정 보 보 호 학 회 지
제 31 권 제3호, 2021. 6

악성코드 특징정보(Feature)의 종류 및 시스템 적용 사례 연구

김 병 재*, 한 상 원*, 이 재 광**

요 약

공격자는 공격을 성공적으로 수행하기 위해 악성코드를 주로 사용하며, 방어자는 공격자의 공격이 완성되는 최종 단계 이전에 악성코드를 탐지하여 무력화 할 수 있도록 노력하는 것이 매우 중요하다. 그래서 이를 선제적으로 식별하고 대응하기 위해 인공지능 분석, 연관분석, 프로파일링 등 다양한 분석 기법이 연구되어지고 있다. 이러한 분석 기법들은 사전에 악성코드의 특징을 파악하고 어떤 악성코드 특징정보를 분석할지 선택하는 것이 가장 중요하다. 본 연구에서는 악성코드 특징정보의 종류와 실제 시스템에 적용한 사례에 대해서 살펴보고자 한다.

Ⅰ. 서 론

컴퓨터와 인터넷의 급속한 발전으로 인간의 삶을 편하고 윤택하게 만들어 주었으나 부작용 역시 급증하고 있다. 분산 서비스 거부 공격, 개인정보 유출, 해킹 등은 대부분 악성코드에 의해 발생하고 있다. 매년 침해사고 및 악성코드의 수가 급증하고 있으며, 공격자는 방어 환경을 우회하고 탐지를 회피하기 위해 악성코드를 지속 고도화해 나가고 있다. 또한, 악성코드 오픈소스와 악성코드 자동화 제작 도구 등 악성코드를 쉽게 구매하거나 제작할 수 있어 누구나 쉽게 공격자가 될 수 있다[1]. 그래서 기업 및 기관에서는 급증하는 대량의 악성코드를 효율적으로 분석하고 대응하기 위해 AI 등 다양한 기법을 적용하고 연구하고 있다. 이러한 연구를 위해 악성코드 및 침해사고의 특징을 정확하게 파악하는 것이 수반 되어야하며, 정확성 및 효율성 향상을 위해 어떤 악성코드 특징정보(Feature)를 선정하여 활용할지 결정[2][3]하는 것이 가장 중요하다. 한국인터넷진흥원은 다양한 침해사고 및 악성코드를 분석하면서 악성코드의 수많은 특징 중 중요도가 높은 특징정보를 선별하여 분류하고 이를 분석 및 대응에 활용하고 있다.

본 연구의 2장에서는 다양한 악성코드 특징정보 중 악성코드를 분류하고 분석하는데 활용 가능한 주요 특징정보의 종류 및 지표에 대해 정의하고 3장에서는 특징정보를 실제 침해사고 분석 대응 업무에 적용하여 AI 분석, 시각화 연관 분석을 수행하는 특징정보 기반 분석 시스템에 대해 다룬다. 그리고 4장은 결론 및 향후 연구 방향으로 구성하였다.

Ⅱ. 특징정보 종류

악성코드가 가지는 세부적인 정보의 유형을 6개 카테고리(메타데이터, 정적정보, 동적정보, 네트워크 정보, ATT&CK 프레임워크, 기타정보)로 총 72가지 특징정보로 분류하여 관리하고 있다. 모든 특징정보는 자동 추출 에이전트를 개발하여 추출 및 정규화하고 저장·관리하고 있다.

한국인터넷진흥원 특징정보 분석 관리 시스템에 적용되어 활용하고 있는 특징정보만 기술, 정의하였으며, AI 모델 개발, 프로파일링, 자동 분석에 활용할 수 있다.

2.1. 메타데이터(Metadata)

악성코드 파일 내 기본적인 메타데이터 정보와 PE 헤더 정보가 포함되어 있다.

Hash는 각 함수를 통해 얻어지는 고정된 길이의 데이터이며, 악성코드를 쉽고 빠르게 식별할 때 사용한

* 한국인터넷진흥원 (선임연구원, kimbyeongjae@kisa.or.kr), 한국인터넷진흥원 (선임연구원, hsw89@kisa.or.kr)
** 한국인터넷진흥원 (팀장, leejk@kisa.or.kr)

〔표 1〕 파일 메타데이터 및 프로파일링 정보

Category	Feature
Hash	MD5
	SHA1
	SHA256
	Import Hash
	Full Fuzzyhash
Basic Info	File Size
	Original FileName
	Entropy
	Timestamp
Compile Info	Packer
	Compiler
	Linker
Digital Sign	Digitally Signed Name
	Digitally Signed S/N
MAC Time	Modify Time
	Access Time
	Create Time
Profilling	Attacker Group
	Family

〔표 2〕 PE(Portable Executable) 정보

Category	Feature
Entry	EntryPoint
	EntryPoint Data
	ImageBase
Resource	Name
	RVA
	Size
	Type
	Language
	Country
Characteristics	Image File Characteristics
Rich Header	Rich Header
Section	Section MD5
	Fuzzy Hash(.data)
	Name
	Virtual Address
	Virtual Size
	Raw Address
	Raw Size
	Entropy

다. Basic Information은 악성코드의 기본 정보를 담고 있으며, Compile Information은 악성코드 제작 패커의 종류와 개발 언어 및 컴파일러의 정보를 포함하고 있다. Digital Sign은 악성코드를 서명한 악용 혹은 변조된 인증서에 대한 정보를 포함하고 있다. MAC Time은 악성코드의 생성, 접근, 수정 시간을 통해 변조 여부 및 침투 시간 등을 확인할 수 있고, Profilling은 기 분석된 샘플에 대해 악성코드 유포 조직을 추적하기 위한 공격자 그룹 및 패밀리 정보를 포함하고 있다.

PE파일 헤더에는 대표적으로 시스템 유형, 파일 특성, 섹션 정보 등 중요한 정보가 포함되어 있으며, 특징정보로 활용할 수 있다[4].

그 중 Entry는 악성코드의 시작지점 및 데이터를 포함하고 있으며, Resource와 Section 정보는 각각의 기본 정보를 포함하고 있다. 또한, PE파일 전체를 Fuzzy Hashing하여 유사도를 비교하는 것보다 성능이 우수하다고 알려진 .data section의 Fuzzy Hash[5]도 포함되어 있다. Characteristics와 Rich Header를 통해 파일의 속성 등을 확인할 수 있다.

2.2. 정적정보(Static Information)

개발경로 및 문자열 등 코드 내에서 확인 가능한 정적 정보가 포함되어 있다.

개발 과정에서 파일 내부의 다양한 string이 포함되어 있으며, 특히 개발 환경을 유추할 수 있는 PDB path, 중복 실행 방지를 위한 MutexName 등이 포함될 수 있다. 그리고 쉽고 빠르게 악성코드의 시그니처를 탐지하고 분류할 수 있는 오픈 소스 및 자체 생성 YARA Rule에 탐지된 탐지결과도 포함하여 관리하고 있다.

〔표 3〕 바이너리 정적 분석 정보

Category	Feature
String	PDB Path
	String
	MutexName
Rule	YARA

2.3. 동적정보(Dynamic Information)

샌드박스를 통해 악성코드 실행 시 파일, 레지스트리, 프로세스 등 다양한 행위 정보를 파악할 수 있고, 주요 정보만 추출 및 관리하고 있다.

File Activities는 악성코드 실행 시 파일을 쓰고 읽고 삭제하는 모든 주요 행위를 포함하고 있으며, Registry Activities는 악성코드가 생성하는 레지스트리 키와 값을 포함하고 있다. Process Activities는 특정 프로세스의 생성, 중단 등을 포함하고 있다.

〔표 4〕 샌드박스 동적 분석 정보

Category	Feature
File Activities	FileCreated
	FileDeleted
	FileWritten
	FileRead
Registry Activities	KeyCreated
	KeyValueCreated
Process Activities	ProcessCreated
	ProcessSuspended
	ProcessTerminated
	ShellExecuted

2.4. 네트워크 정보(Network Information)

악성코드 실행 시 접속하거나 파일 및 메모리 내 삽입되어있는 URL/IP 등의 정보가 포함되어 있다.

네트워크 정보는 악성코드가 실행되면서 실제로 연결되는 IP/URL 정보 외에도 바이너리 파일 내부에 삽입되거나 실행 시 메모리에 삽입된 네트워크 정보도 포함하고 있다.

〔표 5〕 동적/정적 네트워크 분석 정보

Category	Feature
IP/URL	Binaries
	Memory
	Connected

2.5. ATT&CK 프레임워크

미국의 비영리기관 MITRE社에서 개발하여 제공하는 ATT&CK 프레임워크는 실제 발생한 사이버 공격 사례를 기반으로 Tactics, Techniques, Common Knowledge를 통해 공격자의의 전략, 전술 등 사이버 공격 행위를 나타낼 수 있다[6][7].

ATT&CK에는 다양한 기술들이 존재하지만 악성코드가 사용하는 기술들은 각 전술별로 Top3에 밀집되어 있는 것을 알 수 있으며[8], 동일한 악성코드 패밀리 또는 유사 침해사고인 경우 그 특징이 더욱 확연하게 나타난다. 그래서 악성코드의 행위 정보를 Mitre社의 전략, 전술별 기술단위로 추출 및 매칭하여 관리하고 있다. 매칭은 샌드박스 분석 결과 및 시그니처 기반으로 하고있으며, ATT&CK Matrix의 업데이트에 따라 주기적으로 관리가 필요하다.

각각의 단계별로 수십개의 기술들이 포함되어 있으며, 악성코드 실행 시 시그니처 매칭을 통해 각각의 기술 사용 여부를 확인한다. 기술의 개수는 본고 작성 시

〔표 6〕 Mitre社의 ATT&CK 프레임워크

Category	Feature
Initial Access	9 Techniques
Execution	10 Techniques
Persistence	17 Techniques
Privilege Escalation	12 Techniques
Defense Evasion	32 Techniques
Credential Access	14 Techniques
Discovery	22 Techniques
Lateral Movement	9 Techniques
Collection	15 Techniques
Command and Control	16 Techniques
Exfiltration	8 Techniques
Impact	13 Techniques

점으로 기술하였으며, 향후 추가되거나 삭제될 수 있다.

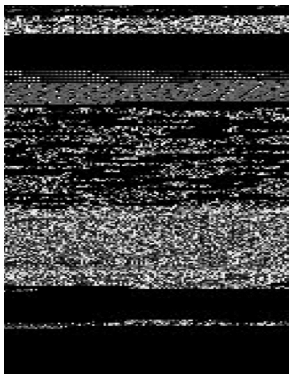
2.6. 기타 정보

악성코드 함수 단위, 이미지처리 등의 기타정보가 포함되어 있다.

PE Image[8]는 악성코드를 0과 1의 이진 문자열을 행렬로 재구성하여 이미지 처리하였으며, API는 악성코드 동작 시 API의 Call Sequence를 벡터화한 값이다. Function Code Block은 악성코드를 실행 함수 단위로 분해한 값이며, Opcode와 Bytecode는 각각 악성코드의 Operation Code와 Byte Code를 나열한 값을 포함한다.

〔표 7〕 기타 분석 정보

Category	Feature
etc	PE Image
	API Sequence
	Function Code Block
	Opcode
	Bytecode



〔그림 1〕 PE Grayscale Image

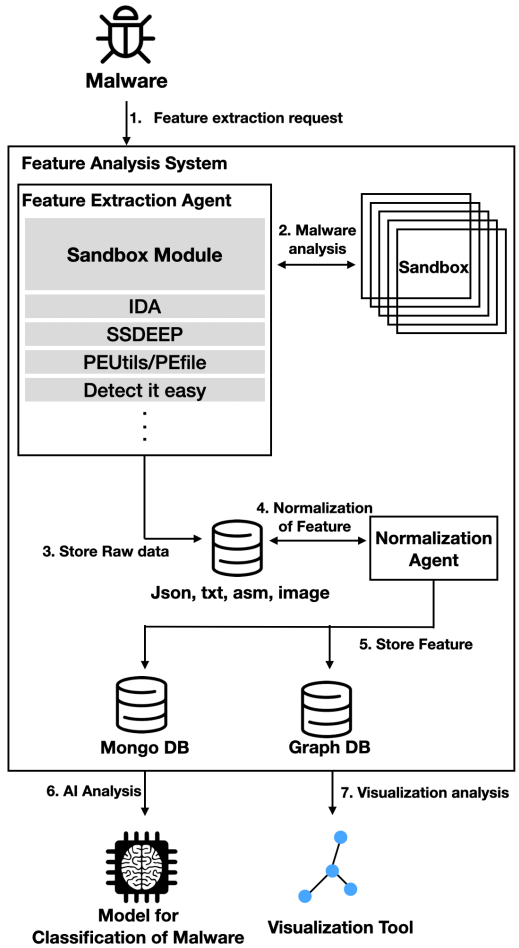
Ⅲ. 활용 사례

2장에서 기술한 악성코드 특징정보를 활용하면 고위험 침해사고 및 악성코드를 식별하고 대응할 수 있다. 한국인터넷진흥원은 신규 유입되는 악성코드 샘플에 대해 식별된 특징정보를 자동으로 추출하고 추출된 특징정보를 활용하여 AI 분석 및 시각화 연관 분석에 적용하였다.

3.1. 특징정보 분석 시스템

악성코드 특징정보를 적용하여 개발한 특징정보 분석 시스템은 그림 2와 같이 신규 악성코드 등록 시 다음과 같은 과정을 통해 악성코드를 분석하고 72개의 특징정보를 추출 및 저장한다.

1. 특징정보 분석 관리 시스템에 악성코드가 등록되면 샌드박스에 동적분석을 요청하고, IDA Python 등 정적 분석 라이브러리를 통해 정적 분석 정보와 메타데이터를 분석한다.
2. 샌드박스를 통해 분석된 동적 분석 정보에서 주요 악성 행위를 선별한다.



〔그림 2〕 특징정보 분석 시스템 구성도

3. 추출된 원시 데이터(메타데이터, 동적/정적 분석 정보)를 추출된 형태(json, txt, asm, image)에 따라 스토리지에 저장한다.
4. 데이터의 중복 제거 및 특징정보 분류를 위해 정규화 에이전트에 특징정보 가공을 요청한다.
5. 가공된 특징정보는 Mongo DB와 Graph DB에 각각 저장된다.
6. 대량의 악성코드 중 랜섬웨어 등 고위험 악성코드를 분류하기 위해 AI 분석을 요청한다.
7. 특징정보간 연관 분석 및 악성코드의 공격 그룹을 식별을 위해 시각화 분석을 요청한다.

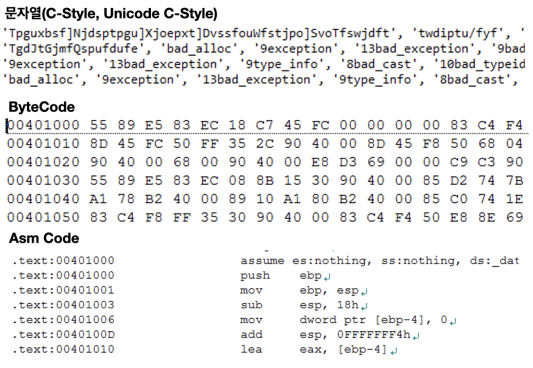
3.1.1. 특징정보 추출

모든 악성코드는 샌드박스와 python 기반 바이너리 분석 에이전트를 통해 분석되고 특징정보가 추출된다.

샌드박스의 동적 분석 결과를 json 형태로 저장하고 파서를 활용하여 file activities, registry activities, process activities를 추출한다.

IDA Python을 통해 악성코드 내부의 문자열(C-Style, Unicode C-Style)과 bytecode 및 opcode를 추출하며, Python의 PEfile을 통해 파일의 메타데이터 정보를 추출한다.

SSDeep 도구를 통해 파일 전체의 fuzzy hash와 .data section의 fuzzy hash를 계산하며 Complie 및 Packer 정보는 Detect it easy 라이브러리를 활용하여 추출한다. 그 외 ATT&CK Matrix 및 프로파일링 정보 등은 동적 분석 정보와 시그니처 기반 탐지를 통해 특징정보를 추출한다.



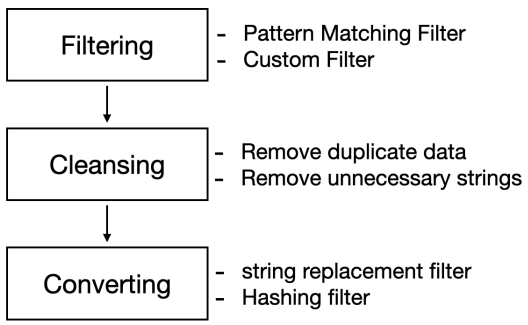
〔그림 3〕 IDA python을 통한 특징정보 추출

3.1.2. 특징정보 정규화

추출된 특징정보는 그림 4와 같이 다음 3단계(Filtering, Cleansing, Converting) 과정을 거쳐 원시 데이터로부터 피처를 추출하고 이를 분석에 적합한 형식으로 가공한다.

1단계는 Filtering 과정은 시그니처 패턴 매칭 필터와 커스텀 필터를 통해 사이즈, 특징정보 내용을 분류하며, 2단계 Cleansing 과정을 통해 동일 악성코드 내 중복 데이터 제거 및 불필요한 문자열을 제거한다. 마지막 3단계 Converting 과정에서 데이터 및 사이즈 일관성을 위해 문자열 치환과 hashing 필터를 적용한다.

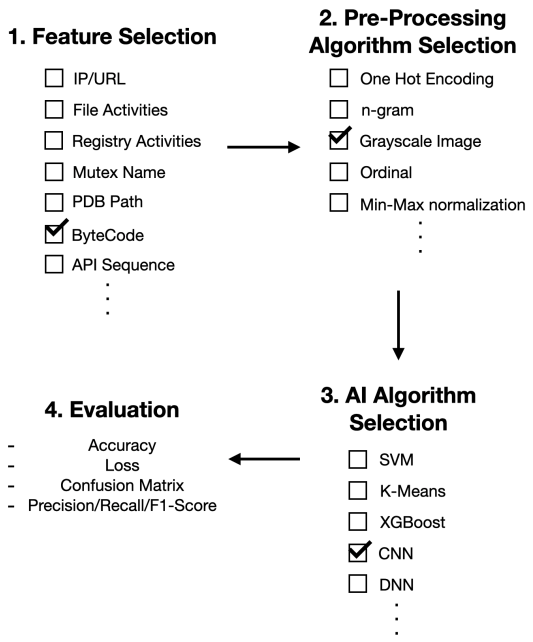
정규화 과정을 통해 가공된 데이터는 모두 몽고DB와 Node/Edge 형태로 그래프DB(Neo4J)에 저장된다.



〔그림 4〕 특징정보 정규화 과정

3.2. AI 분석

고위험 악성코드를 신속하게 분석하고 대응하기 위해 한국인터넷진흥원에서 개발한 CNN AI 분류 모델[9]과 2015년 4월에 종료되고 50개 이상의 연구논문과 데이터 세트가 인용된 Microsoft Malware Classification Challenge(BIG 2015)[10]의 우승팀(“say NOOOOO to overfitttttting”) AI 모델을 시스템에 적용하여 활용하고 있으며 신규 모델을 지속적으로 테스트하고 있다. 자동으로 추출된 특징정보 기반으로 AI 분석 모델에 의해 1차적으로 고위험 악성코드(랜섬웨어, APT 악성코드 등)를 분류할 수 있어 분석 및 대응 시간을 단축할 수 있다. AI 모델 개발에 있어서 특징정보 선택(Feature Selection)이 가장 중요한 요소인데, 앞서 설명한 특징정보를 사전에 추출하여 가공해두면 모델 학습 및 평가가 빠르게 이루어지고 적용될 수 있



〔그림 5〕 특징정보 분석 시스템 AI 학습 과정

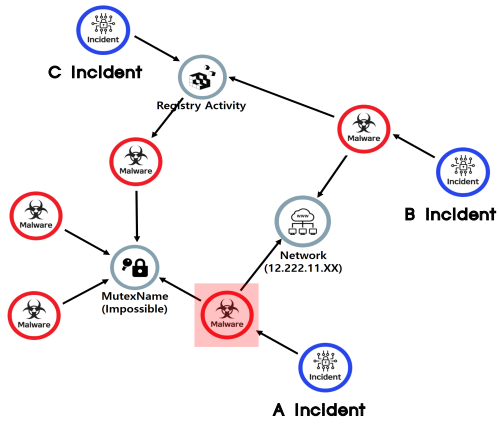
어 용이하다.

사전에 추출된 특징정보를 활용한 AI 학습 과정을 그림 5와 같이 특징정보 목록에서 학습에 사용할 특징정보를 선택하고, 특징정보 형태에 맞는 전처리 알고리즘을 선택한다. 그리고 원하는 알고리즘을 선택하여 AI모델을 학습한다. 학습된 AI 모델을 평가하고 재학습을 반복하여 실무에 적용할지 결정한다.

3.3. 시각화 연관 분석

본 연구 시점인 현재(`21.05)까지 추출된 악성코드 특징정보는 약 1천만개이며, 이를 효율적으로 연관 분석하기 위해 시각화 분석을 적용하였다. 각각의 특징정보를 정규화하여 노드/엣지 기반으로 정보를 생성하여 그래프DB에 저장하고, 저장된 특징정보를 정규화된 데이터 기준에 따라 시각화하였다. 시각화 연관분석을 통해 신규 악성코드의 특징정보가 기존 식별된 고위험 악성코드 특징정보와 유사 및 동일하여 링크가 연결된다면 분석가는 신속하게 신규 악성코드를 파악하고 대응할 수 있다.

그림 6은 신규 악성코드 등록 시 추출된 특징정보 중 뮤텍스이름, 레지스트리 행위, 네트워크 정보가 기



〔그림 6〕 특징정보 시각화 연관 분석

존 분석된 특징정보와 연관성이 확인되어 A사, B사, C사 침해사고가 연관된 침해사고인 것으로 연관 분석결과를 간략하게 표현했다.

Ⅳ. 결론 및 향후 연구 방향

악성코드 특징정보를 식별하고 가공 및 DB에 저장해두는 것은 AI학습, 연관 분석 등을 통해 신규 악성코드의 위협을 식별하고 신속하게 분석 및 대응하는데 있어서 굉장히 중요한 요소이다. 하지만, 기업의 보안담당자 및 데이터 분석가가 악성코드의 세부적인 특징정보에 대해 알기는 쉽지 않다. 그래서 본 연구를 통해 한국인터넷진흥원이 실제로 적용하여 활용하고 있는 악성코드 특징정보에 대해 공유하였고 시스템 적용 사례에 대해 설명하였다.

악성코드의 특징정보는 본 연구에서 기술한 것 외 다양한 특징정보들이 있기 때문에 향후 중요하다고 판단되는 특징정보는 추가적으로 식별하여 적용할 예정이며, 현재 PE 악성코드 외 APK 악성앱, 웹로그 등의 특징정보도 식별하고 다양한 AI 알고리즘에 적용하여 우수한 모델을 개발하는데 중점을 두고 연구중이다.

또한, 수집된 특징정보의 데이터 분석을 통해 EDR 등에 적용할 수 있는 방안에 대해서도 연구한다면 최근 급증하는 랜섬웨어 등의 고위험 악성코드를 탐지하는데 유용하다고 판단된다.

참 고 문 헌

[1] D. Gavrilut, M. Cimpoesu, D. Anton, L. Ciortuz,

"Malware detection using machinelearning, IEEE, 735-741, 2009"
[2] K. Chumachenko, "Machine learning methods for malware detection and classification, XAMK University of Applied Science, 2017
[3] M. Al-Kasassbeh, S. Mohammed, M. Alauthman, A. Almomani, "Feature Selection Using a Machine Learning to classify a Malware", Springer Nature Switzerland, 889-904, 2020
[4] H.S. Anderson, P. Roth,"An open dataset for training static PE malware machine learning models", arXiv preprint arXiv:1804.04637, 2018
[5] 박창욱, 정현지, 서광석, 이상진, "퍼지해시를 이용한 유사 악성코드 분류모델에 관한 연구", 정보보호학회논문지, 1325-1336, 2012.12
[6] MITRE, ATT&CK, "https://attack.mitre.org"
[7] 안명길, 이정륜, "사이버 전투실험 분석을 위한 MITRE ATT&CK 기반의 시스템 구성 및 방법론 연구", 한국컴퓨터정보학회논문지, 31-37, 2020.8
[8] 한국인터넷진흥원, "TTPs#3 공격자의 악성코드 활용 전략 분석", 2020.9
[9] 한국인터넷진흥원, "머신러닝 기반 악성코드 분석 알고리즘 적합성 연구", 2017.8
[10] R. Ronen, M. Radu, C. Feuerstein, E. Yom-Tov, and M. Ahmadi, "Microsoft malware classification challenge", arXiv preprint arXiv:1802.10135, 2018

〈저자소개〉

김 병 재 (Byeongjae Kim)
정회원
2011년 2월 : 서울호서전문학교 사이버해킹보안과 학사
2015년 8월 : 동국대학교 디지털포렌식학과 석사
2016년~현재 : 한국인터넷진흥원 선임연구원
<관심분야> 악성코드 분석, 침해사고 대응, AI 보안, 사이버 위협 프로파일링

한 상 원 (Sangwon Han)
2014년 2월 : 순천향대학교 정보보호학과 학사
2013년~현재 : 한국인터넷진흥원 선임연구원
<관심분야> 악성코드 분석, 침해사고 대응, AI 보안, 사이버 위협 프로파일링

이 재 광 (JaeKwang Lee)
2007년 2월 : 서울대학교 수학과 석사 졸업
2010년 2월~현재 : 한국인터넷진흥원 인터넷침해대응센터 근무(현 종합분석팀장)
<관심분야> 포렌식, 침해사고 조사 기법, 데이터 프로파일링