JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 30 No. 12, pp. 195-204, December 2025
https://doi.org/10.9708/jksci.2025.30.12.195

Machine Learning-based Application-Layer Network Fingerprinting for Dark Web Criminal Service Classification

Jinwoo Shin*, Dong-Won Kang**, Jiyeon Kim***

*Undergraduate Student, Dept. of Cyber Security, Daegu University, Gyeongsan, Korea
**Undergraduate Student, Dept. of Computer Engineering, Daegu University, Gyeongsan, Korea
***Professor, School of Computer and Information Engineering, Daegu University, Gyeongsan, Korea

[Abstract]

As drug and hacking crimes exploiting the anonymity of the Dark Web have rapidly increased, automated technologies capable of detecting them quickly are needed. Previous approaches such as HTML-based static content analysis are vulnerable to content concealment and dynamic changes, and transport layer (L3/4) traffic analysis is limited to network transmission characteristics such as packet size and transmission rate, failing to reflect the structural and content characteristics of websites. To overcome these limitations, this paper collects not only L3/4 metrics but also application layer (L7) metrics, and analyzes effective metrics for classifying service types of Dark Web crime sites through machine learning-based learning. To this end, approximately 12,847 Dark Web sites related to drugs and hacking were collected, and a dataset was constructed by extracting 46 network metrics of L3/4 and L7 generated during site access. Furthermore, machine learning algorithms such as XGBoost, Random Forest, Logistic Regression, and SVM were developed as single models learning metrics of each layer and hybrid models integrating metrics of both layers. As a result, the XGBoost model trained only on L7 metrics recorded the highest performance with F1-scores of 0.996 for drug crime site type classification and 0.933 for hacking crime site type classification.

▸ **Key words**: Dark Web, Cyber Crime, Digital Investigation, Fingerprint, Machine Learning

• First Author: Jinwoo Shin, Corresponding Author: Jiyeon Kim
*Jinwoo Shin (jinu@daegu.ac.kr), Dept. of Cyber Security, Daegu University
**Dong-Won Kang (ehfejddl3979@daegu.ac.kr), Dept. of Computer Engineering, Daegu University
***Jiyeon Kim (jyk@daegu.ac.kr), School of Computer and Information Engineering, Daegu University
• Received: 2025. 11. 14, Revised: 2025. 12. 09, Accepted: 2025. 12. 12.
• This paper is an extended version of the paper presented at the 72nd Summer Conference of the Korea Society of Computer and Information (2025) titled "Extracting Fingerprints of Dark Web Drug Sites via Machine Learning on Multi-Layer Network Traffic."

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

[요 약]

다크웹의 익명성을 악용한 마약 및 해킹 범죄가 급증함에 따라, 이를 신속하게 탐지할 수 있는 자동화된 기술이 필요하다. 기존의 방식인 HTML 기반 정적 콘텐츠 분석은 콘텐츠 은폐 및 동적 변화에 취약하고, 전송 계층(L3/4) 트래픽 분석은 패킷 크기, 전송률 등 네트워크 전송 특성에 국한되어 웹사이트의 구조 및 콘텐츠 특성을 반영하지 못하는 한계가 존재한다. 본 논문에서는 이러한 한계를 극복하기 위해 L3/4 메트릭뿐만 아니라, 애플리케이션 계층(L7) 메트릭을 추가로 수집하고, 머신러닝 기반으로 학습하여 다크웹 범죄 사이트의 서비스 유형 분류에 효과적인 메트릭을 분석한다. 이를 위해 마약 및 해킹 관련 약 12,847개의 다크웹 사이트를 수집하여, 사이트 접속 시 발생하는 L3/4 및 L7의 46개 네트워크 메트릭을 추출하여 데이터셋을 구축하였다. 또한, XGBoost, Random Forest, Logistic Regression, SVM과 같은 머신러닝 알고리즘을 각 계층별 메트릭을 학습하는 단일 모델과 두 계층의 메트릭을 통합 학습한 하이브리드 모델로 개발한 결과, L7 메트릭만 학습한 XGBoost 모델이 마약 범죄 사이트 유형 분류에서 F1-score 0.996, 해킹 범죄 사이트 유형 분류에서 F1-score 0.933을 기록하며 가장 높은 성능을 보이는 것을 확인하였다.

▸ **주제어**: 다크웹, 사이버 범죄, 디지털 수사, 핑거프린트, 머신러닝

I. Introduction

다크웹(Dark Web)은 일반적인 표면 웹(Surface Web)과 달리, Tor(The Onion Router)와 같은 특수 브라우저와 암호화된 통신 프로토콜을 통해 정보 교환이 이루어지는 익명화 네트워크이다. 이러한 다크웹의 익명성은 프라이버시 보호와 표현의 자유라는 긍정적 기능 외에도 마약 거래, 해킹, 불법 무기 거래, 아동 음란물 유통 등 다양한 형태의 사이버 범죄에 악용되고 있다. 특히, 마약과 해킹은 다크웹 범죄 중 가장 높은 비중을 차지하며 점점 더 정교하고 조직적인 형태로 발전하고 있다.

다크웹 범죄 사이트는 판매나 거래를 위한 'Marketplace'뿐 아니라 정보 공유를 위한 'Community', 다른 사이트를 중개하고 특정 서비스를 연결해주는 'Broker' 등의 형태로 운영된다. 이러한 사이트들은 수사 회피를 위해 임시 폐쇄하거나 주소를 변경하는 등의 방법으로 추적을 회피하고 있어 실제 수사 환경에서 이들을 정확하게 식별하고, 수사 효율성을 높이기 위해 임시 폐쇄된 범죄 사이트까지 탐지할 필요가 있다.

기존의 다크웹 분석은 HTML 기반 정적 콘텐츠 분석이나 사이트 구조 분석 위주로 수행되어 왔으나, 이는 사이트 구조의 유사성, 콘텐츠의 은폐성, 동적 변화 등으로 인해 신속하고 정확한 탐지가 어렵다는 한계가 존재한다. 또한, OSI 7계층 중 전통적인 네트워크 및 전송 계층(L3/4) 트래픽 분석은 패킷 크기, 전송률, 프로토콜 정보 등 네트워크 전송 특성에 국한되어 웹사이트의 구조 및 콘텐츠 특성을 반영하지 못한다.

이에 본 연구에서는 L3/4 메트릭뿐 아니라, 애플리케이션 계층(L7)의 메트릭을 기반으로 다계층 네트워크 핑거프린트를 추출하고, 이를 머신러닝 기반으로 학습하여 다크웹 범죄 사이트의 서비스 유형을 자동 분류하는 모델을 제안한다. 특히, 기존 연구들이 주로 L3/4 메트릭에 국한되었던 것과 달리, 본 연구는 L7 메트릭까지 수집하여 다크웹 범죄 사이트의 서비스 유형 분류에 효과적인 메트릭을 식별하여 수사 효율을 높이는 데 의의가 있다.

본 논문의 구성은 다음과 같다. 2장에서는 다크웹 범죄 콘텐츠 수집 기법과 분석에 관한 기존 연구를 살펴본다. 3장에서는 마약 및 해킹 범죄 관련 다크웹 사이트의 유형별 다계층 메트릭 수집 방법과 머신러닝 기반 네트워크 핑거프린트 학습 모델을 설계한다. 4장에서는 제안한 모델의 실험 결과 및 성능 분석을 수행하며, 5장에서 결론 및 향후 연구 방향을 제시한다.

II. Related Works

본 장에서는 다크웹 범죄 콘텐츠를 효율적으로 수집하기 위한 크롤링 기법과 수집된 다크웹 데이터를 분석한 기존 연구들을 살펴본다.

1. Collection Techniques for Criminal Dark Web Content

다크웹은 일반적인 검색 엔진으로 접근할 수 없는 웹 공간으로, Tor와 같은 익명 네트워크를 통해서만 접속이 가능하다. 다크웹 내에서 발생하는 범죄를 추적하고 증거를 수집하기 위해 HTTP 통신 기반 웹페이지 구조 및 데이터를 자동으로 수집하는 크롤링(Crawling) 기술 연구가 다양하게 수행되었다.

다크웹 콘텐츠 수집을 위한 기존 연구로는 로그인 세션 쿠키 및 CAPTCHA(자동 입력 방지 문자)를 동시에 처리하는 크롤러를 개발한 연구[1], 다크웹 포럼 수집에 특화된 크롤러를 개발한 연구[2], 표면 웹, 소셜 미디어 웹, 다크웹을 수집한 후 머신러닝을 적용하여 범죄 관련성이 높은 다크웹 콘텐츠를 효율적으로 수집하는 연구[3]가 있다. 또한, 자연어 처리 및 패턴 인식 기법을 적용하여 키워드와 링크를 기반으로 다크웹 내 보안 위험 데이터를 효율적으로 수집하는 연구[4], 학습 가능하고 재사용 가능하며 확장 가능한 자동화된 크롤링 방법 및 도구를 개발한 연구[5]가 수행되었다.

기존 다크웹 크롤링 연구들[1-5]은 주로 CAPTCHA, 로그인 인증 등 기술적 장애물을 극복하거나 수집 효율성을 개선하는 데 초점을 두었으며 다크웹 페이지 상의 텍스트 또는 HTML Tag를 수집하였다. 이에 비해, 본 연구는 크롤링을 통해 수집된 사이트 접속 시 발생하는 네트워크 트래픽 메트릭을 수집하고, 이를 머신러닝 기반으로 학습하여 다크웹 범죄 사이트의 서비스 유형을 자동 분류하는 핑거프린트를 개발한다는 점에서 차별점이 있다.

2. Analysis of Criminal Dark Web Data

다크웹 데이터 분석 연구는 크게 콘텐츠 기반 분석과 네트워크 트래픽 기반 분석으로 구분할 수 있다.

먼저, 콘텐츠 기반 분석 연구로는 다크웹 데이터의 구조적 패턴을 분석하고 SVM(Support Vector Machine)을 사용하여 콘텐츠 유형을 분류하는 연구[6], 다크웹 DUTA 데이터셋을 활용하여 LLDA(Labeled Latent Dirichlet Allocation) 기반 키워드 가중치를 사용한 TextCNN으로 범죄 다크웹 텍스트를 분류하는 연구[7]가 있다. 이와 유사하게 HTML 문서에서 특정 태그들을 대상으로 TF-IDF(Term Frequency-Inverse Document Frequency)로 임베딩하여 RNN(Recurrent Neural Network)으로 다중 라벨 분류를 시도한 연구[8]가 수행되었으며, 다크웹 콘텐츠에 SVC(Support Vector Classifier), NB(Naive Bayes)를 적용하여 마약, 해킹, 위조 신분증 등으로 분류하는 연구[9]도 진행되었다. 또한, 다크웹 페이지의 텍스트 데이터를 기반으로 NLP(Natural Language Processing)를 이용하여 범죄 콘텐츠를 자동으로 분류하는 연구[10], 다크웹에서 수집된 데이터셋을 활용하여 딥러닝 기반 유형 분류 모델(CNN, LSTM)을 구축하고 범죄와 비범죄 활동을 자동으로 식별하는 연구[11]가 수행되었다. 다크웹과 표면 웹 간 유사 사용자를 식별하기 위해 BERTopic과 저자 속성 분석을 활용한 연구[12], 다크웹 콘텐츠를 자동으로 크롤링한 뒤 수집된 데이터를 SVM, Decision Tree 등의 데이터 마이닝 기법으로 범죄를 분류하는 연구[13], 다크웹 사이트의 HTML 구조 패턴을 분석하여 사이트 유형별로 클러스터링하는 연구[14]도 수행되었다.

네트워크 트래픽 기반 분석 연구로는 약 6,000만 개의 대규모 다크웹 페이지를 수집하여 도메인, 웹그래프 구조, 암호화폐 활동 등을 분석하여 다크웹 서비스 운영 및 거래 패턴을 분석한 연구[15], 다크웹 네트워크 트래픽과 암호화폐 거래 내역을 연계하여 다크웹 활동과 암호화폐 거래 간의 관계를 분석한 연구[16]가 있다. 다크웹 익명망(Tor, I2P, JonDonym 등)에서 수집한 트래픽을 플로우 단위(패킷 크기, 간격)로 특징을 추출한 후 머신러닝을 통해 익명망을 분류하는 연구[17], 다크웹 네트워크 L3/4의 트래픽을 수집하여 공간-시간 융합 및 어텐션 메커니즘(Attention Mechanism)을 결합한 딥러닝 모델을 학습하여 마약 범죄, 무기 거래, 불법 포럼, 금융 사기 등의 범죄 유형을 분류하는 연구[18]가 수행되었다. 또한, 공개 데이터셋(CIC-Darknet2020)을 활용하여 네트워크 L3/4 트래픽의 바이트열을 RGB 이미지로 변환하고 시간 차원을 추가한 3D-CNN 모델을 학습시켜 다크웹 트래픽을 오디오, 이메일, 브라우징, 채팅, P2P(Peer-to-Peer), 파일 업다운로드, 비디오, VoIP(Voice over Internet Protocol) 등으로 분류하는 연구[19], 공개 데이터셋(CICDarknet2020)을 멀티 모달 기반으로 학습하는 연구[20]가 수행되었다.

기존 다크웹 콘텐츠 분석 연구들[6-14]은 주로 다크웹 페이지를 수집하여 텍스트, HTML 구조를 통해 범죄 유형 및 서비스 유형을 분석하였다. 반면, 본 연구는 페이지 콘텐츠가 아닌 다크웹 사이트 접속 시 발생하는 네트워크 트래픽 메트릭을 활용함으로써 직접 페이지에 접속하지 않고도 신속하게 사이트 유형을 분류할 수 있다. 또한, 다크웹 네트워크 트래픽 분석 연구들[15-20]은 주로 공개 데이터셋을 활용하여 L3/4 메트릭 학습이나 사용자 행위 분류에 집중한 반면, 본 연구는 직접 범죄 다크웹에 접속하여 L3/4와 L7 메트릭을 수집하고, 각 계층별 메트릭의 유효성을 실험적으로 검증하여 범죄 다크웹 유형 분석에 가장 효과적인 메트릭 조합을 도출한다는 점에서 차별점을 갖는다. Table 1은 기존 연구와의 차별성을 비교 분석한 표이다.

Table 1. Comparative Summary of Existing Studies and our Proposed Model

Reference	Collect Own Datasets	Classify Criminal Service Types	Use Traffic Metrics	Use L3/4 and L7 Metrics
[7, 11]	-	-	-	-
[1-6, 8-10, 12-13]	✔	-	-	-
[14]	✔	✔	-	-
[15-18]	✔	-	✔	-
[19-20]	-	-	✔	-
Our model	✔	✔	✔	✔

III. Traffic Collection and Network Fingerprinting of Criminal Dark Web

1. Multi-layer Metric Collection from Criminal Dark Web

본 연구에서는 마약 및 해킹 범죄 분야의 전문 자료에 기반하여 키워드를 선정하고, 체계적인 데이터 수집 과정을 통해 관련 다크웹 사이트를 확보하였다. 마약 관련 사이트 수집을 위해서는 미국 마약단속국 발간 속어 자료[21]를 기반으로 표준 키워드 'Drug'과 주요 마약류 6종 및 은어 1종을 선정하였다. 해킹 관련 사이트 수집에는 MITRE ATT&CK 프레임워크의 공격 기술 정의[22]를 참조하여 'exploit', 'RCE', 'rootkit' 등의 키워드를 Table 2와 같이 선별하였다.

Table 2. Dark Web Crime Search Keyword

Crime	Keyword	Description
Drug	drug	Substance affectingbody or mind
	LSD	Hallucinogenic drug
	heroin	Opioid narcotic
	cocaine	Central nervous stimulant
	marijuana	Psychoactive cannabis
	weed	Slang for marijuana
	Opium	Narcotic from poppy plant
	Ketamine	Dissociative anesthetic
Hacking	zeroday	Unknown vulnerability
	0day	Unknown vulnerability
	ddos	Distributed denial-of-service attack
	rootkit	Privileged access malware
	rce	Remote code execution exploit
	credentials	Authentication information
	exploit	Code abusing vulnerability
	malware	Malicious software
	ransomware	File-encrypting malware

데이터 수집은 자체 개발한 다크웹 사이트 수집 프로그램을 활용하여 Torch와 Ahmia 검색엔진에서 실시하였으며 약 3일간의 수집 과정을 통해 마약 관련 URL 6,860개, 해킹 관련 URL 5,987개를 확보하였다. 수집 과정에서 미러(Mirror) 사이트 및 하위 URL과 같은 중복 사이트는 별도로 제거하지 않고 그대로 유지하였다.

이후, 수집된 모든 URL에 직접 접속을 시도하여 'Connection Error'가 발생한 마약 사이트 1,226개, 해킹 1,060개를 식별하였다. 이러한 사이트는 운영 중단, 수사 회피를 위한 임시 폐쇄, 서버 이전 등으로 접속이 불가능하나 검색엔진 색인에는 남아있는 것으로 해석된다.

서비스 유형은 Table 3과 같이, 마약과 해킹 의뢰 및 판매 등 거래가 이루어지는 'Marketplace', 관련 정보를 공유하고 소통하는 'Community', 여러 사이트를 모아 중개하는 'Broker', 범죄와 관련 없는 'Normal'로 구분하였다.

Table 3. Description of Dark Web Service Types

Service Type		Description
Marketplace		Site where operators sell or provide platforms for direct drug transactions
Community		Site for sharing information about drugs, manufacturing, or usage methods among users
Broker		Site promoting or mediating marketplaces and communities
HTTP Status Code	404	HTTP Page Not Found
	408	Request Timeout
	500	Internal Server Error
	503	Server Overload or Maintenance
Normal		Dark web site unrelated to crime activity

본 논문에서는 이와 같이 접속이 불가한 사이트를 제외하고, 접속이 성공한 모든 사이트에 대해 전수 조사를 실시하여 HTML 기반 콘텐츠 특성을 분석하였다. 이를 통해 'Marketplace'는 상품명과 가격, 주문을 위한 Form 및 Button 태그가 있는 점, 'Community'는 댓글을 위한 Form 태그 및 여러 사용자명이 있는 점, 'Broker'은 서로 다른 Domain의 URL을 표시하는 점 등 서비스 별 고유 특성을 식별하여 분류하였다.

다크웹 범죄 사이트는 추적을 피하기 위해 onion 주소를 자주 변경하거나 서버를 일시적으로 중단하는 방식으로 운영된다. 특히, 다크웹 서비스 유형별로 추적 우회 방법과 운영 패턴이 상이할 수 있어, HTTP Status Code 분포 특성이 서비스 유형 구분에도 유용한 지표가 될 수 있다. 본 연구는 이러한 패턴을 반영하기 위해 HTTP Status Code를 별도의 분류 기준으로 포함하고, 각 코드

에 따른 트래픽 핑거프린트를 추출하였다.

마약과 해킹 사이트의 상세한 서비스 유형별 분포는 Table 4와 같다. 다크웹 접속이 성공한 사이트 중, 마약과 해킹 모두 'Marketplace'가 각각 5,330개(95.92%), 3,444개(69.9%)로 가장 높은 비율을 차지하였는데 이는 다크웹이 주로 불법 거래의 플랫폼으로 활용되고 있음을 보여준다.

Table 4. Distribution of Dark Web Service Types by Crime Category

Service Type		Count		Ratio(%)	
		Drug	Hacking	Drug	Hacking
Marketplace		5,330	3,444	95.92	69.90
Community		110	427	1.98	8.67
Broker		35	132	0.63	2.68
HTTP Status Code	404	18	-	0.32	-
	408	-	66	-	1.34
	500	30	-	0.54	-
	502	14	-	0.25	-
	503	3	-	0.05	-
Normal		17	858	0.31	17.41
Total		**5,557**	**4,927**	**100**	**100**

또한, 마약 사이트의 경우, 'Marketplace' 유형이 대부분을 차지하는 반면, 'Community'는 110개(1.98%), 'Broker'는 35개(0.63%), 'Normal'은 17개(0.31%)로 매우 낮은 비율을 보였다. 이는 마약 관련 다크웹이 주로 직접적인 거래 목적으로 운영되고 있음을 나타낸다. 반면, 해킹 사이트는 'Marketplace' 외에도 'Normal' 858개(17.41%), 'Community' 427개(8.67%), 'Broker' 132개(2.68%)로 상대적으로 높은 비율을 나타냈다. 이는 해킹 관련 다크웹이 단순 거래뿐만 아니라 기술 정보 공유, 일반 보안 관련 콘텐츠 제공, 사이트 중개 등 다양한 목적으로 활용되고 있음을 의미한다. 또한, 마약 사이트에서 다양한 에러 코드가 발생하는 것으로 보아, 마약 사이트가 더 빈번한 구조 변경이나 부분 폐쇄를 통해 수사를 회피하는 반면, 해킹 사이트는 상대적으로 안정적인 서버 운영을 유지하고 있음을 알 수 있다.

본 연구에서는 마약과 해킹 범죄 다크웹 서비스 접속 시 발생하는 실제 네트워크 트래픽으로부터 L3/4 메트릭 16개, L7 메트릭 30개를 수집하여 기계학습을 위한 데이터셋을 구축한다. L3/4 메트릭은 Table 5와 같이 총 패킷 수, 고유 IP 수, 평균 TTL(Time To Live), 초당 패킷 수(pps; Packet Per Second) 등으로 구성되며 주로 트래픽의 양적 규모, 네트워크 경로 복잡도, 전송 효율성 등 물리적·전송 계층 수준의 통신 특성을 나타낸다. 특히, 고유 IP 수와 평균 TTL은 다크웹 사이트가 익명성 확보를 위해 여러 중계 노드를 경유하는 특성을 반영하며 초당 패킷 수는 서비스 규모와 활성도를 나타내는 지표로 활용된다.

Table 5. L3/4 and L7 Metric Types

Layers	Metrics
L3/4	Total packets, Source ips, Destination ips, Total bytes, Average packet size, Duration second, pps, bps, Min packet size, Max packet size, Average TTL, Min TTL, Max TTL, Average window, Min window, Max window
L7	Internal links, External links, Images, Form count, Script count, CSS files, Input fields, JS libraries, Meta tags, Title, Content length, Text length, Content type, Character set, Content encoding, Content hash, Content language, HSTS(HTTP Strict Transport Security), CSP(Contents Security Policy), XFO,(X-Frame-Options) Cookies, CORS headers, Security headers, Server, Powered by, ETag, Last modified, Via header, Total time (ms), TTFB(Time to first byte) (ms)

L7 메트릭은 내·외부 링크 수, 이미지 수, 콘텐츠 크기, HSTS(HTTP Strict Transport Security) 헤더 존재 여부, 총 응답 시간 등으로 구성되며 주로 사이트의 구조적 복잡도, 콘텐츠 풍부도, 보안 정책, 서버 응답 성능 등 애플리케이션 계층의 특성을 나타낸다. 특히, 외부 링크 비율이 높은 사이트는 'Broker' 유형일 가능성이 높으며 HSTS 헤더 사용은 상대적으로 높은 보안 수준을 의미한다. 또한, 콘텐츠 크기와 이미지 수는 'Marketplace'와 'Community'를 구분하는 주요 지표로 활용된다.

2. Machine Learning-based Fingerprinting Model for Dark Web Drug Marketplaces

본 연구에서는 다크웹 범죄 사이트 유형을 자동 분류하기 위해 3.1절에서 구축한 데이터셋을 활용하여 XGBoost, Random Forest, Logistic Regression, SVM 기반 학습 모델을 설계한다. 학습 과정에서는 L3/4 메트릭만을 사용한 모델, L7 메트릭만을 사용한 모델, 두 계층 메트릭을 결합한 하이브리드 모델을 각각 구축하여 계층별 분류 성능을 비교 분석한다.

본 연구에서는 데이터셋의 클래스 불균형 문제를 해결하기 위해 SMOTE(Synthetic Minority Over-sampling Technique) 기반 오버샘플링과 클래스 가중치를 병행 적용하였다. SMOTE는 소수 클래스의 합성 샘플을 생성하여 학습 데이터의 균형을 맞추며 클래스 가중치는 모델 학습 시 소수 클래스에 더 높은 중요도를 부여한다. 이를 통해 'Marketplace'와 같은 다수 클래스뿐만 아니라,

'Broker', 'Community' 등 소수 클래스에 대한 분류 성능을 향상시킬 수 있다.

IV. Experimental Results

본 장에서는 제안한 네트워크 핑거프린트 기반 다크웹 탐지 모델의 실험 설정 및 성능 분석 결과를 제시한다. 실험은 마약 및 해킹 두 가지 범죄 유형을 대상으로 수행하였으며 L3/4 메트릭만을 학습한 단일 모델, L7 메트릭만을 학습한 단일 모델, 두 계층 메트릭을 결합한 하이브리드 모델의 성능을 비교 분석한다. 머신러닝 모델로는 3.2절에서 설계한 것과 같이 XGBoost, Random Forest, Logistic Regression, SVM 4가지 알고리즘을 사용한다.

실험에 사용된 각 머신러닝 알고리즘의 하이퍼파라미터는 Table 6과 같다. 하이퍼파라미터는 예비 실험을 통한 경험적 튜닝을 기반으로 설정하였다. Hybrid 모델의 경우 특성 차원이 높아(37개) 과적합 방지를 위해 더 강한 정규화를 적용하였다. 모든 모델에서 클래스 불균형 해소를 위해 balanced class weight를 적용하였다.

Table 6. Hyperparameter Settings

Model	Hyperparameter	L3-4 / L7	Hybrid
XGBoost	n_estimators	100	150
	max_depth	5	6
	learning_rate	0.1	0.08
	subsample	0.8	0.85
	colsample_bytree	0.8	0.8
	reg_alpha	0.3	0.5
	reg_lambda	0.8	1.0
	min_child_weight	2	3
	gamma	0.1	0.2
Random Forest	n_estimators	120	150
	max_depth	8	10
	min_samples_split	3	5
	min_samples_leaf	1	2
	max_features	0.8	0.7
Logistic Regression	solver	saga	saga
	max_iter	2000	3000
	C	0.5	1.0
	penalty	elasticnet	elasticnet
	l1_ratio	0.3	0.5
SVM	kernel	RBF	RBF
	C	1.5	2.0
	gamma	auto	auto

1. Comparative Analysis of Detection Performance across Models

Fig. 1은 마약 및 해킹 범죄 다크웹 사이트에 대한 서비스 유형 분류 성능을 나타낸다. 4가지 머신러닝 알고리즘을 비교한 결과, XGBoost가 두 범죄 유형 모두에서 가장 우수한 F1-score를 기록하였다. XGBoost 모델의 계층별 성능을 분석한 결과, 마약 사이트는 L3/4 단일 모델 0.749, 하이브리드 모델 0.993, L7 단일 모델 0.996을 기록하였으며 해킹 사이트는 각각 0.649, 0.928, 0.933을 기록하였다.

두 범죄 유형 모두 L7 단일 모델이 가장 높은 성능을 보였으며 하이브리드 모델 역시 L3/4 단일 모델 대비 크게 향상된 성능을 기록하였다. 한편, 마약 사이트의 분류 성능(L7: 0.996)이 해킹 사이트(L7: 0.933)보다 높게 나타났는데 이는 Table 4에서 확인된 서비스 유형 분포 차이에 기인한다. 마약 사이트는 'Marketplace'가 95.92%로 가장 높은 비율을 차지하여 상대적으로 단순한 분포를 보이는 반면, 해킹 사이트는 'Marketplace' 69.9%, 'Normal' 17.41%, 'Community' 8.67% 등 다양한 유형이 혼재하고 있어 분류 난이도가 높기 때문으로 해석된다.

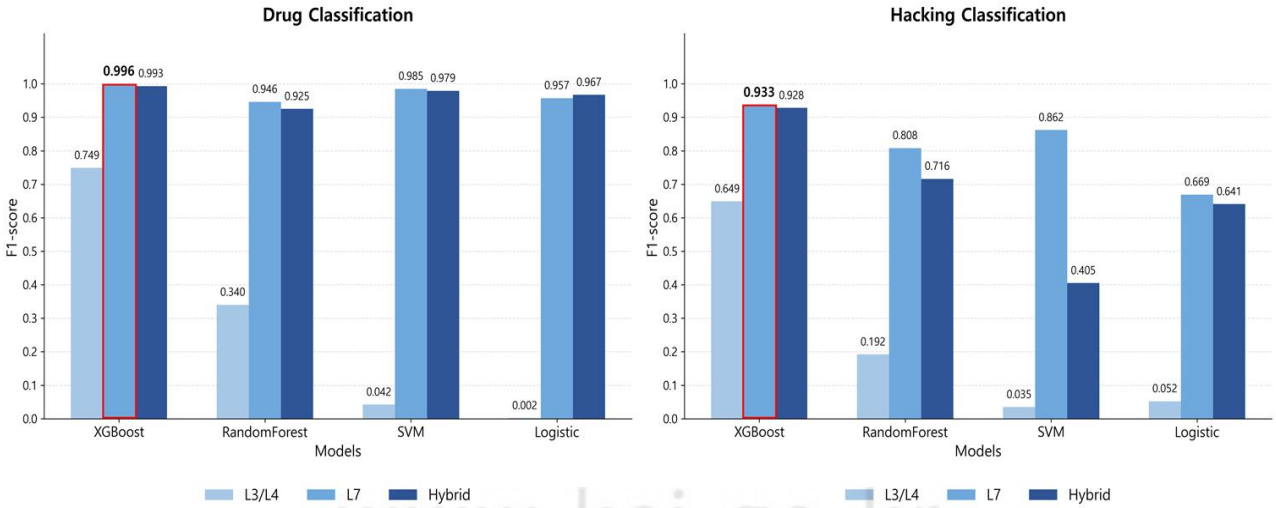

Fig. 1. Performance Comparison of Machine Learning Models for Drug and Hacking Crime Classification

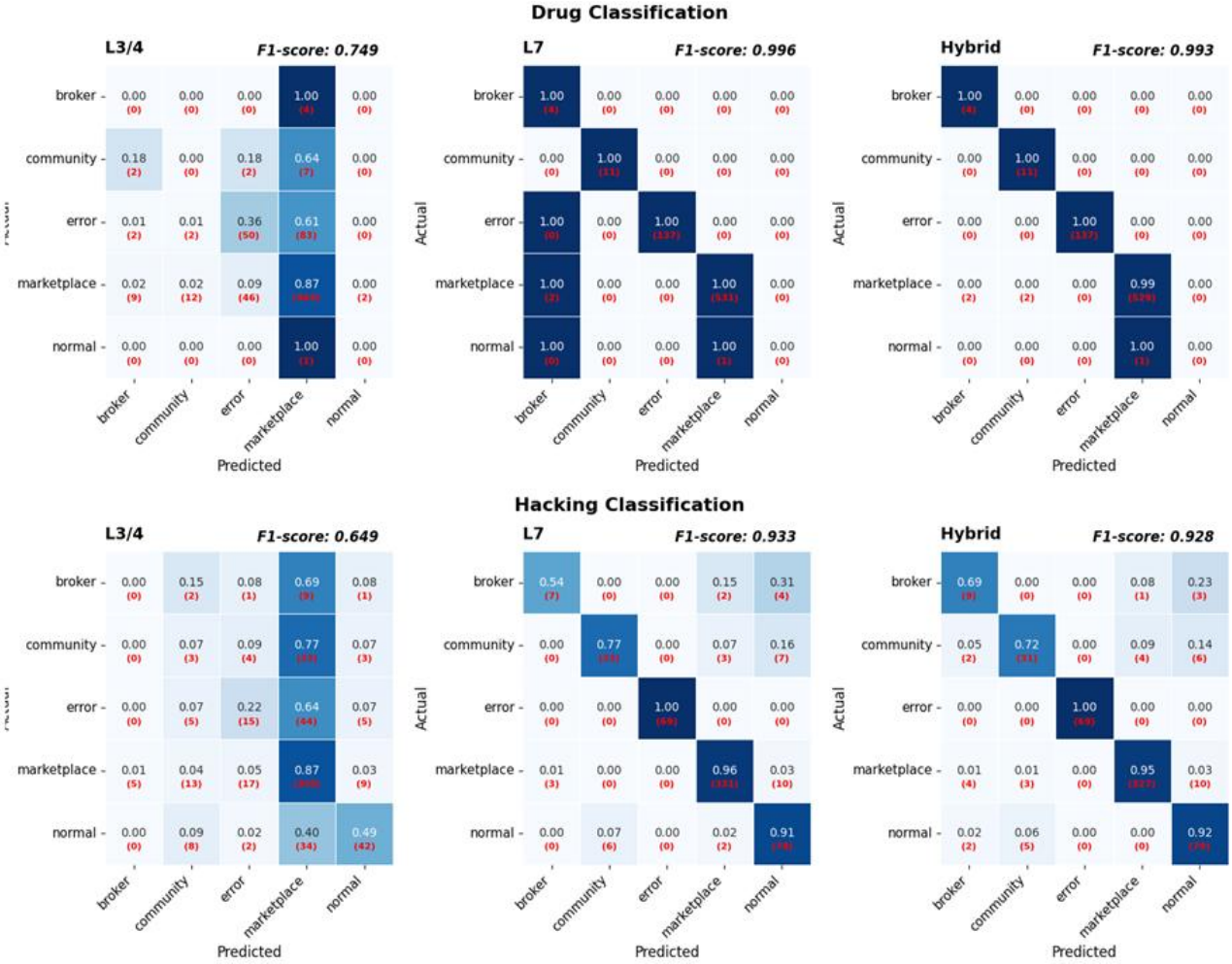

Fig. 2. Confusion Matrix of XGBoost Model for Drug and Hacking Crime Classification

또한, L3/4 단일 모델의 성능이 마약(0.749)과 해킹(0.649) 모두에서 상대적으로 낮게 나타났으며, 이는 물리적·전송 계층 메트릭만으로는 애플리케이션 수준의 서비스 유형을 구분하는 데 한계가 있음을 나타낸다.

2. Model-wise Analysis of Detection Results

본 연구에서는 마약 및 해킹 범죄 다크웹 사이트를 대상으로 OSI 계층별 네트워크 메트릭을 활용한 머신러닝 기반 탐지 성능을 분석하였다. 평가의 신뢰성을 혹보하기 위해 실험에는 3.1절에서 구축된 데이터셋을 학습(80%), 검증(10%), 평가(10%)로 분할하여 서로 다른 랜덤 시드(42~51)를 사용하여 반복 실험을 수행하고, 평균값을 최종 결과값으로 사용하였다.

또한, 데이터 불균형 문제를 완화하기 위해 SMOTE 기반 오버샘플링과 클래스 가중치를 병행 적용하였다. Fig. 2는 마약 및 해킹 범죄 사이트에 대한 XGBoost 모델의 Confusion Matrix를 나타낸다.

마약 사이트의 경우, L7 모델이 F1-score 0.996으로 가장 높은 성능을 기록하였으며 모든 클래스에서 90% 이상의 분류 정확도를 달성하였다. Hybrid 모델은 0.993을 기록하였으나, ‘Normal’ 클래스에서 일부 샘플이 ‘Marketplace’로 오분류되는 경향을 보였다. L3/4 모델은 0.749로 가장 낮은 성능을 기록하였으며 특히, ‘Community’와 ‘Marketplace’ 클래스를 구분하지 못하고 대부분 ‘Marketplace’로 분류하는 편향을 보였다.

해킹 사이트의 경우에는 L7 모델이 0.933, Hybrid 모델이 0.928, L3/4 모델이 0.649를 기록하였다. L7 모델은 ‘Marketplace’ 클래스에서 높은 정확도를 보였으나 ‘Broker’와 ‘Normal’ 클래스 간 일부 오탐이 관찰되었다. Hybrid 모델 역시 유사한 패턴을 보였으며 ‘Community’ 클래스에서 ‘Marketplace’로의 오분류가 일부 발생하였다. L3/4 모델은 ‘Broker’ 클래스 대부분을 ‘Normal’로 오분류하였으며 ‘Community’와 ‘Marketplace’를 명확히 구분하지 못하여 전반적으로 낮은 성능을 기록하였다. 이러한 성능 차이는 각 계층 메트릭의 특성에 기인한다. L3/4 계층은 패킷 수, 전송량, TTL, 윈도우 크기와 같은

물리적·전송 계층 정보를 포함하나 사이트의 기능적 차이를 반영하기 어렵다. 반면, L7 계층은 HTML 구조, 스크립트 수, 응답 시간, 보안 헤더 등 애플리케이션 레벨의 메트릭을 포함하여 사이트의 목적과 유형을 명확히 구분할 수 있다. 이러한 특성은 Fig. 3의 학습 곡선에서도 확인된다. L7과 Hybrid 모델은 약 20 Epochs 이후 손실(Log Loss)이 안정화되었으며 훈련·검증 곡선 간 편차가 거의 존재하지 않았다. 반면, L3/4 모델은 손실이 일정 수준 이상에서 정체되는 과소적합 현상을 보였으며 검증 손실이 안정되지 못하였다.

해킹 사이트의 분류 성능이 마약 사이트보다 낮게 나타난 것은 데이터의 구조적 다양성과 클래스 간 유사성 때문으로 해석된다. 정상적인 기술 커뮤니티와 해킹 관련 포럼은 스크립트 수, 내부 링크 수, 폼 구성에서 유사한 웹 구조를 보이므로 트래픽 기반 구분이 상대적으로 어렵다. 이는 향후 세션 행동 패턴(Session Behavior), 암호화 트래픽 특성(Encrypted Traffic Feature), 또는 비정상 통신 흐름(Anomaly Flow)을 추가 반영함으로써 개선될 수 있다. 결론적으로, 두 범죄 유형 모두에서 L7 모델이 가장 우수한 성능을 기록하였으며 Hybrid 모델은 L3/4 메트릭 추가에도 불구하고 L7 모델과 유사한 수준의 성능을 유지하였다. L3/4 메트릭은 단독 사용 시 낮은 분류 성능을 보였으나 Hybrid 모델에서는 과적합 억제와 일반화 성능 개선에 기여하는 것으로 나타났다.이러한 성능 차이는 계층별 메트릭이 담고 있는 정보의 특성에서 비롯된다. L7 메트릭은 HTML 구조, 스크립트 수, 콘텐츠 길이 등 사이트의 목적과 기능을 직접적으로 반영하는 반면, L3/4 메트릭은 패킷 크기, TTL 등 네트워크 전송 특성만을 포함하여 사이트 간 의미적 차이를 구분하기 어렵다. Hybrid 모델은 두 계층 정보를 결합하면서도 L3/4 정보가 분류 성능에 부정적 영향을 미치지 않도록 조정되었으며, 이를 통해 L7 수준의 성능을 유지하면서도 학습 안정성을 확보하였다. 특히 해킹 사이트의 분류 정확도가 마약 사이트보다 낮게 나타난 것은 서비스 유형의 구조적 다양성과 클래스 간 유사성에 기인한다. Fig. 3의 학습 곡선에서 볼 수 있듯이, L7과 Hybrid 모델은 빠른 수렴과 안정적인 학습을 보인 반면, L3/4 모델은 과소적합 양상을 나타냈다. 이는 물리적·전송 계층 정보만으로는 애플리케이션 수준의 서비스 특성을 구분하기 어려움을 의미하며 다크웹 사이트 유형 분류를 위한 핑거프린트 추출에는 애플리케이션 계층 메트릭이 더욱 효과적임을 나타낸다.

V. Conclusion

본 연구는 다크웹 환경에서 운영되는 마약 및 해킹 관련 불법 사이트를 효과적으로 탐지하고 분류하기 위한 머신러닝 기반 모델을 제안하고 그 성능을 검증하였다. 다크웹은 암호화 기술과 익명 네트워크를 활용하여 범죄 활동의 주요 플랫폼으로 기능하고 있으며 기존 콘텐츠 중심 탐지 방식만으로는 실시간 대응에 한계가 존재한다. 이에 본 연구는 다크웹 페이지 콘텐츠가 아닌 네트워크 트래픽의 계층별 메트릭을 활용한 핑거프린트 추출 방법을 제안하였다.

실험 결과, XGBoost 학습 알고리즘을 사용한 모델이 전반적으로 가장 우수한 성능을 보였으며 학습에 사용된 계층별 메트릭에 따라 성능 차이가 발생하였다. 마약 사이

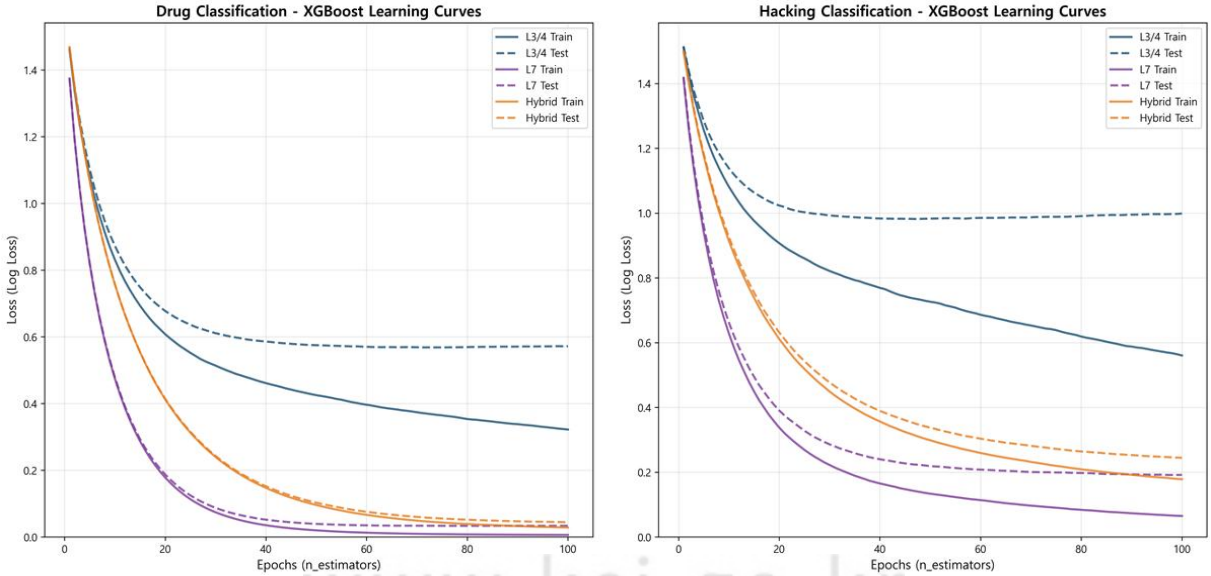

Fig. 3. Learning Curves of XGBoost Model for Drug and Hacking Crime Classification

트 분류에서 L7 단일 모델은 F1-score 0.996을 기록하였으며 해킹 사이트 분류에서는 0.933을 기록하였다. 이는 L3/4 단일 모델 대비 각각 0.247, 0.284 향상된 결과이며 Hybrid 모델 대비 각각 0.003, 0.005 향상된 수치이다. 학습 곡선 및 Confusion Matrix 분석 결과 L7 모델은 약 20 Epochs 이후 손실이 안정되었으며 주요 클래스에서 높은 정밀도와 재현율을 기록하였다.

이러한 결과는 애플리케이션 계층(L7) 메트릭이 다크웹 사이트 핑거프린트 추출에 효과적임을 입증한다. 특히 본 연구는 기존 연구들이 주로 L3/4 중심 분석이나 콘텐츠 기반 분석에 집중한 것과 달리, 직접 범죄 다크웹에 접속하여 L7 메트릭을 수집하고 그 유효성을 체계적으로 검증하였다는 점에서 차별성을 갖는다. 또한, 네트워크 트래픽 메트릭만으로 다크웹 범죄 사이트를 신속하게 분류할 수 있어 수사기관의 선제적 대응체계 구축에 실용적 기여를 할 수 있을 것으로 기대된다.

본 연구는 다음과 같은 한계를 갖는다. 첫째, 데이터 수집 대상이 마약과 해킹 범죄에 한정되어 있어 성범죄, 무기 거래, 자금 세탁 등 다른 범죄 유형에 대한 적용 가능성은 추가 검증이 필요하다. 둘째, 다크웹 트래픽은 암호화 기법 및 네트워크 우회 기술의 발전에 따라 지속적으로 변화하므로 모델의 지속적 성능 유지를 위한 업데이트 전략이 필요하다.

향후에는 영어뿐 아니라, 러시아어, 중국어 등 다양한 언어의 키워드를 추가 수집하기 위한 체계를 구축하고, Torch, Ahmia와 같은 검색엔진에 색인되지 않는 은닉 다크웹을 수집하기 위한 방법론도 개발할 예정이다. 또한, 성범죄, 테러, 불법 무기 거래 등 다양한 범죄 유형을 포함하는 멀티도메인 학습 구조를 적용하며 트래픽 메트릭 외에도 도메인 구조, URL 패턴, TLS 세션 정보 등을 융합하여 다크웹 서비스 유형을 분류하는 멀티모달 탐지 모델을 개발할 계획이다.

ACKNOWLEDGEMENT

This work was supported by 'Tech. Challenge for Future Program Policing([http://www.kipot.or.kr)'/) www.kipot.or.kr)' funded by Ministry of Science and ICT(MSIT, Korea) & Korean National Police Agency(KNPA, Korea). [Project Name : Development of Active Dark Web Information Collection, Analysis and Tracking Technology to Prevent Dark Web Crime / Project Number : RS-2023-00244362]

REFERENCES

[1] R. S. Basheer and B. AlKhatib, “Crawling the Dark Web: A Conceptual Perspective, Challenges and Implementation,” Journal of Digital Information Management, Vol. 17, No. 2, pp. 51-60, April 2019. DOI: 10.6025/jdim/2019/17/2/51-60.

[2] T. Fu, A. Abbasi & H. Chen, “A Focused Crawler for Dark Web Forums,” Journal of the American Society for Information Science and Technology, Vol. 61, No. 6, pp. 1213-1231, June 2010. DOI: 10.1002/asi.21323.

[3] P. Koloveas, T. Chantzios, C. Tryfonopoulos & S. Skiadopoulos, “A Crawler Architecture for Harvesting the Clear, Social, and Dark Web for IoT-Related Cyber-Threat Intelligence,” in Proceedings of the IEEE World Congress on Services (SERVICES), Milan, Italy, pp. 3-8, June 2019. DOI: 10.1109/SERVICES.2019.00016.

[4] R. Jegan, V. Rajavarman & S. Geetha, “AI-Enhanced Dark Web Crawler for Cybersecurity Monitoring,” Tuijin Jishu / Journal of Propulsion Technology, Vol. 45, No. 2, pp. 4395-4404, April 2024. DOI: 10.52783/tjjpt.v45.i02.6660.

[5] M. Campobasso & L. Allodi, “THREAT/crawl: A Trainable, Highly-Reusable, and Extensible Automated Method and Tool to Crawl Criminal Underground Forums,” in Proceedings of the APWG eCrime Researchers Summit (eCrime), Oct. 2022, pp. 1-13. DOI: 10.1109/eCrime57793.2022.10142081.

[6] A. S. Rajawat, P. Bedi, S. B. Goyal, S. Kautish & Z. Xihua, “Dark Web Data Classification Using Neural Network,” Computational Intelligence and Neuroscience, Vol. 2022, Art. ID 8393318, Mar. 2022. DOI: 10.1155/2022/8393318.

[7] G.-Y. Shin, Y. Jang, D.-W. Kim, S. Park, A.-R. Park, Y. Kim & M.-M. Han, “Dark Side of the Web: Dark Web Classification Based on TextCNN and Topic Modeling Weight,” IEEE Access, Vol. 12, pp. 36361-36371, 2024. DOI: 10.1109/ACCESS.2023.3347737.

[8] A. Dalvi, S. Bhoir, N. Naik, A. Kitkaru, I. Siddavatam & S. Bhirud, “A Hybrid TF-IDF and RNN Model for Multi-Label Classification of the Deep and Dark Web,” International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 14, No. 7, 2023. DOI: 10.14569/IJACSA.2023.01407106.

[9] A. Pamuji, “Investigation of the Dark Web Illegal Activities Using Data Mining Approach,” Bulletin of Computer Science & Electrical Engineering, Vol. 4, No. 1, pp. 37-48, Jun. 2023. DOI: 10.25008/bcsee.v4i1.1179

[10] G. Cascavilla, G. Catolino & M. Sangiovanni, “Illicit Darkweb Classification via Natural-Language Processing: Classifying Illicit Content of Webpages based on Textual Information,” in Proc. 19th Int. Conf. Security and Cryptography (SECRYPT), Jul. 2022, pp. 620-626. DOI: 10.5220/0011298600003283.

[11] A. Fayzi, M. Fayzi & K. D. Ahmadi, “Dark Web Activity Classification Using Deep Learning,” arXiv preprint arXiv:2306.

07980, June 2023. DOI: 10.48550/arXiv.2306.07980.

[12] B. V. Babu & K. K. V. Dhatri, “Identifying Similar Users Between Dark Web and Surface Web Using BERTopic and Authorship Attribution,” Electronics, Vol. 14, No. 1, Art. ID 148, 2025. DOI: 10.3390/electronics14010148.

[13] A. H. M. Alaidi, R. M. Al-Airaji, H. T. S. Alrikabi, I. A. Aljazaery & S. H. Abbood, “Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques,” International Journal of Interactive Mobile Technologies (iJIM), Vol. 16, No. 10, pp. 123–139, Oct. 2022. DOI: 10.3991/ijim.v16i10.32707.

[14] V. V. Nair, M. van Staalduinen & D. T. Oosterman, “Template Clustering for the Foundational Analysis of the Dark Web,” in Proceedings of the 2021 IEEE International Conference on Big Data (BigData), pp. 2542-2549, USA, December 2021. DOI: 10.1109/BigData52589.2021.9671936.

[15] Y. Boshmaf, I. Perera, U. Kumarasinghe, S. Liyanage & H. Al Jawaheri, “Dizzy: Large-Scale Crawling and Analysis of Onion Services,” in Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES), Benevento, Italy, pp. 1–11, August 2023. DOI: 10.1145/3600160.3600184.

[16] S. Scharnowski, “Dark web traffic, privacy coins, and cryptocurrency trading activity,” Finance Research Letters, Vol. 67, Article PB, 2024. DOI: 10.1016/j.frl.2024.105875.

[17] A. Montieri, D. Ciuonzo, V. Persico & A. Pescapé, “A Dive into the Dark Web: Hierarchical Traffic Classification of Anonymity Tools,” IEEE Transactions on Network Science and Engineering, Vol. 7, No. 3, pp. 1043-1054, July 2019. DOI: 10.1109/TNSE.2019.2901994.

[18] J. Li & Z. Pan, “Dark Web Traffic Classification Based on Spatial–Temporal Feature Fusion and Attention Mechanism,” Computers, Vol. 14, No. 7, Article 248, 2025. DOI: 10.3390/computers14070248.

[19] J. Li, Z. Pan & K. Jiang, “A Three-Dimensional Convolutional Neural Network for Dark Web Traffic Classification Based on Multi-Channel Image Deep Learning,” Computers, Vol. 14, No. 8, Article 295, 2025. DOI: 10.3390/computers14080295.

[20] J. Zhai, H. Sun, C. Xu & W. Sun, “ODTC: An Online Darknet Traffic Classification Model Based on Multimodal Self-Attention Chaotic Mapping Features,” Electronic Research Archive, Vol. 31, No. 8, pp. 5056-5082, July 2023. DOI: 10.3934/era.2023259.

[21] U.S. Drug Enforcement Administration (DEA), Drug Slang Code Words — DIR-022-18, July 2018. Available: https://www.dea.gov/sites/default/files/2018-07/DIR-022-18.pdf

[22] MITRE Corporation, MITRE ATT&CK® Framework: Enterprise Matrix, 2024. Available: https://attack.mitre.org/

Authors

Jinwoo Shin is an undergraduate student in the Department of Cyber Security, Daegu University, Gyeongsan, South Korea, since 2021. His research interests include cybersecurity, Internet of Things (IoT), and cloud computing.

Dong-Won Kang is an undergraduate student in the Department of Computer Engineering, Daegu University, Gyeongsan, South Korea, since 2021. His research interests include cybersecurity, Dark Web crime investigation, and artificial intelligence.

Jiyeon Kim received the B.S. and Ph.D. degrees in information security engineering from Seoul Women’s University, Seoul, South Korea, in 2007 and 2013, respectively. Dr. Kim was a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering, Carnegie Mellon University, United States, from 2014 to 2017. She is currently an Assistant Professor in the Department of Computer Engineering, Daegu University, Gyeongsan, South Korea. Her research interests include cybersecurity, cybercrime investigation, cloud computing,artificial intelligence, and critical infrastructure protection.