정 보 보 호 학 회 지
제 29 권 제6호, 2019. 12

OSINT기반의 활용 가능한 사이버 위협 인텔리전스 생성을 위한 위협 정보 수집 시스템

김 경 한*, 이 슬 기*, 김 병 익*, 박 순 태*

요 약

2018년까지 알려진 표적공격 그룹은 꾸준히 증가하여 현재 155개로 2016년 대비 39개가 증가하였고, 침해사고의 평균 체류시간(dwell-time)은 2016년 172일에서 2018년 204일로 32일이 증가하였다. 점점 다양해지고 심화되고 있는 APT(Advanced Persistent Threat)공격에 대응하기 위하여 국내외 기업들의 사이버 위협 인텔리전스(CTI; Cyber Threat Intelligence) 활용이 증가하고 있는 추세이다.

현재 KISA에서는 글로벌 동향에 발맞춰 CTI를 활용할 수 있는 시스템을 개발 중에 있다. 본 논문에서는 효율적인 CTI 활용을 위한 OSINT(Open Source Intelligence)기반 사이버 위협 정보 수집 및 연관관계 표현 시스템을 소개하고자 한다.

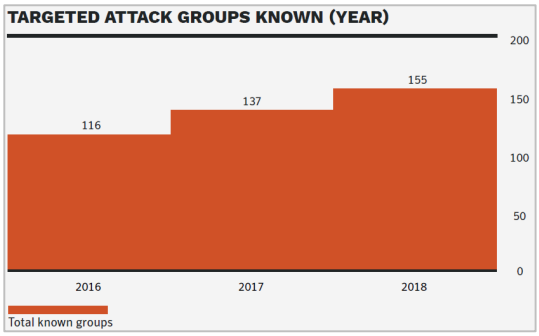
Ⅰ. 서 론

글로벌 보안회사인 Symantec의 2019 사이버 보안 위협 동향 보고서에 의하면 [그림 1]과 같이 2018년까지 알려진 표적공격 그룹은 현재 155개로 2016년 116개 대비 39개 그룹이 증가하였다. 이들 중 제로데이 공격과 같이 최신기법을 사용하는 그룹은 23%에 불과하며, 이는 표적공격 그룹들이 기존의 공격 기법 및 도구들을 재활용한다는 것을 의미한다[1].

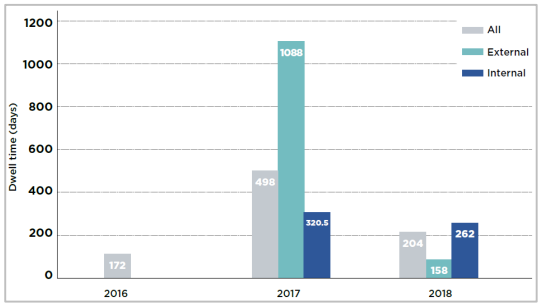
또한, 전 세계적으로 사이버 침해사고가 지속적으로 발생하고 있으며, [그림 2]와 같이 아시아 태평양 지역의 침해사고 평균 체류시간은 2016년 172일에서 204일로 오히려 32일이 증가하였다. 이는 사이버 위협이 보다 집요해지고 고도화되고 있음을 의미한다.

이러한 상황에서 유사/변종 공격에 효과적으로 대응하기 위해서는 기존에 단편적으로 수집 및 공유되었던 데이터들을 하나의 시스템에 수집하여 연관성을 파악하고 유사성을 분석하는 사이버 위협 인텔리전스 기술이 필요하다. 일례로 SANS의 2019 CTI 보고서에 의하면, 금융, 공공, IT 등 다양한 분야에서 종사하는 보안 담당자 중 80.8%가 CTI 활용이 효과가 있다고 응답했다[2,3].

CTI를 효과적으로 활용하려면 정보 공유가 필수적이고, CTI활용의 핵심은 이를 활용하려는 기관의 요구



(그림 1) 알려진 표적 공격 그룹 수[1]



(그림 2) 아시아 태평양지역 침해사고 평균 체류시간(dwell-time)[2]

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00158, 국가 차원의 침해사고 대응을 위한 사이버 위협 인텔리전스 분석(CTI) 및 정보 공유 기술 개발)

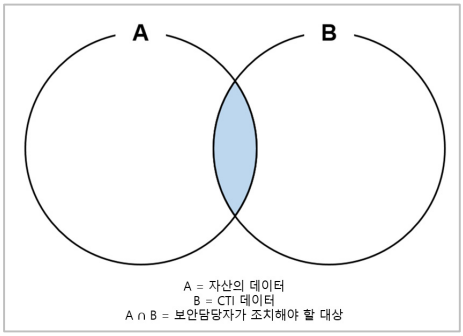
* 한국인터넷진흥원({kookie, sglee, kbi1983, stpark12}@kisa.or.kr)

사항을 정의하여 [그림 3]과 같이 보호대상 자산과 CTI데이터의 연관성을 식별하여 보호조치 대상 및 내용을 도출하는 것이다.

그러나, 국내에서는 국가기관 및 정보공유분석센터(ISAC; Information Sharing & Analysis Center)를 중심으로 제한적인 형태의 공유가 이루어지고 있다. 특히, 공유되는 정보들 중 대부분이 기업의 비밀 및 민감정보 등을 포함하고 있다는 이유로 침해지표(IoC; Indicator of Compromise)와 같이 정제된 단편적인 정보들만 공유되고 있다.

또한, 국내 대다수의 기관 및 기업들은 자산에 대한 정보조차 제대로 식별이 어려운 실정이며, CTI 데이터를 활용한다 하여도 단편적인 정보들을 어떻게 활용할 것인가에 대한 노하우가 부족하다. [표 1]은 자산과 CTI 데이터 연관성 분석 예시를 보여준다.

본 논문에서는 위에서 제시한 한계점들을 해결하기 위하여 현재 KISA에서 개발중인 OSINT를 활용한 사이버 위협 정보 수집 및 연관관계 표현 시스템을 소개하고자 한다.



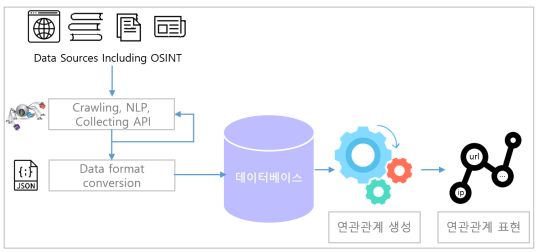
(그림 3) 연관성 분석 예시

〔표 1〕 CTI 데이터 활용 예시

CTI 데이터	자산 데이터	조치 내용
TLS 1.1에서 RC4 사용	구 시스템에서 RC4 사용	구 시스템(XP의 익스플로러) 유지를 위하여 일정 기간 권고 후 RC4 삭제
DELETE 메소드를 통하여 REST API 사용 가능	DELETE 메소드 사용 X	해당 메소드 제거 혹은 IPS, WAF등에서 메소드 제한
FTP 포트 기본 설정 되어 있음	서버에 FTP가 설치되어 있지 않음	서버 조치 X

Ⅱ. OSINT를 활용한 사이버 위협 정보 수집 및 연관관계 표현 시스템

현재 KISA에서 개발중인 사이버 위협정보 수집 및 연관관계 표현 시스템은 1장에서 제시된 한계점을 극복함으로써 활용가능한 CTI를 생성하는것을 목표로 하며, 그 정보를 시각적으로 표현해주는 시스템이다. [그림 4]는 현재 KISA에서 개발중인 시스템(이하 '위협정보 수집 시스템')의 개요를 나타낸다.

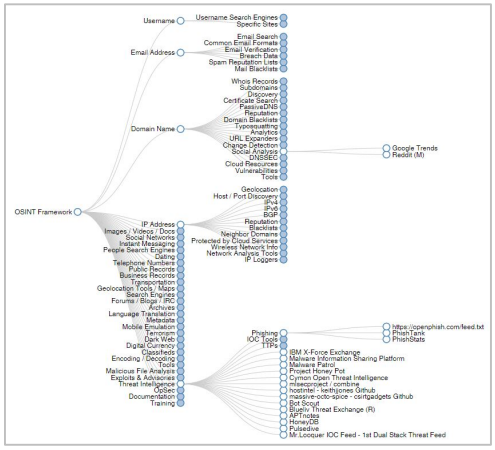


(그림 4) 위협정보 수집 시스템 개요

2.1. OSINT(Open Source Intelligence) 수집 채널

OSINT란 공개된 출처에서 수집할 수 있는 정보들을 지칭하는 말이나, 본 논문에서는 [그림 5]와 같은 유형의 정보들로 사이버 보안 분야에서의 TIS(Threat Intelligence Service) 및 TIP(Threat Intelligence Pltform)등에서 제공되는 정보들을 일컫는다.

'위협정보 수집 시스템'에서는 [표 2]와 같이 24개의 수집 채널들로부터 OSINT를 수집하고 있다.



(그림 5) OSINT 프레임워크〔4〕

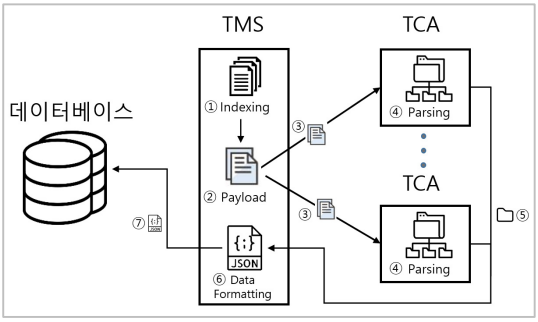
〔표 2〕 OSINT 수집 채널

위협정보 수집 채널	설명
MC-Finder	KISA 악성 컨텐츠 탐지 시스템 (악성코드 유포지, 경유지 정보 등)
상황전파문	KISA 위협정보 공유 시스템
C-TAS	KISA 침해사고 정보공유시스템
MISP	OSINT 공유 플랫폼
Sentinel	KISA 악성코드 유사도 분석시스템
Exploit-DB	취약점 및 Exploit Code
NVD	취약점 정보
DNSBL	DNS 블랙리스트 목록
VirusShare	악성코드 바이너리 및 분석 정보
Zone-h	해킹된 도메인 정보
malwaredomainlist.com	악성 도메인 정보
ThreatCrowd	악성코드 해시, domain, ip 정보 등
EmergingThreat	IP 블랙리스트 정보
TalosIntelligence	취약점 분석 보고서 등
bambenek	침해자원(IP/도메인) 정보
Anti-Hacker-Alliance	침해자원 정보(도메인 활성상태, 연관 IP 등)
Cymon	위협 보고서 등
OpenPhish	피싱사이트 관련 정보
OTX	공개 위협 인텔리전스 커뮤니티
Whois	도메인 등록 정보
IP2Location	IP 지역정보
VirusTotal	정적/동적 악성코드 분석 정보
ThreatMiner	침해사고 분석 및 위협 보고서 등
DNS/PTR	DNS 레코드 정보

2.2. 사이버 위협 정보 수집 프레임워크

현대사회는 정보의 홍수라 불릴만큼 매일 수많은 정보들이 생성되고 있으며, OSINT를 통해 제공되는 정보들 또한 그 양이 매우 방대하고, 형태 역시 상이하다. 이러한 다양한 OSINT를 효율적으로 수집하기 위하여 '위협정보 수집 시스템'은 TMS(Threat Management Server)와 TCA(Threat Collecting Agent)로 구성된다. TMS는 수집 채널별로 수집할 정보량에 따라 TCA를 동적으로 할당하여 작업량을 조절하고, TCA로부터 수집된 정보들을 사전정의된 DB 저장구조에 맞게 변환하여 관리한다. TCA는 TMS로부터 전달받은 환경설정 정보를 기반으로 수집 환경을 구성하여 수집 채널로부터 정보를 수집한다.

'위협정보 수집 시스템'의 수집 과정은 크게 세 단계로 구분되는데, 첫 번째 단계는 작업 채널 및 순서 선정(Indexing) 단계로, 24개의 채널들 중에서 마지막 정보 수집 이후 새로운 정보가 존재하는 채널들을 대상으로 하여 작업량 및 우선순위를 선정한다. 두 번째 단계는 작업 분할(Payload)단계로 Indexing과정에서 산출된 작업량을 기반으로 TCA를 동적으로 할당한다. 세 번째 단계는 데이터 파싱(Parsing)단계로 대상 수집 채널들로부터 데이터를 파싱하고, 파싱된 데이터는 TMS로 전달되어 DB에 저장된다. [그림 6]은 사이버 위협 정보 수집 프레임워크를 나타낸다.



(그림 6) 사이버 위협 정보 수집 프레임워크

2.2.1. 사이버 위협 정보 이력 관리

일반적으로 OSINT 채널을 통하여 제공되는 정보들은 블랙리스트 IP 주소 및 평판정보, 도메인, URL, Hash등과 같은 침해지표로 단편적인 정보들이다. 이러한 정보들은 대부분이 침해사고 분석 이후에 공유되는 과거정보이므로 [그림 7]과 같이 정보를 제공하는 채널별로 해당 사이버 위협 정보의 수집/업데이트 시간, 분류 기준, 해당 정보의 유효성 등 이력이 상이한 경우가 존재한다. 특히, IP 주소, 도메인, DNS 등의 정보는 과거 특정 시점에 침해사고에 활용되었을지라도, 현재는 보안패치 등을 통하여 더 이상 취약한 상태가 아닐 수도 있다. 그러나, 이러한 정보는 일반적인 OSINT 채널을 통하여 수집하기 어렵기 때문에 '위협정보 수집 시스템'에서는 현재 한국인터넷진흥원에서 운영 중인 C-TAS(Cyber Threat Analysis and Sharing System)를 통해 수집한다.

DOWNLOAD	MD5	fdc07234b5d567ea9d6ef9f62f94204b
	SHA1	d884136722b4199c09454f426ca2b939afc07c05
	SHA256	c5dc20694ce11f1c0304eda9275f859c61f43013b0913c6cdbf83df160e0a0f6
SSDeep		192:uqazjUb5nZuHFCnQjxn5Q/hnQleMNnwLlnQOkEnt/jnQTbnRnQOCJVevo7NtLFo1:nmQ/a9ygcQCg
Size		28,110 bytes
File Type		HTML document, Non-ISO extended-ASCII text, with very long lines, with LF, NEL line terminators
Detections		Ad-Aware = Trojan.JS.Redirector.BVR ALYac = Trojan.JS.Redirector.BVR Antiy-AVL = Trojan/JS.Redirector.qa Arcabit = Trojan.JS.Redirector.BVR Avast = HTML:Script-inf [Susp] AVG = HTML:Script-inf [Susp] Avira = HTML/Infected.WebPage.Gen2 Baidu = JS.Trojan.Iframe.x BitDefender = Trojan.JS.Redirector.BVR CAT-QuickHeal = JS/Iframe.AE ClamAV = Html.Trojan.Iframe-6736972-0 Comodo = TrojWare.JS.Agent.SCP@4nwmcf Cyren = JS/Redir.HS DrWeb = HTML.BadLink.1 Emsisoft = Trojan.JS.Redirector.BVR (B) ESET-NOD32 = HTML/ScrInject.B F-Prot = JS/Redir.HS F-Secure = Malware.HTML/Infected.WebPage.Gen2 FireEye = Trojan.JS.Redirector.BVR Fortinet = JS/Redirector.QA!tr GData = HTML.Trojan.IFrame.AM Ikarus = Trojan.JS.IFrame K7AntiVirus = Trojan (0053d33a1) K7GW = Trojan (0053d33a1) MAX = malware (ai score=87) MaxSecure = Trojan.JS.IFrame.ae McAfee-GW-Edition = JS/Iframe.AE McAfee = JS/Iframe.AE Microsoft = Trojan:JS/Iframe.AE MicroWorld-eScan = Trojan.JS.Redirector.BVR NANO-Antivirus = Trojan.Html.Iframe.dczskt Rising = Trojan.Redir@URL!1.A089 (CLASSIC) Symantec = Trojan.Webkit!html Tencent = Html.Win32.Script.504173 VIPRE = Trojan.JS.IFrame.ae (v)
ExIF Data		ContentType = text/html; charset=windows-1251 FileSize = 27 kB FileType = HTML FileTypeExtension = html MIMEType = text/html Title = OsCommerce - ??????? ??????
VirusTotal Report submitted 2019-11-17 15:01:05 UTC		
VirusShare info last updated 2019-11-28 06:55:25 UTC		

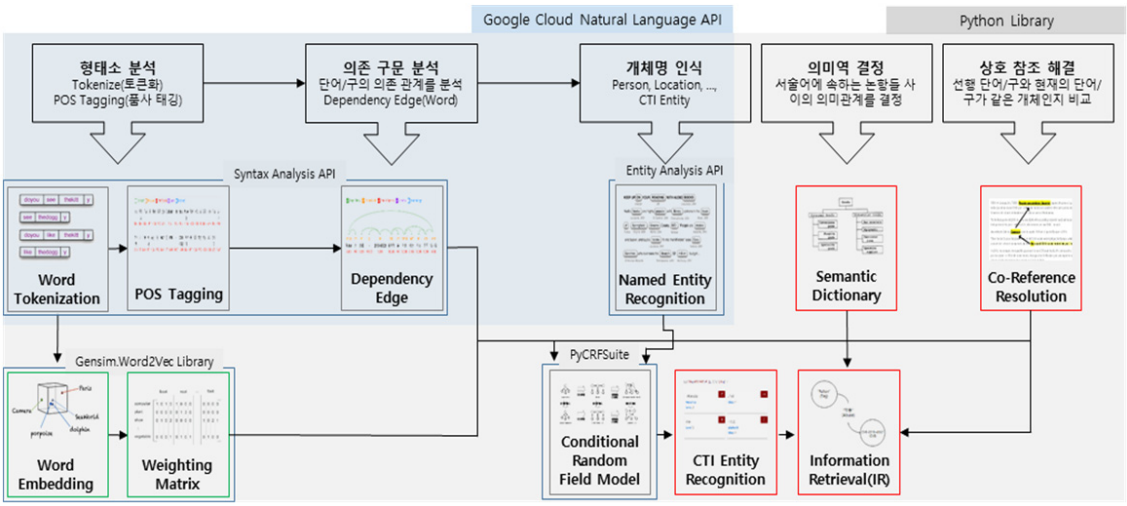
(그림 7) 서로 상이한 수집 이력을 갖는 위협 정보[5]

OSINT 기반의 CTI를 활용하려면 수집된 정보의 이력관리가 매우 중요하다. '사이버 위협정보 수집 시스템'은 재귀적 조회를 통하여OSINT채널을 통해 제공되는 정보들 간의 수집이력을 관리한다. 먼저, 침해지표와 같은 단편적인 위협 데이터가 TCA를 통해 수집되면, TMS는 해당 정보를 기반으로 하여 VirusTotal, OTX, C-TAS 등의 채널에 쿼리를 보내고, 침해지표와 관련된 연관정보들을 DB에 저장한다. 또한, 수집이력 관리를 위하여 주기적으로 DB에 저장되어 있는 정보들을 기반으로 연관정보를 제공하는 채널들에 쿼리를 보내 위협 정보들 간의 싱크를 유지한다.

2.2.2. 자연어 처리를 이용한 침해사고 연관정보 수집

2.2.1절과 같이 사이버 위협 정보의 이력 관리를 함으로써 OSINT를 통해 제공되는 정보들의 한계점을 극복하더라도, 활용 가능한 CTI를 생성하기 위해서는 침해사고에 이용되었던 전략, 기술, 절차를 지칭하는 TTPs(Tactics, Techniques, Procedures)정보가 필수적이다. 해당 정보들은 일반적으로 국내외 보안 벤더사의 침해사고/위협 분석 보고서 또는 TIS를 통해 획득할 수 있다. [그림 8]은 각종 침해사고/위협 분석 보고서로부터 TTPs와 같은 침해사고 연관정보를 자연어 처리 모듈을 이용하여 수집하는 프로세스를 나타낸다.

침해사고/위협 보고서는 일반적으로 pdf, word, hwp의 형태로 공유된다. 이 중 pdf의 경우 이미지 형태로 공유되는 것들도 존재하는데, 이러한 유형의 pdf 파일로부터 pdfminer와 같은 오픈 라이브러리를 통해 텍스트를 추출하는 것에는 한계가 있다. 또한, 보고서 내의 사이버 위협 정보들은 단순 텍스트뿐만 아니라 그림, 표 등 여러 가지가 형태로 존재하기 때문에 '위협정보 수집 시스템'에서는 온전한 정보를 추출하기 위해 OCR(Optical Character Recognition)기능을 제공하는 Google Cloud Vision API를 통해 정보를 추출한다.

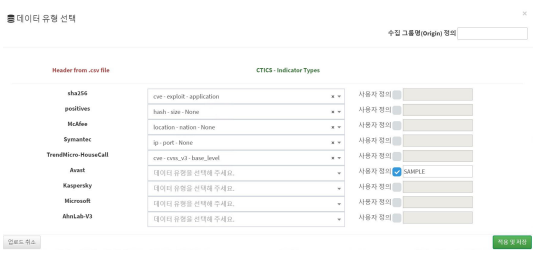


(그림 8) 자연어 처리를 이용한 침해사고 연관정보 수집 프로세스

보고서로부터 텍스트를 추출한 후에는 각각의 사이버 위협 정보들이 어떠한 유형의 정보인지 CRF (Conditional Random Field)를 통하여 개체명 인식을 진행한다. 개체명 인식과 더불어 보고서 내에서 존재하는 동일 위협 정보를 지칭하는 다른 단어(예를 들어, 피해IP와 17x.xxx.xx.54)를 식별하여 상호 참조 문제를 해결한다. 또한, 보고서로부터 TTPs를 추출하기 위하여 품사 태깅(POS; Part of Speech) 정보 및 의존 구문 분석 결과를 기반으로 하여 특정 개체명이 어떠한 서술어에 속하는지 의존관계를 식별한다. 의존관계가 식별된 위협 정보는 해당 정보가 침해를 받은 IP인지, 유포지 IP인지 등과 같이 나뉘게 된다.

2.2.3. 보안 로그 데이터 연동

CTI를 효과적으로 활용하기 위해서는 기관 및 기업에서 운용하고 있는 여러 보안 장비들의 로그로부터 현재 어떠한 유형의 위협이 존재하는지, 해당 위협은 어떠한 유형의 CTI와 연관이 있는지를 식별하는 것이 필요하다. 그러기 위해서는 기업이 보관하고 있는 여러 가지 유형의 보안로그와 CTI 데이터를 비교분석 할 수 있어야 한다. '위협정보 수집 시스템'은 기업의 보안로그 데이터를 csv형태로 입력받아 사전 정의된 데이터 유형에 맞춰 DB에 저장할 수 있도록 한다. 만일, 사전 정의된 데이터 유형에 맞지 않는 데이터가 존재할 경우 [그림 9]와 같이 사용자 정의를 통하여 새로운 유형을 정의할 수 있다.



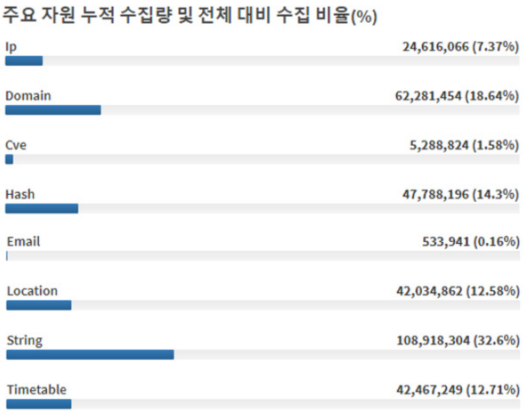
(그림 9) 보안로그 데이터 유형 정의 화면

Ⅲ. 결 론

전 세계적으로 사이버 공격이 점점 더 지능화되고 고도화 되는 가운데 위협 대응 비용을 절감하고 신속하게 대응하기 위하여 CTI 활용 요구가 증가하고 있다. 본 논문에서는 현재 KISA에서 개발 중인 활용 가능한 CTI 생성을 위한 사이버 위협정보 수집 시스템을 소개하였다.

해당 시스템을 통하여 현재까지 수집된 사이버 위협 정보 및 연관정보는 총 769,357,338건 이며, IP, Domain, CVE, Email, Hash 등과 같은 주요 사이버 위협 정보 유형별 수집량은 [그림 10]과 같다.

본 논문에서 제시하는 시스템을 통하여 보다 용이하게 사이버 위협 정보를 수집하여 활용가능한 CTI를 생성하여 산업 전반에 걸쳐 활용함으로써 진일보하는 사이버 위협 대응에 기여하기를 기대한다.



(그림 10) 주요 사이버 위협 정보 유형별 수집량

참 고 문 헌

[1] ISTR Volume 24 2019, https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf, Symantec

[2] M-Trends 2019, https://www.fireeye.com/current-threats/annual-threat-report.html, FireEye

[3] The Evolution of Cyber Threat Intelligence(CTI): 2019 SANS CTI Survey, sans.org/reading-room/whitepapers/threats/paper/38790, SANS

[4] OSINT Framework, https://osintframework.com/

[5] https://virusshare.com/

[6] 김병익, 김낙현, 이슬기, 조혜선, 박준형, "기계학습 기반의 잠재적 사이버 위협 자동 분석 시스템", 대한전자공학회 하계학술대회, pp.368-371, 2018

[7] 이슬기, 조혜선, 김낙현, 김병익, 박준형, "토픽 모

델링 기반 사이버 위협정보 분류방안", 한국인터넷정보학회 추계학술발표대회, 19(2), pp.225-226, 2018

[8] 임원식, 윤명근, 조학수, "KOSIGN:정보보호제품 관점의 사이버 위협정보 공유 체계", 한국정보보호학회학회지, 28(2), pp.20-26, 2018

[9] U. Noor, Z. Anwar, U. Noor, Z. Anwar and Z. Rashid, "An Association Rule Mining-Based Framework for Profiling Regularities in Tactics Techniques and Procedures of Cyber Threat Actors," 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), 2018, pp.1-6, 2018

[10] 김병익, 이슬기, 김경한, 박순태, "사이버 공격 확산 방지 및 신속한 대응을 위한 사이버 위협 인텔리전스 분석 기술", 한국정보처리학회 추계학술발표대회, 26(2), pp.420-423, 2019

[11] 박순태, 김병익, 이슬기, "사이버위협인텔리전스(CTI) 기술을 활용한 공격자 프로파일링 및 대응", 국방과보안, 1(1), pp.131-154, 2019

〈저자소개〉

김 경 한 (KyeongHan Kim)
정회원
2015년 9월 : 순천향대학교 정보보호학과 졸업
2017년 9월 : 순천향대학교 정보보호학과 석사
2017년 9월~현재 : 한국인터넷진흥원 주임연구원
<관심분야> 자연어 처리, 인공지능 알고리즘, 국제 표준, 사이버 위협 프로파일링

이 슬 기 (Seulgi Lee)
정회원
2013년 2월 : 충남대학교 컴퓨터공학과 졸업
2019년~현재 : 고려대학교 빅데이터응용및보안학과 석사과정
2012년 10월~현재 : 한국인터넷진흥원 선임연구원
<관심분야> 네트워크 보안, AI 보안, 소프트웨어 보안

김 병 익 (Byungik Kim)
정회원
2010년 2월 : 아주대학교 정보및컴퓨터공학과 졸업
2018년~현재 : 을지대학교 의료정보보호학과 석사과정
2010년 7월~현재 : 한국인터넷진흥원 선임연구원
<관심분야> 시스템보안, 의료보안, 위협정보연관분석

박 순 태 (SoonTai Park)
정회원
1992년 2월 : 단국대학교 전자계산학과 졸업
1998년 8월 : 국민대학교 정보과학대학원 정보통신학과 석사
2010년 8월 : 전남대학교 대학원 정보보안협동과정 박사
2000년 4월~현재 : 한국인터넷진흥원 팀장
<관심분야> IT보안성 평가, 정보보호 인력 양성, 정보통신기반보호, 조직 정보보안/개인정보보호 실무, 정보보호 R&D