 취약점을 악용하여 권한을 얻는 것이다. 이런 공격은 공격자가 네트워크를 통해서 로컬로 패킷을 보낼 수 있을 경우에만 가능하다.

2.3 User to Root (U2R) 공격

이 공격은 공격자가 취약점이 있는 로컬에 이미 또 다른, 주로, R2L 공격을 통해 접속해있을 때 가능한 공격이다. 공격자는 일반 사용자 권한으로 공격을 시작해 root권한까지 얻게 된다.

2.4 Probing 공격

Probing은 네트워크를 스캔하면서 취약점을 발견하는 공격이다. 공격자는 네트워크에 존재하는 서비스 맵을 활용해서 공격 포인트를 찾아낸다.

<표 1> 사이버 공격의 분류

공격 범주	공격 타입	설명
DoS	▪ Neptune ▪ Pod	- SYN 포트를 flooding을 통해 공격한다. - 잘 못된 패킷을 연속해서 보내 crash를 유도한다.
R2L	▪ Spy ▪ Ftp_write	- 취약점을 공격해 시스템의 중요한 정보를 유출한다. - FTP 서버를 이용해 로컬 시스템에 접근한다.
U2R	▪ Buffer_overflow ▪ rootkit	- 버퍼오버플로우를 통해 root 권한을 얻는다. - 관리자 권한을 얻는다.
Probing	▪ Portsweep	- 포트들을 스캔해 호스트에 있는 서비스들을 알아낸다.

3. 최신 탐지 기법의 종류

본 장에서는 많이 사용되고 있는 공격 탐지 방법들을 소개하고 특징을 간략히 설명하겠다.

2.1 Intrusion Detection Systems (IDS)

이 탐지 방법은 호스트 안에서 모니터링을 통해 공격을 감지한다. 이 탐지 시스템은 기술적인 면에서 크게 두 가지로 나눌 수 있다. 기존에 알려진 공격을 탐지하는 Signature 기반과 비정상 행위를 탐지하는 Anomaly 기반이 존재한다.

2.2 Signature를 기반으로 탐지

이 탐지 기술이 현재 널리 사용되고 있는 백신 프로그램의 기본 방식이다. 공격의 특징을 분석해서 그 것을 signature로 인식한 후, 데이터 로그에 같은 특징이 나타나는지 감시하는 방법이다. 이미 알려진 공격을 탐지하는 것에 용이하다.

2.3 Anomaly를 기반으로 탐지

호스트에서 비정상적인 행동이 일어나는지 감시하는 공격 탐지 방법이다. 공격이 일어날 경우 정상적이지 않은 패턴이 발견될 것이라는 전제 아래 작동한다. 정적 탐지 방식과 동적 탐지 방식으로 나눌 수 있는데, 정적 방식은 소프트웨어적인 부분만 확인한다.[1] 반면, 동적 방식은 네트워크 트래픽 등에서 비정상적인 행위를 확인한다.

2.4 딥러닝을 활용한 탐지

존재하는 공격들의 99%는 기존 공격이 조금 변형된 것이다.[4] 그렇기 때문에 방대한 데이터에서도 특징들을 잘 분석해내는 딥러닝이 보안 기술로 활용되는 것은 당연한 수순이다. 기존 방식에서는 잡아내지 못 했던 새로운 공격도 더 높은 정확성으로 잡아낼 수 있다. 5장에서 좀 더 자세히 다루도록 하겠다.

4. DDoS 공격의 종류 및 최신 탐지 기법

본 장에서는 DDoS 공격의 종류를 간략하게 구분하고 최신 탐지 기법을 소개하도록 하겠다.

DoS 형태의 공격은 네트워크 공격들 중 3분의 1 이상을 차지하고 있다.[5] DDoS 공격은 하나의 공격자가 여러 좀비 PC들을 만들어서 반복되는 공격을 동시다발적으로 시행하는 것이다. 이 공격은 크게 세 가지로 분류된다. 이 중 Volumetric 공격이 65% 이상의 점유율로 가장 흔하다.[6] 이 공격의 예로는 Flooding 공격을 들 수 있다. 서버가 감당할 수 없는 양의 연결을 시도해서 서버의 리소스를 모두 사용해 작동을 정지시키는 것이다. 두 번째 형태는, Application Layer 공격이다. 이는 서버에 특이한 요청을 보내거나 또는 거대한 파일을 전송시켜서 서버를 마비시키는 것이다. 이 공격의 경구 서버 입장에서는 서버가 정상적으로 작동하고 있다고 인지할 위험성도 있다. 마지막으로는 State-exhausting 공격인데 이는 20% 정도의 점유율을 가지고 있다. 예로는 ping-of-death 공격을 들 수 있다.

DDoS 공격 탐지 방법으로는 크게 세 가지 방법이 있다. Signature Based Approach(SBA), Anomaly Based Approach(ABA) 그리고 Entropy Based Approach(EBA). SBA는 위에 언급한 대로 알려진 공격을 막는 것에 용이하지만, Zero-Day 공격을 방어하지는 못하는 한계를 가지고 있다. 그렇기 때문에 제안 된 ABA는 서버의 트래픽 상태를 감시하고 이상 현상이 발생되면 알람을 띄우는 방식으로 진화했다. 다만, 정상 상태라는 기준이 잘 정해져 있지 않을 경우 잘 못된 알람을 보낼 우려가 있다. 그래서 가장 적합한 방식은 EBA이다. 이 방식은 서버의 특성들의 엔트로피의 변화를 감지하면서 과도한 변화가 있을 경우 공격으로 감지하는 탐지 기법이다. 그러나 이 방법 역시 상당한 계산 시간과 메모리 사용이 강요되기 때문에 이를 간략화 하는 방법이 필요하다.

Flow는 특정 시간 동안에 소스와 데스티네이션 간에 전송되는 일련의 IP 패킷들을 말한다.[8] EBA 방식의 한계인 긴 계산 시간을 단축시키기 위해서 각각의 패킷을 고려하지 않고, 일정 시간동안 축적된 flow를 기준으로 계산을 진행 할 수 있다. 동시에 엔트로피의 계산 역시 flow를 기준으로 변화를 탐지한다. 이러한 방법이 Fast Entropy Approach[3]이다. 공격이 있을 시에는 한 flow가 압도적으로 지배할 것이기 때문에 엔트로피 값이 크게 감소할 것이고 이때를 탐지하는 방법이다.

많이 사용되고 있는 Flooding 공격을 막는 데에는 적당한 역치의 설정이 아주 중요하다. 만약 역치가 너무 높으면 사소한 공격은 감지를 못 할 것이고, 역치를 너무 낮게 설정하면 정상적인 서버의 과부화도 공격으로 취급될 것이다. 즉, 엔트로피 값에 따라서 유동적으로 변화하는 역치를 설정해야 한다. 변화하는 역치의 기준은 서버 상태를 분석한 이후 알고리듬화 해서 적용하게 된다.

5. 딥러닝의 활용 및 성능 확인 지표

본 장에서는 딥러닝의 학습에 사용되는 사이버 공격의 특징들과 그 성능을 평가할 수 있는 지표들을 알아본 후 마지막으로 기존 방식과 딥러닝의 성능 차이를 확인해 보겠다.

딥러닝 기술의 핵심은 많은 데이터양을 사람이 아닌 머신 스스로 특징을 찾아내서 학습한다는 점이다. 보통 8:2로 나뉜 학습 데이터와 테스트 데이터를 통해서 여러 가지 딥러닝 알고리듬을 실험해 본 후 사용한다. 기존 방식의 머신러닝은 데이터가 딥러닝보다 현저히 적다는 점에서 차이를 찾을 수 있다. 딥러닝이 적용된 탐지 엔진은 두 가지 모듈을 가지고 있다: Pre-trained 모듈, Online-Learning 모듈.[5] 전자는 미리 생성된 데이터로 학습을 하는 것이고, 후자는 모니터링을 통해 실시간 생성되는 데이터로 모듈을 업데이트하는 것이다. 이렇게 실시간으로 모듈을 업데이트하면 신생되는 공격도 막을 가능성이 높아진다.

DDoS 공격 역시 딥러닝을 활용해서 탐지하는 것이 최신 방어법이다. 그렇다면, DDoS 공격의 특징에 대해서 알아보자. 첫째, SSH Brute-force 공격의 경우, 한 번 SSH에 접속될 때 마다 Diffie-Hellman 키(D-H Key)[7]가 생성된다. 따라서 공격이 진행되는 동안에는 D-H Key의 발생이 압도적으로 많다. 둘째, DDoS DNS(Domain Name Server) 공격의 경우 Inbound / Outbound 비율의 변화가 중요 특징이다. DNS는 도메인 이름을 IP 주소로 바꿔주는 일을 하는데, 공격 중에는 수많은 IP 주소 변경 요청이 서버로 되돌아가도록 되기 때문에 그 비율이 현저히 떨어지게 된다. IP 스푸핑을 이용해서 요청이 되돌아가도록 만든다. 마지막으로, Transmission Control Protocol(TCP) 공격의 경우를 살펴보면, Synchronize Packet(SYN)과 Acknowledgement(ACK) 비율의 변화가 크게 나타난다. 정상 사용 중에는 SYN가 ACK에 비해 100배 정도 많이 나타난다. SYN는 세션이 만들어 질 때에만 나타나지만, ACK는 패킷을 보내는 것만으로도 생성되기 때문이다. 하지만, 공격이 진행되는 동안에는 수많은 세션을 생성하지만 클라이언트 쪽에서 아무런 활동을 하지 않기 때문에 SYN/ACK 비율이 비정상적으로 높다. 이런 특징 점들을 바탕으로 머신러닝 모델을 학습해서 공격을 탐지할 수 있다.

감시 시스템이 공격을 탐지해서 알람을 띄울 때 발생될 수 있는 경우의 수는 네 가지이다. 탐지 시스템이 올바르게 작동한 경우는 True Positive (TP) 이다. 이는 표 2에 정리되어 있다.

<표 2> 실제와 예상에 따른 알람

Actual / Prediction	Normal	Attack
Normal	True Negative (TN)	False Negative (FN)
Attack	False Positive (FP)	True Positive (TP)

이 네 가지 경우를 조합해서 감시 시스템의 성능을 판단할 수 있다. Accuracy는 시스템이 모든 경우에 대하여 올바르게 탐지 한 경우를 나타낸다. Detection Rate(DR)과 Recall의 경우는 공격이 얼마나 탐지되었는지를 나타내며, False Alarm Rate(FAR)는 상인 상태에서 공격이라고 잘못 탐지한 경우를 타나낸다. Precision은 모든 알람들 중 올바른 알람의 비율을 알려준다. 마지막으로, F1 지수는 Precision과 Recall의 조화평균을 나타낸다. F1 값이 높을록 더 좋은 알고리듬이다. 식은 아래 그림 1에 정리되어 있다.

$$Accuracy = \frac{TP+TN}{FP+FN+TP+TN} \qquad Precision = \frac{TP}{TP+FP}$$

$$Detection\ Rate = \frac{TP}{TP+FN} \qquad Recall = \frac{TP}{TP+FN}$$

$$False\ Alarm\ Rate = \frac{TP}{TN+FP} \qquad F1 = 2 * \frac{Precision * Recall}{Precision+Recall}$$

(그림 1) 탐지 성능 판단 지표들의 정의

마지막으로 기존에 쓰였던 방식과 인풋 데이터가 훨씬 늘어난 딥러닝의 성능 비교를 보겠다. 현재 쓰이고 있는 방법이란, 인풋 데이터양이 적은 머신러닝을 지칭한다. 표 3에서 확인 할 수 있듯이 Accuracy, DR, 그리고 FAR까지 모두 훨씬 나은 결과를 보여준다. 특별히 FAR이 대폭 향상되는데, 이는 알람이 뜰 때마다 확인해야하는 관리자의 입장에서는 가장 중요한 성능지표 중 하나이다.

<표 3> 기존 방식과 딥러닝의 성능 비교[4]

	Accuracy (%)	DR (%)	FAR (%)
기존 방식	95.22	97.50	6.57
딥러닝	99.20	99.27	0.85

6. 결론

본 논문에서 다양한 사이버 공격과 탐지 기법, 그 중에서도 DDoS 공격과 딥러닝을 이용한 탐지 법을 살펴보았다. 각각의 공격과 탐지 법은 표를 통해 확인할 수 있다. IoT가 보편화 되면서 모든 기기들이 인터넷을 통해 연결되어 있는 사회에서 DDoS 공격은 가장 위협적이다. 가장 많이 사용되는 공격 방법은 Volumetric 공격으로써 여러 PC에서 동시다발적으로 공격이 진행되어 근원을 파악하는 것이 상당히 어렵다. 이를 막는 방법으로는 EBA를 많이 사용하는데, 여전히 변종 공격에는 약한 모습을 보이고 있다. 그럼으로, 앞으로 각광 받을 방어 기법은 딥러닝을 활용해서 사람의 손을 거치지 않고 공격을 감지하고 막는 것이다. 조금씩 변형되어 가는 DDoS 공격 역시 탐지가 가능하기 때문이다. 딥러닝을 통해서 변형된 공격도 낮은 FAR으로 탐지하는 연구가 계속되어야 할 것이다.

7.. ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었으며 (IITP-2017-2015-0-00403), 2017년도 정부(미래창조과학부)의 재원으로 정보 통신 기술 진흥 센터의 지원을 받아 수행된 연구 (No.2016-0-00078, 맞춤형 보안 서비스 제공을위한클라우드기반지능형보안기술개발) 및 2017년도 두뇌 한국 21 플러스 사업에 의하여 지원되었음.

참고문헌

[1] Raiyn J. A survey of cyber attack detection strategies. International Journal of Security and Its Applications. 2014;8(1):247-56.

[2] Liu J, Lai Y, Zhang S. FL-GUARD: A Detection and Defense System for DDoS Attack in SDN. InProceedings of the 2017 International Conference on Cryptography, Security and Privacy 2017 Mar 17 (pp. 107-111). ACM.

[3] David J, Thomas C. DDoS attack detection using fast entropy approach on flow-based network traffic. Procedia Computer Science. 2015 Jan 1;50:30-6.

[4] Diro AA, Chilamkurti N. Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems. 2017 Sep 1.

[5] He Z, Zhang T, Lee RB. Machine Learning Based DDoS Attack Detection from Source Side in Cloud. InCyber Security and Cloud Computing (CSCloud), 2017 IEEE 4th International Conference on 2017 Jun 26 (pp. 114-120). IEEE.

[6] Worldwide infrastructure security report. Available: http://www.arbornetworks.com/images/documents/WISR2016_EN_Web.pdf

[7] Blake-Wilson S, Menezes A. Authenticated Diffie-Hellman key agreement protocols. InSelected Areas in Cryptography 1998 Aug 17 (Vol. 1556, pp. 339-361)

[8] David J, Thomas C. Intrusion Detection Using Flow-Based Analysis of Network Traffic. Advances in Networks and Communications. 2011:391-9.