KIPS Trans. Comp. and Comm. Sys.
Vol.11, No.4 pp.127~132
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2022.11.4.127

Classification of Service Types using Website Fingerprinting in Anonymous Encrypted Communication Networks

Dongyoung Koo†

ABSTRACT

An anonymous encrypted communication networks that make it difficult to identify the trace of a user's access by passing through several virtual computers and/or networks, such as Tor, provides user and data privacy in the process of Internet communications. However, when it comes to abuse for inappropriate purposes, such as sharing of illegal contents, arms trade, etc. through such anonymous encrypted communication networks, it is difficult to detect and take appropriate countermeasures. In this paper, by extending the website fingerprinting technique that can identify access to a specific site even in anonymous encrypted communication, a method for specifying and classifying service types of websites for not only well-known sites but also unknown sites is proposed. This approach can be used to identify hidden sites that can be used for malicious purposes.

Keywords : TOR Network, Machine Learning, Website Fingerprinting, Service Types, Classification

익명 암호통신 네트워크에서의 웹사이트 핑거프린팅을 활용한 서비스 유형 분류

구 동 영†

요 약

토르 (Tor, The Onion Router)와 같이 다수의 가상 컴퓨터 및 네트워크를 경유함으로써 이용자의 인터넷 접속에 대한 추적을 어렵게 하는 익명 암호통신 네트워크는 데이터 송수신 과정에서의 사용자 및 데이터 프라이버시 보호를 그 운영목적으로 하고 있다. 하지만 이러한 익명 암호통신 네트워크를 통한 불법 콘텐츠 공유 및 무기거래 등 부적절한 용도로의 악용 및 오용에 있어, 기존의 탐지 기법을 적용하거나 적절한 대응책을 마련하기에는 어려움이 따른다. 본 논문에서는 익명 암호통신에서도 특정 사이트에 대한 접근 정보를 높은 정확도로 유추할 수 있는 웹사이트 핑거프린팅 (website fingerprinting) 기법을 확장하여, 특정 사이트 뿐 아니라 알려지지 않은 사이트에 대해서도 서비스 유형을 특정하고 분류하는 방법을 강구함으로써 악의적 목적에 활용될 수 있는 은닉 사이트 또는 잠재적 불법 사이트에 대한 식별 방안을 제시한다.

키워드 : 토르 네트워크, 머신러닝, 웹사이트 핑거프린팅, 서비스 유형, 분류

1. 서 론

정보통신기술의 발달과 더불어 공개 채널을 사용하는 인터넷의 대중화로 검열 및 도청 등에 따른 사용자 및 데이터에 대한 프라이버시 침해 우려가 커지면서, 데이터 송신 전에 암호화를 수행하는 암호통신 기술을 적용하는 사례 또한 급격히 증가하고 있다[1]. 하지만 암호통신은 개인정보보호를 넘어 악의적 목적을 지닌 사용자에 의한 불법 콘텐츠 공유, 마약 및 무기류 거래, 개인정보 매매, 악성코드 배포 등에 악용될 가능성이 증가함에 따라 암호통신 트래픽에 대한 분석의 중요성 또한 꾸준히 증가하고 있다. Zscaler에 따르면 암호통신을 이용한 공격이 해마다 314%씩 증가하는 추세를 보이며[2], SonicWall에 따르면 암호통신을 이용한 위협은 전년대비 21% 증가한 것으로 보고되었다[3]. 이는 사용자 및 데이터 프라이버시 보존을 위한 암호통신을 수행하는 경우라 하더라도 악의적 목적으로 운영될 수 있는 사이트에 대한 특정과 해당 사이트에 접근하는 사용자 트래픽 분석의 필요성을 반증하는 것이라 할 것이다.

대표적 익명 암호통신 프로젝트인 토르(TOR: The Onion Routing)에서는 전 세계에 물리적으로 분산된 다수의 가상 컴퓨터 및 라우터를 주기적으로 변경하여 경유하면서 계층화된 암호 패킷을 복호화함으로써 네트워크 우회 및 익명화를 제공한다. 하지만 최근 소개된 웹사이트 핑거프린팅(website fingerprinting) 기법은 머신러닝 기법을 활용하여 사용자가 접속하는 웹사이트를 높은 확률로 식별할 수 있다. 하지만 학습된 특정 웹사이트에 대한 접근 및 학습되지 않은 사이트에

※ 본 연구는 한성대학교 교내학술연구비 지원과제임.
† 종신회원 : 한성대학교 전자정보공학과 조교수
Manuscript Received : January 14, 2022
Accepted : February 9, 2022
* Corresponding Author : Dongyoung Koo(dykoo@hansung.ac.kr)

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/ licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

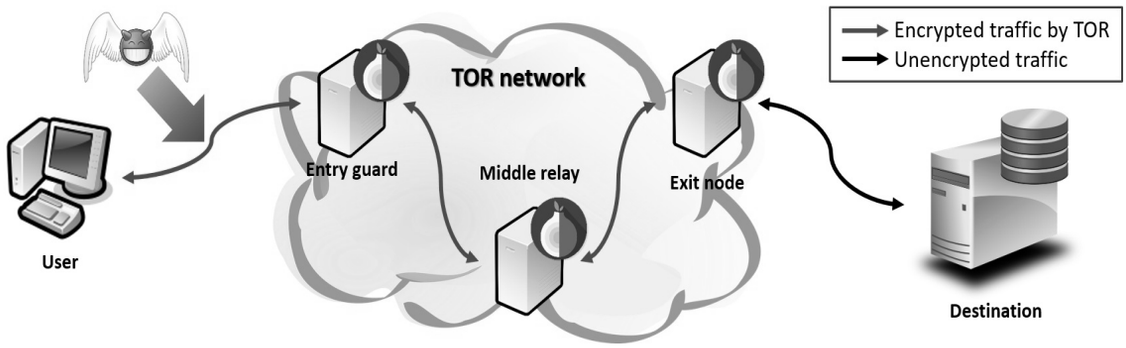

Fig. 1. Tor Network and Analysis Point

대한 접근 여부만을 판단하기 때문에 딥 웹(deep web) 등 대중에게 알려지지 않거나 사전에 데이터를 수집하기 어려운 웹 사이트에 대해서는 사용자가 접근하는 사이트가 어떤 서비스를 제공하는지 특정하기 어렵다는 제한이 있다. 따라서 본 논문에서는 특정 사이트에 대한 식별에서 나아가 웹사이트의 서비스 유형에 따른 특징을 분석하고 분류하는 방법을 모색함으로써 알려지지 않은 불법 또는 악용 가능성이 있는 사이트에 대한 분류를 수행한다.

2. 관련 연구

암호통신 네트워크 분석은 포트 기반 분석, 심층패킷분석, 트래픽 분석 등으로 분류될 수 있다. 포트 기반 분석은 동적 포트 할당에 따라 신뢰도가 떨어지며, 심층패킷분석(deep packet inspection, DPI)은 암호통신으로 인한 페이로드의 접근이 제한되어 높은 정확도를 확보하기 어렵다. 트래픽 분석은 통계적 특성 및 행위 정보를 입력으로 하는 머신러닝 기법을 활용한 분석 기법으로 전문가에 의한 특징 추출 및 각 데이터에 대한 레이블링(labeling)이 수작업으로 이루어지는 것에 기반하고 있어, 충분한 데이터 확보가 어렵다는 제약이 있다. 이에 수작업 기반 트래픽 분석에서의 효율성 저하를 극복하기 위한 암호통신 분석 연구가 활발히 이루어지고 있다.

2.1 익명 암호통신 네트워크

대표적인 익명 암호통신 네트워크인 토르(Tor)는 Fig. 1과 같이, 사용자가 궁극적으로 접속하고자 하는 웹사이트가 아닌 엔트리 노드(entry node)에 접속을 하면 토르 네트워크 내의 여러 중계 노드(relay node)를 거쳐 최종 노드(exit node)에서 사용자의 요청 및 응답을 대신 주고받기 때문에 사용자가 접근하는 웹사이트를 직접 확인하기 어렵게 한다.

2.2 딥러닝 활용 암호통신 트래픽 분석

딥러닝을 활용한 암호통신 트래픽 분석은 암호문에 대한 복호화 과정을 거치지 않은 상태에서 수집된 트래픽의 특징만을 학습함으로써 암호문을 분류하는 것으로, 정확도 향상을 위한 다수의 연구가 발표되었다. Chen et al.[4]은 수작업에 의한 특징 추출 및 오프라인 분석의 한계점 극복을 위하여 합성곱 신경망(CNN)을 적용하여 온라인 트래픽 분석이 가능한 프레임워크를 개발하였다. Wang et al.[5]은 분할 정보 기반의 머신러닝을 활용하는 경우 발생하는 국부 최적해 문제를 해결하기 위하여 1차원 CNN을 활용하여 암호통신에서 비선형 관계로 표현되는 사용자 행동을 특정하는 오프라인 분석 기법을 제시하였다. Lotfollahi et al.[6]은 머신러닝에서 수작업 의존에 따른 전반적 특징 추출의 한계점 개선을 위하여 CNN과 SAE (stacked auto-encoder)의 두 가지 딥러닝 기법을 활용하여 자동으로 VPN을 이용한 통신을 식별하는 'deep packet' 기법을 제시하였다. 딥러닝 학습을 위한 데이터셋의 불균형 문제를 해결하기 위한 연구도 다양하게 이루어지고 있는데, Vu et al.[7]은 AC-GAN (auxiliary classifier-GAN)을 활용하여 불균형 데이터셋을 합성한 후 SVM, 의사결정 트리, random forest를 통하여 학습하고 트래픽을 식별하는 방법을 제시하였다. Rezaei와 Liu [8]는 인위적 스크립트로 만들어져 레이블된 데이터가 실제 트래픽과 다른 분포를 가지는 제약을 극복하고자 샘플링된 시계열 특징을 이용한 준지도 CNN 기반 딥러닝 기법을 통하여 오프라인으로 암호통신을 분석하는 기법을 제시하였다. Aceto et al.[9]은 모바일 기기에서 일관된 데이터셋의 확보가 어려운 상황에서 패킷의 헤더와 페이로드로부터 CNN, LSTM, SAE, MLP를 활용한 딥러닝 학습을 통하여 온라인 분석을 수행하고 사용된 애플리케이션 식별을 위한 연구를 수행하였다.

2.3 웹사이트 핑거프린팅

웹사이트 핑거프린팅은 익명화된 사용자의 활동 식별을 목적으로 트래픽 분석을 통하여 개별 세션에 대한 익명 암호통신 네트워크에서 사용자가 접근하는 웹사이트를 추정하는 기법이다. 웹사이트 트래픽을 수집하여 패킷 크기, 패킷 방향,

도착 간격 시간 등의 특징 정보를 이용하여 머신러닝 기법을 활용한 훈련 과정을 거치는데, 분류기의 성능 발전에 따라 식별 성공률이 높아지고 있다.

본 연구에서는 사용자가 접근하고자 하는 최종 목적지 웹사이트의 서비스 유형을 알 수 없는 상황에서도 수집된 트래픽 데이터셋으로부터 특정 웹사이트를 접근하는 트래픽이 제공하는 서비스의 유형을 분류하는 것을 목적으로 한다.

3. 제안 방법

웹사이트 핑거프린팅은 토르와 같은 익명 암호통신 네트워크를 이용하는 사용자가 접근하는 특정 웹사이트를 유추하지만, 제안 기법은 사용자가 접근하는 웹사이트의 서비스 유형을 파악하는 것을 그 목적으로 한다. 따라서, 동일 서비스 유형에 속하는 웹사이트 사이에서의 공통점을 머신러닝 기법을 이용하여 학습하기 때문에, 기존 웹사이트 핑거프린팅에서와 같이 동일 서비스 유형에 속하는 서로 다른 웹사이트의 특징들은 학습 모델에서 제외될 필요가 있다.

3.1 데이터셋 수집

Alexa 상위 100개 사이트 중에서 4가지(지도, 뉴스, 쇼핑, 스트리밍) 대표 서비스 유형에 대한 분류를 시도한다. Ubuntu 20.04 LTS 64bit에서 동작하는 3개의 데스크톱을 성북구(2대)와 노원구(1대)에 배치하여 각 장치가 Tor 엔트리 노드와 통신하는 트래픽을 수집하였으며, 지도 서비스 등과 같이 이미지 정보의 송수신이 다량 발생하는 실제 트래픽 환경을 고려하여 Chromium에서 torsocks를 활용하여 패킷을 수집하였다. 웹사이트마다 30초 동안 접속한 패킷을 100회 반복 수집하여 총 9,800개 트래픽 데이터셋을 학습하였다. 접속된 트래픽 정보는 오프라인 분석을 위하여 .pcap 확장자를 가지도록 TCPdump를 이용하여 파일로 저장하였으며, Tor 네트워크의 출구 노드는 20분 간격으로 경로를 재설정하는 기본 설정을 유지하며, 다양한 사용환경을 고려하여 각 웹사이트에 대한 접근을 라운드-로빈(round-robin) 방식으로 수집하였다. 웹사이트가 제공하는 서비스 유형 학습을 위하여 학습 단계에서는 웹사이트 단위와 웹사이트가 속하는 서비스 유형에 따른 레이블링을 별도로 적용하여 학습하였으며, 웹사이트의 서비스 유형은 Table 1과 같다.

3.2 전처리

본 연구에서는 [10,11]에서 활용한 트래픽 내에서 각 패킷의 방향성만 고려한 1차원 특징 벡터를 활용하였으며, Sirinam et al.[11]의 연구에서 논의되었던 성능향상을 위한 패킷 방향과 함께 패킷 크기까지 고려할 수 있도록 머신러닝 학습을 위한 고정 길이 입력으로 트래픽마다 5,000개의 패킷까지만 고려하였다.

3.3 학습 데이터셋 레이블링

학습 및 검증, 테스트 과정에서 기존 연구와 제안 기법의 성능 분석을 위하여 아래와 같이 각 사이트에 대한 레이블링을 수행하였다.

- 웹사이트 각각에 대한 레이블
- 4개 서비스 유형에 대한 레이블
- 특정 서비스(지도/뉴스/쇼핑/스트리밍) 해당 여부에 대한 이진 레이블

3.4 학습 결과

1) 웹사이트별 트래픽 패턴

기존 연구에서의 동일 웹사이트에서의 트래픽은 유사한 패턴을 보임을 확인할 수 있는데, 날씨 정보를 제공하는 darksky.net 웹사이트에 대하여 서로 다른 시점에 웹사이트 접속 트래픽을 수집하여 시각화한 결과는 Fig. 2와 같다. x축은 시간의 경과를 나타내며, y축은 해당 시점에 송수신되는 데이터의 양을 의미하며, 사이한 시점에 수집한 트래픽이라

Table 1. Websites from Alexa top 100 According to Service Types

Service Type (# of websites)	Websites
Map (6)	www.openstreetmap.org, www.google.com/maps, www.mapquest.com, www.bing.com/maps, thetruesize.com, map.kakao.com
News (4)	edition.cnn.com, www.nyutimes.com, www.bbc.com, www.tribunnews.com
Shopping (14)	www.amazon.de, www.rakuten.co.jp, www.amazon.co.uk, www.ebay.com, www.aliexpress.com, www.flipkart.com, www.amazon.co.jp, www.1688.com, www.etsy.com, www.tmall.com, world.taobao.com, www.amazon.com, global.jd.com, www.amazon.in
Streaming (9)	www.panda.tv, www.twitch.tv, www.livejasmin.com, www.aparat.com, www.yy.com, www.youtube.com, www.netflix.com, www.bongacams.com, www.zhanqi.tv

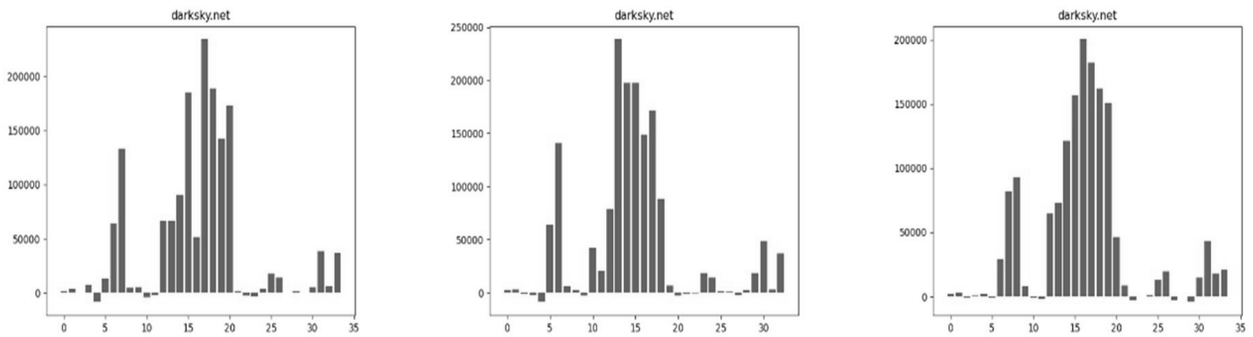

Fig. 2. Traffic Pattern of Darksky.Net (Weather Service)

하더라도 동일 웹사이트 접근에서는 유사한 트래픽 패턴을 보이는 것을 확인할 수 있다.

2) Open World에서의 웹사이트/서비스 분류

Open World (OW) 분류는 학습 데이터에 사용되지 않은 웹사이트를 입력으로 하여, 분류 정확도를 파악하는 것으로, 전세계 모든 웹사이트에 대한 데이터셋을 충분히 확보하지 못하는 상황에서 알려지지 않은 웹사이트의 식별 및 해당 웹사이트가 제공하는 서비스 유형을 유추하기 위한 방법이다. 일반적인 웹사이트 핑거프린팅에서는 주요 관심 웹사이트를 선정하여 특정 웹사이트의 접근 여부만을 판별하는 이진 분류를 적용하는 반면, 본 연구에서는 학습에 사용되지 않은 웹사이트라 하더라도 서비스 유형에 따라 학습된 모델을 이용하여 제공되는 서비스 유형을 유추하는 것을 목표로 한다.

Fig. 3은 패킷 방향만을 고려한 경우로, 서비스별 분류 정확도는 0.68250 (29 epoch)으로 가장 높게 나타났다. epoch은 모든 학습 데이터셋에 대하여 완료 학습 사이클의 수를 의미하며, 전체 데이터셋을 사용하여 몇 번의 반복 학습을 거쳤는지를 epoch의 수로 표현하였다. 나머지는 특정 서비스 유형별 이진 분류를 수행할 때의 결과로 여러 유형을 한 번에 분류하는 것보다 상대적으로 높은 정확도를 보인다. 지도 서비스는 0.77250 (28 epoch), 뉴스 서비스는 0.94500 (27 epoch), 쇼핑 서비스는 0.6500 (25~27 epoch), 스트리밍 서비스는 0.74250 (27 epoch)의 정확도를 보였다.

Fig. 4는 패킷 방향과 패킷 크기를 함께 고려한 분류 정확도로, 학습에 사용한 웹사이트 중에서 임의 선택한 웹사이트에 접근하는 새로운 트래픽을 분류하는 Closed World (CW)와는 달리 패킷 방향만 고려한 경우와 유사하거나 낮은 정확도를 보였다. 4개 서비스의 유형 분류는 29 epoch에서 0.68250, 서비스별 이진 분류에서는 지도 서비스가 29 epoch에서 0.79000, 뉴스는 28 epoch에서 0.92250, 쇼핑은 30 epoch에서 0.99000, 스트리밍은 28 epoch에서 0.75750의 정확도를 보였다.

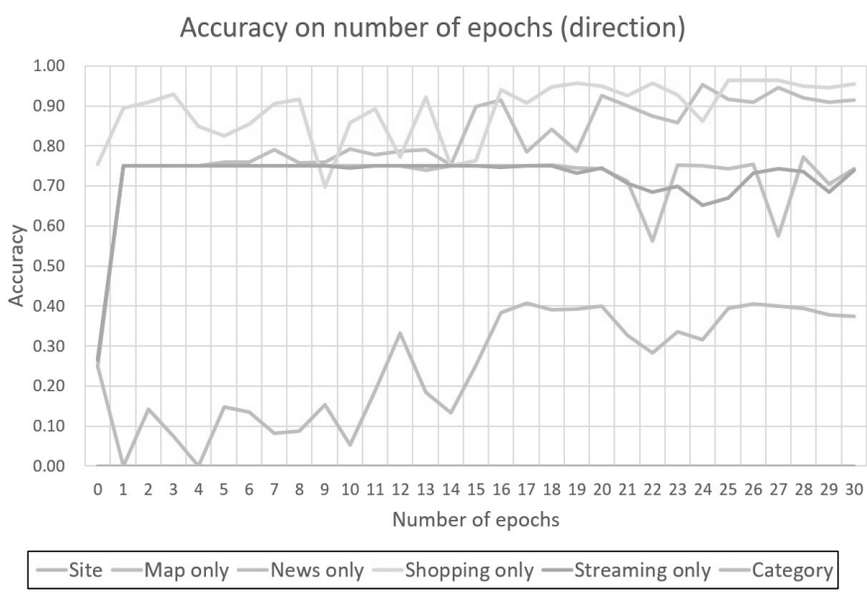

Fig. 3. Classification Accuracy for Specific Service Types using Packet Directions Only (OW)

Table 2. Classification Accuracy According to Combination Ratio

비율	0:1	8:1	4:1	2:1	1:1	1:2	1:4	1:8	1:0
정확도	0.7475	0.7775	0.7800	0.7650	0.7525	0.7450	0.7200	0.7050	0.2775

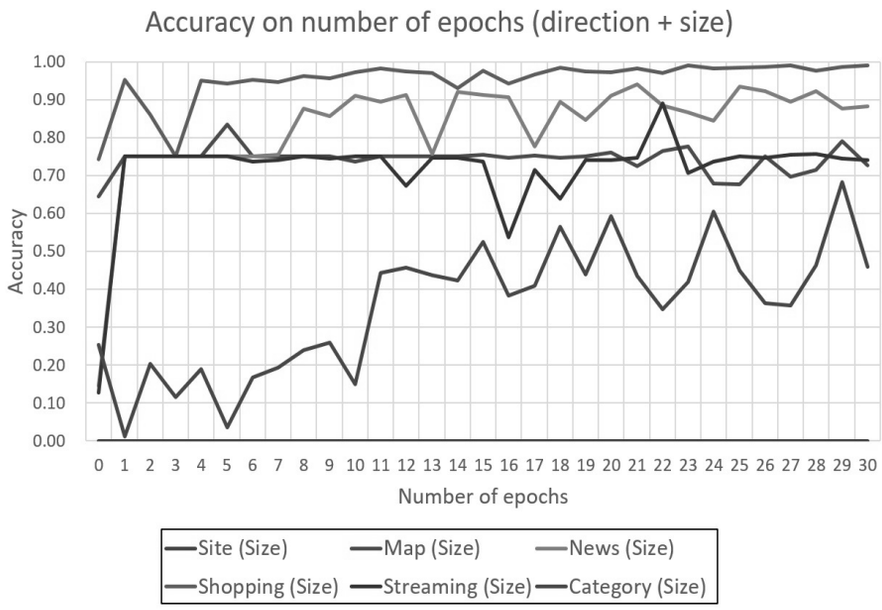

Fig. 4. Classification Accuracy for Specific Service Types using both Packet Directions and Size (CW)

3) Open World에서의 서비스 분류 정확도 개선

Open World (OW)에 대한 분류 성능은 Closed World (CW)에 비하여 일반적으로 낮은 정확도를 나타내는데, Sirinam et al.[11]에서는 Open World에서도 90.7%의 높은 정확도를 보이는 것에 비하여 본 실험에서는 37.5~46.0%의 낮은 정확도로 선행 연구 결과와 큰 차이를 보인다. 이는 [11]에서 사용한 데이터셋의 10%로 학습을 수행한 데이터셋의 부족에 따른 것과 동일 서비스를 제공하는 웹사이트의 편차에 따른 것으로 볼 수 있다.

서비스 유형별 예측 정확도를 높이기 위하여 서비스 유형별 이진 분류기와 4개 서비스 유형에 대한 분류의 예측 확률을 조합하여 예측 정확도를 높이고자 하였다. 서비스 유형별 이진 분류에서 가장 높은 확률과 가장 낮은 확률은 전체 예측에 큰 영향을 미치므로 배제하고 전체 서비스 유형 분류기와 서비스별 이진 분류기에서의 예측 확률의 비율 (α:β)을 달리하여 정확도를 Table 2와 같이 계산하였다. 서비스 유형 분류와 서비스별 이진 분류기의 반영 비율을 4:1로 설정한 경우, 분류 정확도가 최대 78%까지 높아지는 것을 확인하였다.

4. 결 론

본 연구에서는 암호통신 네트워크에서 사전에 학습된 특정 웹사이트의 접속여부를 판단하는 웹사이트 핑거프린팅을 확장하여, 암호통신을 이용하는 사용자가 접근하는 웹사이트에서 제공하는 서비스 유형을 분류하는 방법을 모색하였다. 실험을 통하여 서비스 유형과 각 서비스 이진 분류기를 조합함으로써 Open World에서도 최대 78%의 정확도로 서비스 유형을 분류할 수 있음을 확인하였다. 향후 패킷의 다양한 특성 등의 조합을 통하여 특정 웹사이트 뿐 아니라 알려지지 않은 웹사이트에서 제공하는 서비스 유형에 따른 분류 정확도를 높일 수 있는 기법에 대한 기초가 될 것으로 생각하며, 향후 연구에서는 다양한 앙상블 기법을 적용하여보다 많은 서비스 유형에 대한 암호통신에서의 서비스 유형 분류 연구를 수행할 계획이다.

References

[1] N. Shah, “The challenges of inspecting encrypted network traffic,” Fortinet [Internet], https://www.fortinet.com/blog/industry-trends/keeping-up-with-performance-demands-of-encrypted-web-traffic. 2020.08.04.

[2] N. Wodecki, “Zscaler’s 2021 encrypted attacks report reveals 314 percent spike in HTTPS threats,” Zscaler [Internet], https://www.zscaler.com/press/zscalers-2021-encrypted-attacks-report-reveals-314-percent-spike-https-threats. 2021.08.28.

[3] SonicWall, "2021 Cyber Threat Report: Mid-Year Update," SonicWall [Intenet], https://webobjects2.cdw.com/is/content/CDW/cdw/on-domain-cdw/brands/sonicwall/mid-year-2021-cyber-threat-report.pdf

[4] Z. Chen, K. He, J. Li, and Y. Geng, "Seq2Img: Sequence-to-Image based approach towards IP traffic classification using convolutional neural networks," *IEEE International Conference on Big Data (BIGDATA)*, pp.1271-1276, 2017.

[5] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-End encrypted traffic classification with one-dimensional convolution neural networks," *IEEE Internationa Conference on Intelligence and Security Informatics (ISI)*, pp.43-48, 2017.

[6] M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Soft Computing*, Vol.24, No.3, pp.1999-2012, 2020.

[7] L. Vu, C. T. Bui, and Q. U. Nguyen, "A deep learning based method for handling imblanced problem in network traffic classification," *International Symposium on Information and Communication Technology (SoICT)*, pp.333-339, 2017.

[8] S. Rezaei and X. Liu, "How to achieve high classification accuracy with just a few labels: A semi-supervised approach using sampled packets," *arXiv e-prints, arXiv-1812*, 2019.

[9] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Mobile encrypted traffic classification using deep learning," *Network Traffic Measurement and Analysis Conference (TMA)*, pp.1-8, 2018.

[10] T. Wang and I. Goldberg, "Improved website fingerprinting on tor," *ACM Workshop on Privacy in the Electronic Society (WPES)*, pp.201-212, 2013.

[11] P. Sirinam, M. Imani, M. Juarez, and M. Wright, "Deep fingerprinting: Undermining website fingerprinting defenses with deep learning," *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp.1928-1943, 2018.

구 동 영

https://orcid.org/0000-0003-3283-5494
e-mail : dykoo@hansung.ac.kr
2009년 연세대학교 컴퓨터.산업공학(공학사)
2012년 한국과학기술원 전산학(공학석사)
2016년 한국과학기술원 전산학(공학박사)
2016년 ~ 2017년 고려대학교 컴퓨터학과 연구교수
2017년 ~ 현 재 한성대학교 전자정보공학과 조교수
관심분야 : Information Security, Applied Cryptography, Network Security, Cloud/Fog/Edge Computing Security