Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.34, NO.4, Aug. 2024 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2024.34.4.577

인공지능 기반 개체명 인식 기술을 활용한 보안 위협 정보 식별 방안 연구*

김 태 현,[1†] 임 준 형,[2] 김 태 은,[1] 엄 익 채[3‡]
[1,2]한국인터넷진흥원 (연구원, 팀장), [3]전남대학교 (교수)

A Study on the Identification Method of Security Threat Information Using AI Based Named Entity Recognition Technology*

Taehyeon Kim,[1†] Joon-Hyung Lim,[2] Taeeun Kim,[1] Ieck-chae Euom[3‡]
[1,2]Korea Internet & Security Agency (Researcher, Manager),
[3]Chonnam National University (Professor)

요 약

새로운 기술이 개발 됨에 따라, 랜섬웨어를 만들어 주는 AI 기술 등장과 같은 새로운 보안 위협도 증가되고 있다. 이러한 보안 위협에 대응하기 위해 XDR와 같은 신규 보안장비가 개발되었지만, 단일 보안장비 환경이 아닌 다양한 보안장비를 함께 사용하는 경우 필수 데이터 식별 및 분류를 위해 수많은 정규표현식을 만들어야 하는 어려움이 존재한다. 이를 해결하기 위해 본 논문에서는 다양한 보안장비 사용 환경에서 인공지능 기반 개체명 인식 기술을 도입하여 위협 정보 식별을 위한 필수 정보 식별 방안을 제안한다. 보안장비 로그 데이터를 분석하여 필수 정보를 선정한 뒤, 정보의 저장 포맷과 인공지능을 활용하기 위한 태그 리스트를 정의하였고, 인공지능을 이용한 개체명 인식 기술을 통해 필수 데이터 식별 및 추출 방안을 제안한다. 다양한 보안장비 로그 데이터와 23개의 태그 기반 개체명 인식 시험 결과 태그별 f1-score의 가중치 평균이 Bi-LSTM-CRF는 0.44, BERT-CRF는 0.99의 성능을 보인다. 향후 정규표현식 기반의 위협 정보 식별·추출 방안과 인공지능 기반의 위협 정보 식별·추출 방안을 통합하는 프로세스를 연구하고 신규 데이터 기반으로 프로세스를 적용해 볼 예정이다.

ABSTRACT

As new technologies are developed, new security threats such as the emergence of AI technologies that create ransomware are also increasing. New security equipment such as XDR has been developed to cope with these security threats, but when using various security equipment together rather than a single security equipment environment, there is a difficulty in creating numerous regular expressions for identifying and classifying essential data. To solve this problem, this paper proposes a method of identifying essential information for identifying threat information by introducing artificial intelligence-based entity name recognition technology in various security equipment usage environments. After analyzing the security equipment log data to select essential information, the storage format of information and the tag list for utilizing artificial intelligence were defined, and the method of identifying and extracting essential data is proposed through entity name recognition technology using artificial intelligence. As a result of various security equipment log data and 23 tag-based

Received(03. 08. 2024), Modified(05. 08. 2024), Accepted(06. 07. 2024)
* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2021-0-00356,AI·빅데이터 기반 사이버 보안 오케스트레이션 및 자동 대응 기술 개발)
† 주저자, thkim@kisa.or.kr
‡ 교신저자, iceuom@chonnam.ac.kr(Corresponding author)

entity name recognition tests, the weight average of f1-score for each tag is 0.44 for Bi-LSTM-CRF and 0.99 for BERT-CRF. In the future, we plan to study the process of integrating the regular expression-based threat information identification and extraction method and artificial intelligence-based threat information and apply the process based on new data.

Keywords: Log data analysis, Named entity recognition, Security essential information, Security equipment

I. 서 론

현재 ChatGPT와 같은 생성형 AI 기술의 발전과 빅데이터 처리·분석 기술이 고도화됨에 따라 반복적이고 규칙적인 작업의 자동화와 맞춤형 광고, 추천 시스템 등 인터넷 서비스 사용자의 편의성이 크게 향상되었다. 그러나 이러한 새로운 기술적 편의성이 동반되면서 이를 활용한 보안 위협이 증가되었고, 이에 대응하기 위해 보안 분야에서는 SOAR(Security Orchestration, Automation and Response), EDR(Endpoint Detection and Response), XDR(eXtended Detection and Response) 등 다양한 보안장비 및 기술이 발전하였다. SOAR는 보안 이벤트를 자동화하고 조율하여 신속한 대응을 가능하게 하며, EDR은 단말기에서의 위협을 탐지하고 대응하는 데 중점을 둔다. XDR은 여러 보안 도구와 데이터 소스를 통합하여 보안 위협에 대한 종합적인 대응 방안을 사용자에게 제공한다.

대다수 기업은 새로운 보안 위협을 방어하기 위해 신규 보안장비들과 함께 기존 방화벽, 침입차단시스템, 침입탐지시스템 등 멀티 벤더 보안장비를 혼합하여 사용한다[1].

하나의 보안장비에서 동일한 보안 위협이더라도 들어올 때마다 문자열이 상이하기 때문에 특정 보안 위협을 탐지하기 위해서는 많은 정규표현식이 필요하다[2]. 대표적인 침입 탐지 장비인 IDS(Intrusuion-Detection System)를 보더라도, HIDS는 호스트 보안을 담당하고 NIDS는 네트워크 보안을 담당하고 있어 IDS라는 한 종류의 보안장비이지만 기능에 따라 사용하는 장비가 다양해지고 있다. 이처럼 서로 각기 다른 영역을 보안하는 보안장비가 늘어날수록, 데이터 분석을 위해 필요한 정규표현식은 기하급수적으로 증가한다. 특정 협력 관계를 맺은 업체들에 한하여 통합 데이터 분석이 가능한 경우도 있지만, 보편적으로 이러한 문제를 자체적으로 해결하기에 많은 어려움이 있다. 따라서 본 논문에서는 다양한 보안장비 환경에서 데이터 분석을 원활하게 진행하기 위해 인공지능 기반의 개체명 인식 기술을 활용한 로그 데이터의 필수 정보 식별 및 추출 방안을 제시하고자 한다.

인공지능을 통한 보안 위협 정보 변환 방법은 모델 구조마다 상이하지만 다양한 연산 과정을 통해 정보가 어떤 위협 정보인지 판단하므로, 단순 정규표현식 패턴 기반의 변환 방법보다 변환 속도가 느리다. 하지만 기존에 없는 보안 위협 정보가 들어온다면, 이를 탐지하고 분석할 수 있는 형태로 변환하기 위해 정규표현식을 새로 만들어야 하고, 복잡한 정보를 담고 있는 보안 로그라면 사용자가 해당 정보를 파악하는데도 상당한 시간이 소요된다는 단점이 있다. 이처럼 다양한 보안 위협을 식별하기 위해 수많은 정규표현식을 만들어 내야 하는 어려움을 해결하기 위해 인공지능 기반의 개체명 인식 기술을 활용하여 보안 위협 정보를 식별하는 연구를 진행하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 다양한 보안장비 데이터 및 보안 로그 표준, 정규표현식을 이용한 로그 데이터 정보 추출 관련 연구를 소개하며, 3장에서는 보안 위협 데이터 식별·추출 방안을 제안한다. 4장에서는 제안한 보안 위협 데이터 식별 추출 방안의 성능 시험과 그 결과를 소개하며, 5장에서 결론을 맺는다.

II. 관련 연구

본 장에서는 국내외 보안 위협 정보 표준과 동향에 대해 분석하고, 기존 log parser 및 보안관제 분야에 인공지능 활용 연구를 소개한다.

2.1 다양한 보안장비 데이터와 보안 로그 표준

보안장비 로그 데이터는 attack type, port, ip, date, action과 같은 정보로 구성되어 있어, 일반적인 자연어와는 형태가 다르다. 보안장비 제조사들은 보안장비 포맷 정보를 공개하기에는 다양한 위험이 존재하기 때문에 보안장비 원본 로그 데이터 포맷을 공개하지 않는다. 이러한 이유로 데이터 확보의 어려움이 존재한다. 또한 보안장비에 어떤 정보를

포함할지는 보안장비를 만드는 업체의 선택사항이기 때문에, 각 업체에 보안장비가 어떤 위협 정보를 포함하고 있는지 정해진 바가 없다. 관련하여 국내외 표준을 분석해 보면, 국제인터넷표준화기구(IETF), 미국 표준연구소(NIST) 표준화 기구의 표준 조사 결과 보안 로그 데이터 글로벌 표준은 CybOX[3], STIX/TAXII[4]가 존재한다. CybOX는 사이버 보안상 악성코드 분석, 보안 운영, 포렌식, 로그 통합 분석 등 다양한 보안 분야에서 사용할 수 있는 플랫폼 기반 정보를 제공한다. STIX/TAXII는 로그 통합 분석한 내용, 포렌식, 악성코드 등의 위협정보 내용을 TTP(Tactic, Techniques and Procedure) 기반으로 활용할 수 있도록 정보를 표현한다. 국내의 경우 한국정보통신기술협회(TTA)의 관련 표준에 따라 로그 정보 표준이 정의되어 있다. 국내 TTA표준은 보안 로그 포맷에 중점이 아닌 일반 이벤트 로그에 중점을 두거나, 보안 로그에 들어갈 수 있는 정보의 유형을 정의한다[5-8]. 국내 표준은 보안 로그 포맷에 대한 명확한 표준이 존재하지 않으며, 국외 CybOX, STIX/TAXII 표준은 TTP로 활용할 수 있는 악성코드 해시값, 사이버 킬체인 등 다양한 정보를 표준으로 정의한다. 이는 위협 정보를 파악하는 과정에서 불필요한 데이터가 증가하는 문제를 야기한다.

2.2 Log Parser

로그 파싱 및 분석에 사용되는 가장 기본적인 방식은 원본 로그 데이터에서 공통부분을 추출하여 이를 로그 템플릿으로 지정하고, 나머지 부분을 로그 매개 변수로 처리한다[9]. 예를 들어 Table 1에서 좌측 Type 값인 “src_ip”, “src_port”와 같은 정보를 로그 템플릿으로 지정하고 그에 해당하는 값 즉, value 열에 있는 “192,168.0.1”,“443” 등을 매개변수로 정의한다.

Table 1. Information mapping Table format

Type	value
src_ip	192.168.0.1, 192.168.0.2,...
src_port	21,80,443,...
dst_ip	192.168.0.3, 192.168.0.4,...
dst_port	21,80,443,...
stime	19910101-11:30:31,19910101-12:31:31,...
etime	19910101-13:30:31,19910101-14:41:41,...

Table 2. Identification results using regular expressions

UFW log data	Identification result
Nov 10 12:34:56 host name kernel: [UFW BLOCK] IN=eth0 OUT= MAC=00:11:22:33:44:55:66:77:88:99:aa:bb:cc SRC=192.168.1.1 DST=192.168.1.2 LEN=40 TOS=0x00 PREC=0x00 TTL=64 ID=1234 PROTO=TCP SPT=12345 DPT=80 WINDOW=1024 RES=0x00 SYN URGP=0	stime:Nov 10 12:34:56 src_ip: 192.168.1.1 src_port: 12345 dst_ip: 192.168.1.2 dst_port: 80 etc: {“hostname”, “UFW BLOCK”, “kernel”, “IN=eth0”, “OUT”, “MAC=00:11:22:33:44:55:66:77:88:99:aa:bb:cc”, “LEN=40”, “TOS=0x00”, “PREC=0x00”, “TTL=64”,...}.

필드명이 없거나 인식이 안 되는 데이터는 사용자가 직접 정규표현식을 생성하는데, Table 2의 좌측 UFW(Uncomplicated Firewall)를 식별하기 위해 “^(\w{3} \d{1,2} \d{2}:\d{2}:\d{2}) (\S+) kernel: \[(UFW [A-Z]+)\] IN=(\w+) OUT = MAC=([0-9a-fA-F:]+) SRC=([0-9.]+) DST=([0-9.]+) LEN=(\d+) TOS=0x(\w+) PREC=0x(\w+) TTL=(\d+) ID=(\d+) PROTO=(\w+) SPT=(\d+) DPT=(\d+) WINDOW=(\d+) RES=0x00 SYN URGP=(\d+)” 와 같은 정규표현식을 생성하여 매개 변수를 추출한다. 해당 정규표현식을 통한 식별 결과를 그룹 지어 그룹명을 정의해 주면 Table 2의 identification result와 같다.

그 외에 김희두 등은 범죄 수사 분야에 사전학습 언어모델을 도입하여 개체명 인식을 적용하였다. 다양한 범죄 정보를 자동으로 파싱하기 위해 한글 데이터셋을 기반으로 모델을 학습시킨 뒤 한글 기반의 언어모델인 KoBERT, KoELECTRA에 파인 튜닝하여 개체명 인식 연구를 진행하였다. 본 연구는 한글이 아닌 특수문자, 영어, 숫자 등 다양한 문자가 포함된 로그이므로, 구분할 요소가 더 복잡하다[10].

2.3 보안관제 분야 인공지능 활용 연구

현정훈 등[11]은 ELK Stack을 활용한 보안관제 시스템을 제안했다. 해당 논문에서는 단일 보안장

Table 3. information list included in security equipment

equipment / elements	A	B	C	D	E	F	G	H	I	J
date	O	O	O	O	O	O	O	O	O	O
detection equip	O	O	O	O	O	O	O	O	O	O
source ip	O	O	O	O	O	O	O	O	O	O
source port	O	O	O	O	O	O	O	O	O	O
destination ip	O	O	O	O	O	O	O	O	O	O
destination port	O	O	O	O	O	O	O	O	O	O
ip address type	O	X	O	X	X	X	X	O	O	O
session id	O	O	X	O	O	O	O	X	X	O
country information	X	O	X	O	X	X	O	X	X	X
risk level	O	O	X	O	O	X	X	O	X	O
action(response)	O	O	O	O	O	X	O	O	O	X
detection policy	O	O	O	O	O	O	O	O	O	O
allowed packet count	X	O	O	O	X	X	O	X	X	O
blocked packet count	X	O	O	O	X	X	O	X	X	O

비 환경에서 많은 양의 보안 이벤트가 발생하였을 때 ElasticSearch를 이용하여 검색엔진을 구성하였고, LogStash를 이용하여 발생한 이벤트를 수집하여 분석할 수 있는 포맷으로 변환한 뒤, 데이터 저장소로 데이터를 전달한다. Kibana를 이용하여 저장된 데이터를 시각화해 주는 구조를 제안하였다. 이로 인해 고가의 솔루션 도입 없이 해당 기업에 최적화된 침해 대응 기술을 확보할 수 있고, 패턴 위주의 침해사고 사후 대응 체계가 아닌 실시간 보안 위협 데이터 기반으로 행위 위주 탐지 및 대응 체계를 제안했다.

고광수 등[12]은 패턴 기반 보안관제 체계의 오탐, APT 공격 탐지 등 문제점을 제시하였고, 이를 해결하기 위해 인공지능 기반의 통합보안관제 체계를 제안했다. 해당 논문은 인공지능 기술을 보안관제에 적용하여 보안 위협 이벤트 탐지 시 이벤트의 정·오탐과 위험도를 판단한다. 이로 인해 잠재적 위험 감소 등 예방의 관점에서 보안성 제고 방안을 제안했다.

현재 보안관제 분야에서 이상 행위 탐지 및 대응을 위해 인공지능을 활용하여 고도화하는 방안들이 연구되고 있지만, 실제로 활용되고 있는 이기종 보안장비의 서로 다른 포맷을 가진 이벤트가 들어오는 환경에 대해 고려가 되고 있지 않다.

III. 제안하는 보안 위협 데이터 식별·추출 방안

3.1 식별을 위한 필수 정보 선정

보안 위협을 포함하고 있는 다양한 로그 데이터를 활용하기 위해서는 서로 다른 포맷의 로그 데이터를 하나의 포맷으로 통일시켜야 한다. 국내외 점유율이 높은 벤더사 별 보안장비의 데이터를 분석한 뒤, 필수 정보를 선정한다.

Table 3에 제시한 보안장비별 포함하고 있는 정보를 보면 “date”, “detection eqip.”, “source ip”, “source port”, “destination ip”, “destination port” 총 6가지 정보는 분석에 사용된 모든 보안장비가 포함하고 있다는 것을 알 수 있었다. 이 중에서 “detection equip.”은 위협 데이터로 활용하기에는 탐지 장비명은 위협 정보 식별에 적합한 데이터가 아니므로 배제하였다.

해당 필수 정보를 표현하는 방법도 source ip를 ‘s_ip’, ‘source_ip’, ‘src_ip’와 같이 보안장비마다 상이하였다. 이런 부분을 해소하기 위해 새로운 정보 표현 포맷을 Table 4와 같이 수립하였다.

Table 4. Information representation format

Type	Object
date	{stime}:"2023-01-01T10:55:13.123456"
	{etime}:"2023-01-01T10:55:13.234567"
source ip	{{src_ip}:"192.168.0.1"
source port	{src_port}:"443"
destination ip	{dst_ip}:"192.168.0.1"
destination port	{dst_port}:"443"
etc	{etc}:"A, B, C"

3.2 데이터 식별을 위한 인공지능 모델 설계

불특정 다수의 보안장비 로그 데이터를 분석해야 하는 환경에서 정규표현식을 활용한다면 정규표현식 생성에 소모되는 시간, 인력 등 비효율적인 문제가 존재한다. 이러한 문제를 극복하기 위해, 본 논문에서는 인공지능 기반 개체명 인식(NER: Named Entity Recognition) 기술을 통해 사용자 개입 없이 자동으로 ip, port 등 위협 정보를 식별하는 시스템을 제안한다. 본 장에서는 개체명 인식 기술에 적합한 인공지능 태깅, 알고리즘 순서로 학습 모델을 설계하였다.

3.2.1 다양한 보안 데이터 식별을 위한 태깅 기법

다수의 태깅 기법이 존재하지만, 본 연구에서는 ip, port 등 개체명 인식에 적합한 태깅 기법인 BIO 태깅 기법을 사용한다. BIO 태깅 기법은 의미있는 개체의 시작점을 "B(Begin)"로 태깅하고, 의미있는 개체의 내부를 "I(Inside)"로 태깅한다. 그리고 의미 없는 개체를 "O(Outside)"로 태깅한다. 앞서 선정한 필수 정보를 기반으로 식별하고자 하는 정보를 Table 5와 같이 선정하였다.

B(Begin)로 시작하는 태그 11개, I(Inside)로 시작하는 태그 11개, 마지막으로 O(Outside)로 시작하는 태그 1개 등 총 23개로 구성되어 있으며, 필수 정보 요소 중 date의 다양한 로그 표준을 분석해 보았을 때, start_time 즉, 최초 탐지 시간과 end_time 종료 시간을 나누어 표준화되어 있다. 따라서 태깅 리스트 또한 구분하여 설계하였다. 이 외에 ip의 경우도 start_ip, destination_ip뿐만 아니라, NAT 환경에서의 ip, port를 포함하고 있는 보안장비가 많기 때문에, 해당 필수 정보 또한 'B-SRC-NAT-IP','I-SRC-NAT-IP'와 같은 방식으로 정의하였다.

Table 5. BIO tag list

BIO classification	Required fields
B(Begin)	SRC-IP, SRC-PORT, SRC-NAT-IP, SRC-NAT-PORT, DST-IP, DST-PORT, DST-NAT-IP, DST-NAT-PORT, IP-PROTOCOL, START-TIME, END-TIME
I(Inside)	
O(Outside)	O

3.2.2 데이터 식별을 위한 인공지능 알고리즘 모델 선정

다양한 분야에서 인공지능 기반의 개체명 인식을 위해 LSTM(Long Short-Term Memory)과 이전 단어나 패턴에 기반하여 결과값을 도출하는 LSTM의 한계점을 극복한 Bi-LSTM(Bidirectional LSTM)에 대한 많은 연구가 진행되고 있다[13,14]. 로그 데이터의 특성상 특히 필드 데이터의 위치가 중요하다. 예를 들어, soure_ip 다음에 destination ip가 오는 성질이 대표적이다. 많은 보안장비가 존재하지만, 본 연구에서 사용한 다양한 보안 로그 데이터들에 한해서는 해당 규칙을 벗어난 데이터는 존재하지 않았다. 따라서 Bi-LSTM의 대표적인 기술인 양방향 상태 전파가 높은 확률로 필수 정보 개체명 인식을 할 수 있다. 그리고 로그 데이터의 전체 필드도 필드 간에 문맥이 존재하기 때문에, BERT의 문맥 이해 기술과 다양한 보안장비 로그 데이터 확보 문제를 해결하기 위해 사전 훈련된 인공지능 알고리즘 특징을 활용하여 개체명 인식의 좋은 성능이 기대된다.

개체명 인식을 위해 BIO 태그 기법을 사용했기 때문에 태그 간 상관관계가 존재한다. 이런 경우 CRF(Conditional Random Fields) Layer를 추가하여 조건부 확률을 계산하여 최적의 레이블을 찾는 기능 때문에 성능을 개선한 연구 결과가 다수 존재한다[15,16]. 태그 간의 상관관계가 존재하기 때문에 CRF를 모델에 추가하여 개체명 인식 성능을 향상시키고자 한다.

설계한 인공지능 모델의 구성 요소는 두 모델에

공통적으로 CRF layer, Dropout layer를 포함하고 있으며, Dropout layer는 일부 뉴런을 무작위로 비활성화하여 모델의 복잡성을 줄여 특정 훈련 샘플의 의존성을 방지하고, 모든 특징을 균일하게 학습되도록 하여 과적합을 방지한다.

비교하고자 하는 두 인공지능 모델의 아키텍쳐 차이는 Bi-LSTM-CRF의 경우 Embedding layer, Dense layer, Bi-LSTM layer가 존재하며, BERT-CRF 모델은 BERT, 선형 분류기로 구성되어 있다.

3.3 식별된 보안 위협 데이터 추출

태그별로 추출할 데이터 매핑 테이블에 JSON형식으로 정의한다. {"src_ip":"B-SRC-IP", "I-SRC-IP", "I-SRC-IP", "I-SRC-IP", "src_port":"B-SRC-PORT","I-SRC-PORT","I-SRC-PORT","I-SRC-PORT", ...} 와 같이 매핑 테이블을 참고하여 추출한 태그 데이터를 재조합한다. 예를 들어 원본 데이터에 source ip:192.168.0.1가 포함된 경우 B-SRC-IP:192, I-SRC-IP:168, I-SRC-IP:##0, I-SRC-IP:##1 이라고 식별하고, 이를 조합하여 'src_ip':'192.168.0.1'으로 추출한다. 이와 같은 방식으로 port, date 정보를 추출한다.

IV. 개체명 인식 성능 시험

본 장에서는 Bi-LSTM-CRF, BERT-CRF의 시험환경 구성 및 인공지능을 활용한 필수 위협 정보 인식률 측정 결과와 태그 개수에 따른 개체명 인식 성능 결과를 설명한다.

4.1 시험환경 구성

본 시험에 사용한 HW/SW 정보는 Table 6과 같다.

데이터 셋의 경우 로그 포맷의 형태가 다른 FW 9종, IPS 3종, WAF 2종 총 14종의 서로 다른 보안장비에서 로그 포맷별 10개씩 추출하고, 필드 값 수치를 사용 가능한 범위 내에 데이터를 나열한 후, 일정 수만큼 랜덤하게 필드 값을 변경하여 데이터를 증식하였다. 데이터는 Table 7과 같이 학습·테스트 데이터로 구분하여 시험을 진행하였고, 데이터 누출(Data leakage) 현상을 방지하기 위해 F/W 1종을 제외한 13종의 보안장비 로그 데이터를 학습 데이터로 사용하였고, 학습에 사용되지 않은 FW 1종의 보안장비 로그 데이터를 테스트 데이터로 사용하였다.

Table 6. Simulation environment information

	Category	Version
SW	OS	Windows 11
	Program Language	Python 3.10.12
	Package Manager	Conda 23.7.3
	Notebook Platform	Jupyter 1.0.0
HW	CPU	i9-12900H
	RAM	32GB
	GPU	RTX 3080Ti

Table 7. Simulation data

Original		140 data
Data Augmentation		31,667 data
Total	Train	29,370 data
	Test	2,437 data

14종의 보안장비 데이터가 얼마나 다양한 포맷으로 이루어져 있는지 확인하기 위해 각 로그에 대하여 TF-IDF(Term Frequency-Inverse Document Frequency)를 기반으로 일부 단어를 추출하여 이를 기반으로 텍스트를 벡터화하고, 벡터화된 텍스트를 이용하여 코사인 유사도를 계산하였다. 14x14행으로 표

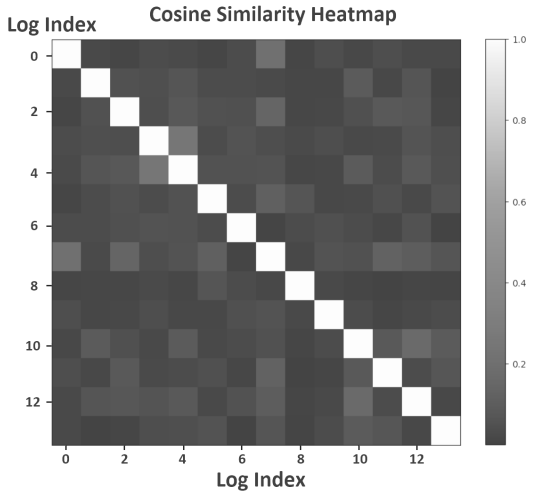


Fig. 1. Visualizing log similarity

현된 행렬에서 대각선 요소는 자기 자신의 유사도이므로 제외하고 평균을 계산해 보았을 때, 0.0465(이하 소수점 반올림)라는 결과를 확인할 수 있었다. IP나 날짜 데이터와 같은 필드 값 부분이 일부 동일하게 표현된다는 점을 감안했을 때, 14개의 보안장비 데이터는 전혀 다른 데이터임을 알 수 있다.

동일한 포맷의 데이터를 증식시킬 수도 있으나, 본 연구의 목적이 다양한 포맷을 가진 보안 위협 정보를 식별하고자 하는 연구이므로, 증식한 데이터를 추가 증식하지 않고 시험을 진행한다.

인공지능 모델 데이터 학습 시 Fig.2과 같이 검증 데이터의 손실함수 값이 Bi-LSTM-CRF는 10회까지 순차적으로 감소하여, 과적합이 발생하지 않았다는 것을 확인할 수 있었고, BERT-CRF는 4회부터 손실함수 값이 증가하여, epoch 4회부터 과적합이 발생함을 확인하였다. 따라서 본 시험은 epoch 설정을 Bi-LSTM-CRF는 10회, BERT-CRF는 3회로 설정하였다. 성능지표는 accuracy, recall, precision 등 다양한 지표가 존재하지만, 테스트 데이터의 클래스 간 불균형으로 인해 각 클래스 f1-score의 weighted-avg를 성능지표로 잡았다.

또한, 태그 개수의 영향을 알아보기 위해 로그 메시지 내에 메시지 내용을 뜻하는 'MSG' 관련 태그, 대응 정보를 뜻하는 'ACT' 관련 태그 등 24개의 태그를 추가하여 태그별 f1-score의 weighted avg를 비교하고자 한다.

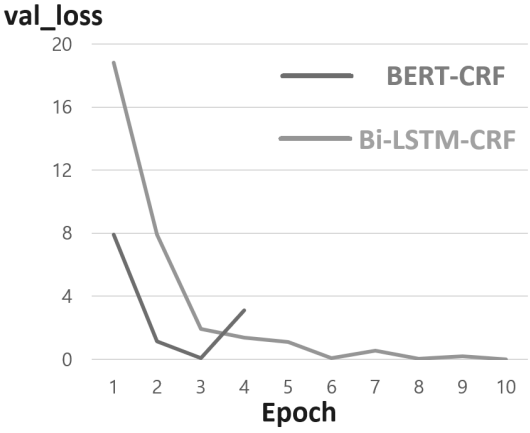


Fig. 2. Validation loss values by epoch

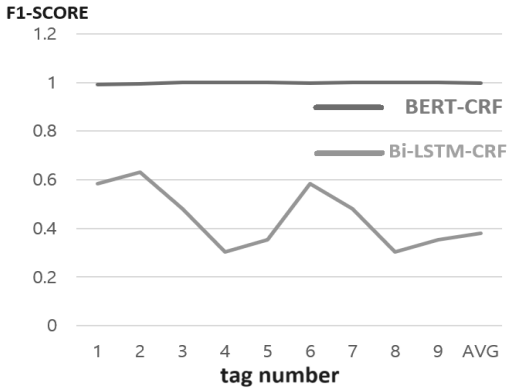


Fig. 3. Performance of f1-score by tag

4.2 성능 평가 결과

Fig.3은 태그별 f1-score 지표이며, x 축의 마지막 weighted avg는 f1-score의 가중치 평균이다. 총 23개의 BIO tag list 중 테스트 데이터에는 x 축의 9개의 BIO tag가 존재하였다. "1.B-DST-IP, 2.B-DST-PORT, 3.B-SRC-IP, 4.B-SRC-PORT, 5.B-START-TIME, 6.I-DST-IP, 7.I-SRC-IP, 8.I-SRC-PORT, 9.I-START-TIME"이 존재하였으며, Bi-LSTM-CRF의 accuracy는 epoch 1회 기준 약 0.89에서 epoch 3회부터 약 0.99로 향상되었으며, BERT-CRF 경우 1회 기준 약 0.99가 나왔다. 그리고 인공지능 알고리즘별 필수 위협 정보 인식률 분석 결과, Bi-LSTM-CRF의 태그별 f1-score 가중치 평균 기준 약 0.44가 나왔으며, 개체명 인식 성능지표가 유난히 낮은 클래스는 PORT 개체명 인식과 TIME 관련 개체명 인식이 f1-score 0.4 이하로 낮은 성능을 보였다. BERT-CRF의 f1-score 가중치 평균 기준 9개 태그 명을 약 0.99 인식하는 결과를 보였다. 이를 통해 다양한 보안장비 로그 데이터에 포함된 위협 정보를 식별하기에는 BERT가 적합하다는 것을 알 수 있다.

Fig.4 는 태그 개수에 따른 성능 평가에 관한 결과이다. 학습시킨 태그 개수가 23개일 때 Bi-LSTM-CRF의 f1-score 가중치 평균은 약 0.44이며, BERT-CRF의 f1-score 가중치 평균은 0.99이다. 학습 데이터의 태그를 47개까지 증가시켜 테스트 진행 결과 Bi-LSTM-CRF의 f1-score 가중치 평균은 0.38이며, BERT-CRF의 f1-score 가중치 평균은 0.43이다. 시험 결과를 통해 태그 개수가 적을수록 위협 정보를 높은 확률로 인식할 수 있다는 것을 알 수

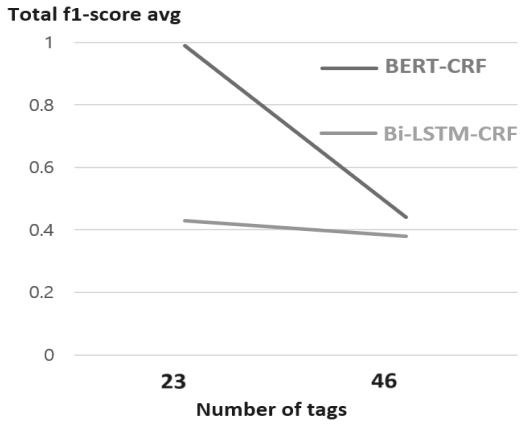


Fig. 4. Performance by tag count

있다. 태그 수가 많을 경우, 테스트 데이터 중 MSG 태그로 인식해야 하는 값을 outside 태그로 잘못 인식하는 등 인식을 제대로 못 하는 경우가 다수 존재하여 성능이 저조했지만, 태그 수가 감소할 경우, 잘못 인식하는 경우가 현저히 감소하여, 위협 정보 인식률 성능 향상에 영향을 미쳤다.

V. 결 론

본 연구에서는 생성형 AI, 빅데이터 처리·분석 기술이 고도화됨에 따라 새로운 보안 위협이 발생하게 되었고, 이를 대응하기 위해 다양한 신규 보안장비가 등장하였다. 신규 보안장비와 기존 보안장비를 혼용하여 사용하다 보니, 위협 정보 표현 포맷이 달라 기존 데이터 분석 기법인 패턴 기반의 위협 정보 식별이 어려워졌다. 이를 해결하기 위해 본 논문에서는 인공지능 기반 개체명 인식 기술을 제안하였다.

위협 정보 선정을 위해 10개의 보안장비가 포함하고 있는 정보를 분석하여 5가지 필수 정보를 선정하였고, 필수 정보를 저장하기 위해 필드명, 필드 값 포맷을 정의하였다. 또한, 개체명 인식을 위해 BIO 태깅 기법을 채택하였고, BIO 태깅을 위한 필수 정보 기반으로 총 23개의 태그 리스트를 정의하였다. 인공지능 알고리즘 선정은 Bi-LSTM과 BERT 두 가지 모델의 개체명 인식 성능을 비교하며, 두 모델의 출력층에. CRF 단계를 추가하여 개체명 인식의 성능을 높이도록 설계하였다.

시험을 위해 부족한 데이터 문제를 해결하기 위해 다양한 업체의 보안장비 원본 로그 데이터를 증식시켰고, train data가 test data 성능 측정에 영향을 주는 "Data leakage" 현상을 방지하기 위해 13종의 보안장비 데이터 27,370개를 BIO 태깅하여 학습/검증 데이터로 활용하고 나머지 1종의 보안장비 2,437개 데이터를 BIO 태깅하여 테스트 데이터로 사용하였다. Bi-LSTM-CRF, BERT-CRF 두 모델의 필수 정보 개체명 인식 결과는 epoch 3회 이상 기준 accuracy는 약 0.99로 비슷한 성능을 보였으나, 그 외에 precision, recall, f1-score 등 다양한 성능지표에서는 약 0.6 이상 큰 차이로 BERT-CRF가 우수한 성능을 보였다. 이는 부족한 학습 데이터 이슈 때문에 Bi-LSTM 모델이 현저히 낮은 성능을 보인다는 것을 알 수 있었다. 그에 반해 BERT 모델은 대규모 데이터셋을 사전에 학습한 모델로 이러한 문제를 극복하기에 유용한 해결책이 된다. 하지만 태그 개수가 증가하거나 보안장비 로그 데이터에 source ip, destination ip뿐만 아니라 sensor ip, translation ip 등 다양한 ip 정보를 학습·테스트 데이터로 활용할 수 있다면, ip 관련 태그 개체명 인식 성능이 현저히 낮아진다. 향후 다양한 보안장비 이벤트를 포함한 양질의 데이터를 수집하여, Bi-LSTM 모델 성능 개선연구를 진행할 예정이다.

또한, 구조가 단순하고 소규모 보안장비를 사용하는 환경에서는 정규표현식 기반의 변환 방법이 인공지능을 활용한 방법보다 변환 속도가 빠르다. 하지만 기존에 없는 유형의 정규표현식이나 복잡한 정보를 담고 있는 보안 로그라면 사용자가 직접 패턴 분석한 뒤, 정규표현식을 작성해야 하므로 개발자 또는 시스템 운영자가 직접 개입해야 한다는 단점이 존재한다. 정규표현식을 이용한 변환 방안과 인공지능을 활용한 변환 방법의 장점을 합치는 프로세스를 정규표현식을 이용한 위협 정보 식별 기술과 인공지능 기반 위협 정보 개체명 인식 기술을 병합한 통합프로세스에 적용해 볼 예정이다.

References

[1] Cisco, "Asia Pacific CISO Benchmark Study," Feb. 2019.

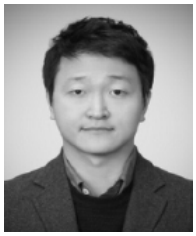
[2] H. K. Kim and K. H. Rhee, "An Analysis of System Log using Regular Expressions," Korea Information Processing Society, 27(1), pp. 154-156, May 2020.

[3] S. Barnum, R. Martin, B. Worrell and I.Kirilov, "The CybOX language specification," The MITRE Corporation, Apr. 2012.
[4] Intruduction to STIX, "STIX" https://oasis-open.github.io/cti-documentation/stix/intro.html, Feb. 2024.
[5] Telecommunications Technology Association, "TTAK.KO-12.0242 Session Information Message Exchange Format," Information and Communication Organization Standard (Korean Standard), Jul. 2014.
[6] Telecommunications Technology Association, "TTAK.KO-12.0279 Security Information Message Exchange Protocol," Korea Communications Standards, Dec. 2015.
[7] Telecommunications Technology Association, "TTAK.KO-12.0256 System Information Message Exchange Format for Security Control," Information and Communication Organization Standard (Korean Standard), Dec. 2014.
[8] Telecommunications Technology Association, "TTAK.KO-12.0229 Extended Intrusion Detection Message Exchange Format," Korea Communications Commission, Dec. 2013.
[9] Y. Liu and D. Zhang, "UniParser: A Unified Log Parser for Heterogeneous Log Data," Proceedings of the ACM WEB Conference 2022, pp. 1893-1901, Apr. 2022.
[10] H. D. Kim and H. S. Lim, "A Named Entity Recognition Model in Criminal Investigation Domain using Pre-trained Language Model," Korea Convergence Society, 13(2), pp. 13-20, Feb. 2022.
[11] J. H. Hyun and H. J. Kim, "Security Operation Implementation through Big Data Analysis by Using Open Source ELK Stack," Journal of Digital Contents Society, 19(1), pp. 181-191, Jan. 2018.
[12] K. S. Ko and I. J. Jo, "Application of Integrated Security Control of Artificial Intelligence Technology and Improvement of Cyber-Threat Response Process," The Journal of the Korea Contents Association, 21(10), pp. 59-66, Oct. 2021.
[13] J. H. Kim and J. Y. Kim, "Comparative analysis of performance of BI-LSTM and GRU algorithm for predicting the number of Covid-19 confirmed cases," Journal of the Korea Institute of Information and Communication Engineering, 26(2), pp. 187-192, Feb. 2022.
[14] S. J. Ko, H. Y. Yun, and D. M. Shin, "Electronic Demand Data Prediction using Bidirectional Long Short Term Memory Networks," Journal of Korea Software Appraisal Association, 14(1), pp. 33-40, Jan. 2018.
[15] S. H. Na and J. W. Min, "Character-Based LSTM CRFs for Named Entity Recognition," Proceedings of KIISE Conference, pp. 792-731, Jun. 2016.
[16] Z. Huang, W. Xu, and K. Yu. "Bidirectional LSTM-CRF models for sequence tagging." arXivpreprint arXiv:1508.01991, Aug. 2015.

〈저자소개〉

김 태 현 (Taehyeon Kim) 정회원
2022년 2월: 건양대학교 정보보호학과 학사 졸업
2021년 9월~현재: 한국인터넷진흥원 주임연구원
2023년 3월~현재: 전남대학교 정보보안융합학과 석사과정
〈관심분야〉 정보보호, 인공지능 보안, 시스템 보안, 네트워크 보안

임 준 형 (Joon-Hyung Lim) 정회원
2001년 2월: 서경대학교 전산정보관리학과 석사 졸업
2000년 10월~현재: 한국인터넷진흥원 인프라보안기술팀 팀장
〈관심분야〉 정보보호, 인공지능, 이동통신 보안, 블록체인



김 태 은 (Taeeun Kim) 정회원
2005년 2월: 백석대학교 정보통신공학부 학사 졸업
2007년 2월: 숭실대학교 컴퓨터공학과 석사 졸업
2013년 7월~현재: 한국인터넷진흥원 책임연구원
〈관심분야〉 정보보안, 네트워크 보안, 융합 보안

엄 익 채 (Ieck-chae Euom) 종신회원
2003년 8월: 전남대학교 컴퓨터정보학부
2015년 2월: 한국과학기술원 소프트웨어대학원 석사 졸업
2019년 2월: 전남대학교 정보보안협동과정 박사 졸업
2019년 10월~현재: 전남대학교 시스템보안연구센터 소장, 데이터사이언스대학원 교수
〈관심분야〉 제어시스템 보안, 스마트그리드 보안, 원자력 보안, 취약점 분석, 차세대인프라보안, 스마트시티 공장 보안, AI기반 이상징후 탐지, 지능형 보안