https://doi.org/10.33778/kcsa.2019.19.2.113

공개출처정보를 활용한 사이버공격 데이터베이스 구축방안 연구★

신 규 용*, 유 진 철**, 한 창 희**, 김 경 민**, 강 성 록***, 문 미 남****, 이 종 관*****

요 약

인터넷과 정보통신기술의 발달로 매일 대량의 공개출처정보(Open Source Intelligence, OSINT)가 발생하고 있다. 최근에는 전체 정보의 95%가 공개출처정보에서 나온다고 할 정도로 공개출처의 활용도가 높아졌다. 이러한 공개출처정보는 잘 정제되어 활용된다면 매우 효과적인 고가치 정보로 활용될 수 있다. 일례로 ISVG나 START 프로그램은 테러나 범죄와 관련된 공개출처정보를 수집해 테러리스트 색출이나 범죄예방에 활용해 많은 효과를 거두고 있다. 하지만 사이버공격과 관련된 공개출처정보는 기존의 테러나 범죄와 관련된 공개출처정보와는 달리 공격자, 공격목적, 피해범위 등을 명확히 식별하기 어렵고, 자료 자체가 상대적으로 정형화되지 않았다는 특징이 있다. 이러한 이유 때문에 공개출처정보를 활용해 사이버공격에 대한 데이터베이스(Database, DB)를 구축하고 활용하기 위해서는 기존의 방식과는 전혀 다른 새로운 접근방식이 요구된다. 따라서 본 논문에서는 공개출처정보를 활용해 사이버공격 데이터베이스를 구축하는 방법론을 제시하고 향후 활용방안에 대해 토의하고자 한다.

A Study on Building a Cyber Attack Database using Open Source Intelligence (OSINT)

Kyuyong Shin*, Jincheol Yoo**, Changhee Han**, Kyoung Min Kim**, Sungrok Kang***, Minam Moon****, Jongkwan Lee*****

ABSTRACT

With the development of the Internet and Information Communication Technology, there has been an increase in the amount of Open Source Intelligence(OSINT). OSINT can be highly effective, if well refined and utilized. Recently, it has been assumed that almost 95% of all information comes from public sources and the utilization of open sources has sharply increased. The ISVG and START programs, for example, collect information about open sources related to terrorism or crime, effectively used to detect terrorists and prevent crime. The open source information related to the cyber attacks is, however, quite different from that in terrorism (or crime) in that it is difficult to clearly identify the attacker, the purpose of attack, and the range of damage. In addition, the data itself of cyber attacks is relatively unstructured. So, a totally new approach is required to establish and utilize an OSINT database for cyber attacks, which is proposed in this paper.

Key words : Open Source Intelligence, OSINT, Cyber Attacks, OSINT Database

접수일(2019년 4월 4일), 수정일(1차: 2019년 6월 24일), 게재확정일(2019년 6월 30일)

★ 본 논문은 2018년 국군사이버사령부(11-1290000-000742-01)와 2019년 육사 사이버전연구센터 지원에 의해 연구되었음.

* 육군사관학교 컴퓨터과학과(주저자)
** 육군사관학교 컴퓨터과학과
*** 육군사관학교 심리경영학과
**** 육군사관학교 수학과
***** 육군사관학교 컴퓨터과학과(교신저자)

1. 서 론

지식(knowledge)은 가공의 목적에 따라 자료(data), 첩보(information), 정보(intelligence)로 구분된다[1]. 이때 데이터란 이론을 세우는 기초가 되는 가장 기본적인 사실 혹은 바탕이 되는 자료로 특정 목적에 의해 평가되거나 가공되지 않은 단순한 사실을 의미한다. 생정보(生情報)라고도 불리는 첩보는 목적성을 가지고 의도적으로 수집된 사실로서 단편적이고 불규칙적이며 불확실성을 내포한다. 반면 정보(intelligence)는 특정한 상황에서 정책적 목적을 가지고 첩보가 분석되고 평가되어져 만들어진 체계화된 지식으로 적시성과 목적성을 갖는다. 즉, 일반적으로 수집된 데이터가 가공되어 첩보가 되고, 이 첩보는 다시 사용되는 목적에 맞춰 분석되어 정보로 재생산되는 것이다.

일반적으로 정보기관에서 정보를 수집하는 방법은 비밀정보수집(covert collection)과 공개정보수집(overt collection)으로 나뉜다[2]. 이때 비밀정보수집은 정보기관의 전통적인 정보수집 방법으로 인간정보(HUMINT : HUMan INTelligence), 영상정보(IMINT : IMage INTelligence), 신호정보(SIGINT : SIGnal INTelligence), 계측정보(MASINT : Measure And Signature INTelligence), 그리고 기술정보(TECHINT : TECHnical INTelligence) 등을 포함한다. 이러한 비밀정보수집은 정보수집을 위한 과도한 비용, 비도덕적인 정보활동으로 인한 윤리문제, 객관성 및 정보 적시성의 결여 등의 문제점을 지니고 있다[3]. 반면 공개정보수집은 공공에서 접근할 수 있는 모든 공개정보 출처(신문, 방송, 간행물, 온라인 매체 등)를 이용하여 정보를 획득하는 방법으로, 공개출처정보(OSINT)가 대표적이다[4]. 공개출처정보는 누구나 접근이 가능한 공개정보를 이용하여 정보활동을 하는 것이기 때문에 비밀정보수집 방법이 가지는 과도한 비용문제와 윤리문제를 해결할 수 있고, 보다 객관적이고 정확한 분석이 가능하며, 언제 어디서든지 즉시 활용할 수 있어 정보활동의 최적의 대안으로 각광받고 있다.

지금까지 공개출처정보는 주로 테러리즘과 범죄분야에 많이 활용되어 왔으며[5], ISVG[6], START[7], 그리고 IDW[8] 등이 그 대표적인 예이다. 먼저 ISVG(Institute for the Study of Violent Groups)는 폭력적인 극단주의와 범죄에 대한 지식 향상을 위해 데이터 수집 및 정보서비스를 제공하는 연구센터로 공개출처정보를 모아서 고유한 데이터베이스 구조를 적용하여 테러집단에 대한 정보를 제공한다. START(Study of Terrorism and Responses to Terrorism)는 미국과 전 세계에 일어나는 테러의 원인과 결과에 대해 연구 및 교육을 실시하여 국가 안보정책 수립에 필요한 기초 자료를 제공한다. 마지막으로 IDW(Investigative Data Warehouse)는 FBI에서 운영하는 중앙 집중식 데이터베이스로 범죄수사를 위해 법 집행기관, 범죄 집행 네트워크, 공공 범죄기록 등에 대한 정보를 데이터베이스로 구축해 자료로 제공한다. 공개출처정보를 활용한 ISVG, START, IDW는 테러공격과 같은 폭력적 위협을 예측하고 대응책을 마련하는데 매우 효과적인 것으로 알려져 있다[5].

이와 같이 테러리즘과 범죄분야에 대한 공개출처정보의 높은 활용성에도 불구하고, 사이버공격 분야에 대한 공개출처정보의 활용은 매우 미진하다. 물론 사이버공격 관련사건과 인물 그리고 조직 등에 대한 데이터베이스를 구축한 뒤 사이버공격자의 개별 특성과 유형분석을 통해 사이버방어 작전에 활용한 예[9]가 있기는 하지만, 사이버공격 자체에 대한 공개출처정보 활용에 대한 연구는 매우 제한적인 것이 현실이다. 따라서 본 논문에서는 공개출처정보를 활용한 사이버공격 데이터베이스 구축방안과 향후 활용방안을 제시한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 공개출처정보를 활용한 데이터베이스 구축사례에 대해 ISVG와 START를 중심으로 분석한다. 3장에서는 공개출처정보를 활용한 사이버공격 데이터베이스 구축과정을 설명하고, 4장에서는 사이버공격 데이터베이스 변수 선정에 대해 설명한다. 5장에서는 3장과 4장을 통해 설계된 사이버공격 데이터베이스에 데이터를 입력하고 관리하기 위한

코드북(codebook)을 제안한다. 마지막으로, 6장에서는 본 논문에 대한 결론을 도출하고, 향후 연구 방향을 제시한다.

2. 공개출처정보 데이터베이스 구축사례 분석

9.11 테러를 포함한 대테러 전쟁을 수행하면서 국가별로 공개출처정보에 대한 관심이 급증하였고, 현재는 미국과 나토 그리고 이스라엘 등이 공개출처정보를 적극 활용하고 있다. 이들 국가들은 공개출처정보의 수집, 데이터 정제, 데이터베이스 구축 그리고 분석에 이르는 전반적인 과정에 대한 구체적인 방법론을 연구하고 있으며, 범죄예방 등의 목적으로 활용하고 있다. 본 장에서는 공개출처정보를 활용한 사이버공격 데이터베이스 구축을 위한 참조모델로 활용될 수 있는 사례로 ISVG 프로그램과 START 프로그램을 각각 살펴본다[2, 5].

2.1 ISVG 프로그램

ISVG는 뉴 헤븐 대학(University of New Haven)의 비영리 단체로, 폭력적인 극단주의에 대한 정보력을 높이기 위해 데이터를 수집하며 테러 및 범죄 예방과 관련된 의사결정을 지원하기 위해 관련 기관에 해당 데이터와 분석결과를 제공한다. 2004년부터 시작된 ISVG 프로그램은 다양한 언어권의 연구원들을 활용하여 약 11개 언어로 작성된 자료를 검색 및 수집하여 데이터베이스를 구축하는데, 이미 약 15만 건 이상의 데이터가 축적되어 있는 것으로 알려져 있다.

ISVG 데이터베이스에서는 테러공격에 대한 사건 정보, 테러리스트 정보, 그리고 테러 조직 정보 등이 자료로 입력된다. 첫째, 테러공격 사건들은 21개의 사건유형들로 세분화된다. 이때 21개의 유형들은 각각 방화, 암살, 무장공격, 생물테러, 폭탄테러, 휴전, 화학테러, 민사재판, 형사재판, 통신, 사이버테러, 자금조달, 하이재킹, 인질납치/유괴, 인질석방, 군사작전, 핵 테러, 경찰활동, 강도, 밀거래, 반달리즘(vandalism)/사보타주(sabotage) 등이다. 이때 각각의 테러공격 사건들은 유형에 따라 사건일자, 시간, 장소, 피해정도, 사용된 무기의 종류, 공격주체와 대상 등의 구체적인 변수에 따른 정보들이 입력된다. 둘째, 테러리스트 개인에 대한 프로파일은 이름(혹은 가명), 사진과 국적, 여권등록 국가, 현재 또는 이전 거주지, 관련된 테러나 범죄사건, 관련된 조직, 현재의 상태(생사여부, 투옥여부 등), 교육정도, 가족관계, 방문한 국가, 범죄경력, 그리고 특징들과 일반 정보들이 각각 입력된다. 마지막으로 테러조직의 경우 조직명과 조직구성, 구조, 자금조달, 관련테러사건, 활동지역, 관련된 인물, 하부조직, 이념, 주요 공격대상, 다른 세력과의 연계도 등에 관한 자료가 분류되어 입력된다[11].

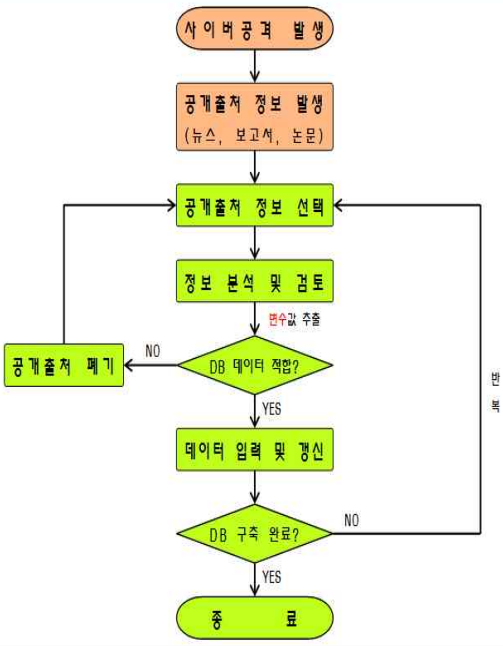
이러한 테러 사건, 테러리스트, 테러조직에 대한 각각의 자료들은 서로 데이터베이스 상에서 연계되어 교차검색을 할 수 있도록 연동되어 있으며, 데이터의 시각화 및 분석 기능을 제공하고 있다.

2.2 START 프로그램

START는 미국 국토안보부의 지원으로 메릴랜드 대학교에서 시행 중인 프로그램으로 테러와 관련된 다양한 정보를 수집하여 효과적인 대테러 활동의 보장을 그 목적으로 한다. START는 다양한 데이터베이스를 구축하여 운용하고 있는데, 이 가운데 테러사건에 대한 데이터베이스는 GTD(Global Terrorism Database)로 공개된 정보를 기초로 전 세계에서 일어났던 테러사건들에 대한 관련 자료들(사건정보, 공격정보, 무기정보, 피해정보 등)이 저장된다. 국내·외에서 발생한 테러 사건에 대한 체계적인 데이터를 포함하고 있으며 현재 11만 건 이상의 정보가 저장되어 있다. 각 사건별로 사건 발생 날짜와 위치, 사용된 무기 및 대상의 성격, 사상자의 수, 그리고 범인의 신원에 관한 자료들이 제공된다. GTD의 경우 인터넷에 입력된 데이터 및 데이터베이스 구조가 모두 공개되어 있다. 따라서 누구나 자료를 열람할 수 있으며, 많은 연구자들이 이를 연구에 활용하고 있다.

3. 사이버공격 데이터베이스 구축과정

이번 장에서는 공개출처정보를 활용한 사이버공격 데이터베이스 구축과정에 대해 설명한다. 본 연구에서 제안하는 공개출처정보를 활용한 사이버공격 데이터베이스 구축과정은 (그림 1)에서 보는 바와 같다.



(그림 1) 사이버공격 데이터베이스 구축과정

3.1 사이버공격 공개출처정보의 종류

(그림 1)에 제시된 바와 같이 사이버공격이 발생하면 다양한 유형의 공개출처정보가 발생하는데, 본 연구를 통해 식별된 사이버공격 공개출처정보는 ① 뉴스 (TV, 신문, 라디오 기사 등), ② 논문 (심사과정을 거쳐 발행되는 학술지 또는 학술대회 논문), ③ 보고서 (정부기관 또는 보안업체에서 발행한 보고서), ④ 소셜 미디어 (개인 및 단체 포함), ⑤ 블로그 (개인, 단체에서 운용하는 블로그), ⑥ 기타 (API, 웹 스파이더 등의 크롤링(crawling) 및 표준 RSS 방식 등)로 구분된다.

3.2 사이버공격 데이터베이스 구축과정

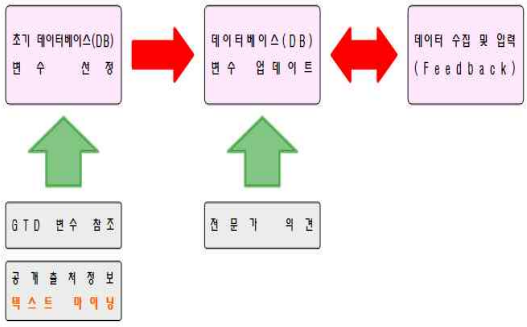
공개출처정보를 활용한 사이버공격 데이터베이스 구축은 사이버공격이 발생한 이후에 생성되는 공개출처정보의 선택으로부터 시작된다. 이때 선택된 공개출처정보는 제일 먼저 분석 및 검토 단계를 거치는데 일반적으로 공개출처정보는 동일한 정보에 대해서도 기술하는 형태, 방식, 용어 등이 다르기 때문에 데이터베이스 입력을 위한 정형화 과정이 필요하다. 이때 정형화 과정을 통해 사이버공격 데이터베이스 구축을 위한 변수를 결정하게 되는데, 이 과정에 대해서는 4장에서 자세히 설명한다.

정보 분석 및 검토 과정을 통해 사이버공격 데이터베이스에 입력될만한 가치가 인정되면 데이터베이스에 입력하며, 그렇지 않을 경우 해당 공개출처정보는 폐기된다. 이때 하나의 공개출처정보(ex. 뉴스)로부터 사이버공격 데이터베이스 각 필드를 채우기에 충분한 정보를 획득하지 못하는 경우에는 다른 검색방법을 동원해 동일한 사이버공격에 대한 다른 공개출처정보를 찾아 보완한다. 만약 분석한 공개출처정보가 이미 입력된 사이버공격에 대한 내용일 경우 해당 사이버공격에 대한 내용을 갱신한다.

(그림 1)에서 보는 바와 같이 공개출처정보를 활용한 사이버공격 데이터베이스 구축은 통계분석을 위한 데이터가 충분히 확보되어 더 이상의 공개출처정보 입력이 필요 없을 경우 종료하며, 그렇지 않을 경우에는 공개출처정보 선택부터 데이터 입력 및 갱신과정을 반복한다.

4. 사이버공격 데이터베이스 변수 선정

공개출처정보를 활용한 사이버공격 데이터베이스가 효과적이기 위해서는 데이터베이스에 들어가는 정보 자체도 중요하지만, 그 정보를 묘사 혹은 표현하는 데이터베이스 변수가 더 중요하다. 왜냐하면 데이터베이스 변수 자체가 제대로 선정되지 못하면 그 변수를 이용해 입력된 정보를 활용한 분석도 그 효과를 보장할 수 없기 때문이다.



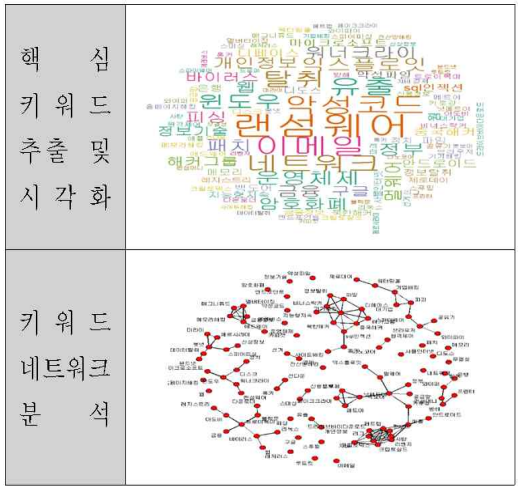
(그림 2) 사이버공격 데이터베이스 변수의 선정과정

본 논문에서는 (그림 2)에서 보는 바와 같이 GTD 변수를 참조한 결과와 사이버공격 관련 공개출처정보에 대한 텍스트 마이닝(text mining)을 통해 초기 변수를 선정하였고, 선정된 초기 변수에 대한 전문가의 의견을 반영하였으며, 데이터의 수집 및 입력 과정을 통해 지속적으로 보완함으로써 최종적으로 사이버공격 데이터베이스 변수를 선정하였다.

4.1 공개출처정보 텍스트 마이닝(text mining)

공개출처정보를 활용한 사이버공격 데이터베이스 구축을 위한 첫 번째 단계는 공개출처정보에 대한 텍스트 마이닝을 통해 초기 데이터베이스 변수를 선정하는 것이다. 초기 데이터베이스 변수 선정을 위해 사이버분야에 전문성을 가진 기자들에 의해 작성되고 있는 보안뉴스[1] 사이트의 사이버공격 기사가 활용되었다. 즉, 2017년 2월부터 8월 사이에 보안뉴스에 보도된 사이버공격 관련 뉴스 111개를 선정해 텍스트 마이닝 기법을 우선적으로 적용했다.

텍스트 마이닝의 첫 번째 단계는 각 기사에 대해 공란 처리, 대·소문자 통일, 문장부호 제거, 유사어 처리 및 불용단어 제거 등의 정형화 작업이다. 다음에는 정형화된 111개의 문서 간 유사도에 따른 그룹화를 진행한 후, 정성적 분석으로 도출된 50여 개의 사이버공격 관련 핵심 키워드를 추출해 사이버공격 데이터베이스 초기 변수로 선정하였다. 이때 각 키워드 간의 상관관계를 확인하기 위해 워드 클라우드를 활용해 핵심 키워드를 시각화하고, 키워드 간의 네트워크를 분석하였다. (그림 3)은 핵심 키워드 추출, 시각화, 그리고 네트워크 분석결과의 예를 보여준다.



(그림 3) 핵심 키워드 시각화 및 네트워크 분석

4.2 데이터 수집 및 입력을 통한 변수 업데이트

앞서 4.1절에서 사이버공격 공개출처정보에 대한 텍스트 마이닝을 통해 결정된 초기 데이터베이스 변수 50개는 사이버공격 관련 공개출처정보의 내용을 정확히 담기에는 불완전하기 때문에 (그림 2)에서 보는 것과 같이 실제 데이터를 수집하고 입력하는 과정을 통해 보완해야 할 필요가 있다. 따라서 연구진은 사이버공격 공개출처정보에 대한 데이터 수집 및 입력 과정을 반복하여 초기 데이터베이스 변수를 보완하였다. 이때 (연구기간의 제한사항 및 신뢰성과 전문성이 확보된 출처를 활용하기 위해) 2017년 1월부터 12월까지의 보안뉴스 기사들 중 사이버공격 관련 기사 134개를 선별해 데이터베이스를 구축하였다[2]. 최초에는 텍스트

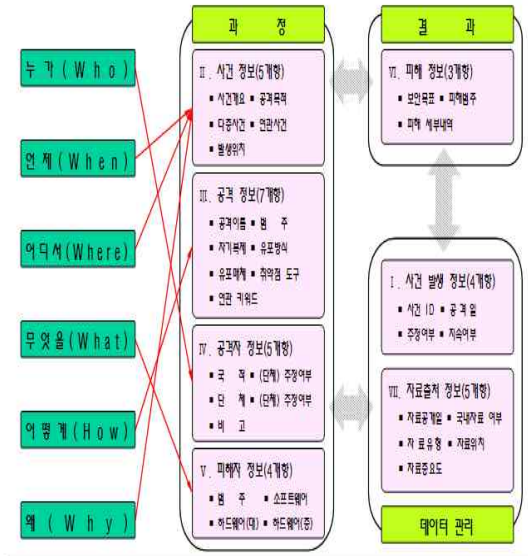
1) https://www.boannews.com/

2) 본 연구에서는 테스트(text) 위주의 공개출처정보 활용방안을 연구하였으며, 오디오나 비디오와 같은 데이터에 대한 연구는 향후 연구로 남긴다.

마이닝을 통해 결정된 초기 데이터베이스 변수 50개를 활용해 데이터베이스에 데이터를 입력하되 필요한 경우 일부 변수를 추가하거나 제외하고, 연관된 변수들을 그룹화하는 과정을 반복하였다. 이런 과정을 거쳐 134개의 기사를 모두 입력하였을 때에는 최종적으로 (그림 4)에서 보는 바와 같이 7개 범주에서 33개의 변수가 결정되었다.

4.3 선정된 사이버공격 데이터베이스 변수들

공개출처를 활용한 사이버공격 데이터베이스는 (그림 4)에서 보듯이 크게 ① 사건발생 정보, ② 사건 정보, ③ 공격 정보, ④ 공격자 정보, ⑤ 피해자 정보, ⑥ 피해 정보, 그리고 ⑦ 자료출처 정보 등 7가지 범주 33개 변수로 구분된다. 이때 사건발생 정보와 자료출처 정보는 데이터 관리를 위한 범주이고, 피해 정보는 해당 사이버공격에 의해 훼손된 보안목표 및 피해의 세부 내용에 대한 내용이다. 사건 정보, 공격 정보, 공격자 정보, 그리고 피해자 정보는 사이버공격 및 사이버공격자 그리고 사이버공격으로 인한 피해자 정보를 묘사하기 위한 범주이다.



(그림 4) 선정된 7개 영역 33개 변수

이와 같이 134개의 보안뉴스 기사를 입력하면서 최종적으로 결정된 7개 범주 33개 변수를 활용하면 (그림 4)에서 보듯이 누가(who), 언제(when), 어디서(where), 무엇을(what), 어떻게(how), 왜(why)라고 하는 육하원칙에 의거 사이버공격을 묘사할 수 있다는 장점이 있다.

5. 사이버공격 데이터베이스 코드북

3장과 4장에 제시된 바와 같이 공개출처정보를 활용해 사이버공격을 위한 데이터베이스 설계를 마친 이후에는 사이버공격과 관련된 다양한 공개출처정보를 확보해 입력해야 한다. 이때 사이버공격에 대한 공개출처정보를 데이터베이스에 입력하기 위해서는 몇 가지 제한사항이 있고, 이를 해결하기 위해서는 데이터베이스 입력을 위한 명확한 가이드라인이 정립되어야 한다. 그렇지 않으면 입력된 데이터가 정화되지 않아 신뢰성이 떨어질 수 있다. 따라서 본 장에서는 공개출처정보를 활용한 사이버공격 데이터베이스 구축 시 발생할 수 있는 제한사항에 대해 알아보고, 이를 극복하기 위한 사이버공격 데이터베이스 코드북(cyber attack database codebook) 작성방안을 제안한다.

5.1 사이버공격 데이터베이스 입력 간 제한사항

사이버공격에 대한 공개출처정보는 누구나 쉽게 작성할 수 있다. 이때 작성된 공개출처정보가 사이버분야의 전문가에 의해 작성된 경우라면 정보의 정확성에 대한 신뢰도를 기대할 수 있겠지만, 그렇지 않은 경우도 다수 존재한다. 만약 공개출처정보가 기술보고서나 학술논문이라면 심의기관이 존재하지만, 블로그를 포함한 대다수의 공개출처정보는 정보를 검증하는 별도의 인증기관이 존재하지 않는다. 따라서 사이버공격에 대한 공개출처정보를 데이터베이스로 입력하기 전에 반드시 검증이 필요하다.

또한 동일한 사이버공격에 대한 내용을 기술하고 있는 공개출처 정보라 하더라도, 저자들의 상

이한 배경지식 및 무분별한 용어 사용으로 같은 현상에 대한 표현이 일치하지 않는 경우가 종종 발생한다. 일례로 바이러스(virus), 웜(worm), 그리고 트로이 목마(trojan horse)는 명확히 다른 개념의 사이버공격이지만, 많은 사이버공격 공개출처정보에서 동일한 의미로 사용하거나 혼용하는 사례가 있어 명확한 구분이 필요하다.

5.2 사이버공격 데이터베이스 코드북의 작성 및 활용

앞서 5.1절에서 살펴보았듯이 사이버공격에 대한 공개출처정보를 데이터베이스로 입력할 때에는 전문성을 가진 사람들에 의해 명확한 기준을 가지고 정제된 데이터가 입력되어야 한다. 이를 위해 본 논문에서는 사이버공격 데이터베이스 코드북의 작성 및 활용을 제안한다. 사이버공격 데이터베이스 코드북은 사이버공격에 관한 공개출처정보를 데이터베이스로 입력할 때 명확한 기준을 제시하고, 데이터베이스를 효과적으로 관리하기 위한 일종의 매뉴얼이다.

사이버공격 데이터베이스 코드북은 사이버공격 데이터베이스에서 사용되는 모든 변수 및 변수별 입력 값을 차례대로 정의한다. 또한 항목별 변수의 의미와 사용목적을 설명함으로써 데이터 입력의 모호성을 제거하고, 초보자나 비전문가도 코드북만 보면 입력을 할 수 있도록 다양한 입력 사례들을 추가 기술하여야 한다. 사이버공격 데이터베이스 코드북에 대한 자세한 내용은 사이버사령부 연구보고서[10]를 참조하기 바란다.

5.3 사이버공격 데이터베이스 코드북 관리

사이버공격은 하루가 다르게 진화하고 있기 때문에 공격기법이 발전함에 따라 기존의 항목으로는 설명이 어려울 수 있다. 이런 경우에는 변수 및 입력 값의 추가 또는 수정이 불가피하다. 반대로 활용도가 저조하거나 불필요한 항목이 있을 경우 자료를 삭제할 필요가 있다. 이때 데이터베이스의 변수에 대한 추가, 수정 및 삭제 간 오류가 발생할 수 있기 때문에 코드북 변경에 대해서는 명확한 이력관리와 형상관리가 요구된다.

6. 결론 및 향후 연구방향

본 논문에서 우리는 공개출처정보를 활용한 사이버공격 데이터베이스 구축방법론을 제안하였다. 이를 위해서 데이터 마이닝 기법을 활용해 100여 개의 사이버공격 기사로부터 50여 개의 초기 데이터베이스 변수를 추출하였고, 134개의 사이버공격 관련 공개출처정보를 입력하는 과정을 통해 최종적으로 7개 범주 33개의 데이터베이스 변수를 확정하였다. 또한 효과적이고 명확한 데이터베이스 구축을 위한 코드북을 제시하였으며, 이 코드북을 활용해 2017년도에 보안뉴스에 제시되었던 사이버공격 기사들에 대한 초기 데이터베이스를 구축하였다.

향후 우리는 구축된 사이버공격 데이터베이스에 대한 기술통계 분석을 통해 제안한 공개출처를 활용한 사이버공격 데이터베이스 구축방안을 보완하고, 나아가 새로이 구축될 사이버공격 데이터베이스를 활용해 사이버공격을 예측할 수 있는 예측모델을 개발할 것이다.

본 연구에서 제안한 사이버공격 데이터베이스가 지속적으로 업데이트되고, GTD와 유사하게 구축된 데이터베이스를 인터넷에 공개해 누구나 해당 자료를 활용할 수 있도록 한다면, 집단지성을 통해 사이버공격 경향을 보다 쉽게 예측하고 분석할 수 있을 것이다. 따라서 정부기관, 연구소, 그리고 학교기관이 연계하여 공개형 사이버공격 데이터베이스를 구축할 것을 제안한다[12].

참고문헌

[1] Jin Gui Min, "National Intelligence Studies, 9th edition", Baeum, Jan. 2019.

[2] Wanhee Lee, Minwoo Yun, and Jung Seok Park, "Intelligence in the Internet Era: Understanding OSINT and Case Analysis", Korean Security Science Review, No. 34, pp. 259-278, 2013.

[3] Byungchul Cho, "A System for National Intelligence Activity Based on All Kinds of OSINT(Open Source INTelligence) on the Internet", Journal of Information and Security, Vol. 3, No. 2, pp. 41-55, June 2003.

[4] Woong Chun, "Open Source Intelligence in the Information Age", Journal of National Intelligence Studies, Vol. 1, No. 1, pp. 151 - 172, July 2008.

[5] Minwoo Yun, "Construction of Database for Terrorism and Crime through OSINT", The Korean Association of Criminal Psychology, Vol. 13, No. 2, June 2017.

[6] Mabrey Daniel, "Analyzing Terrorist Activities through Operational & Associational Coding of Events: Introducing the Institute for the Study of Violent Groups' Relational Database", ISVG center, 2010.

[7] START, "Global Terrorism Database Codebook : Inclusion Criteria and Variables", START Center. University of Maryland, 2017.

[8] FBI Office of the Program Management Executive, "Security Concept of Operations (S-CONOPS), Investigative Data Warehouse (IDW) Program", Electronic Frontier Foundation, Nov. 2004.

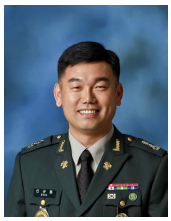
[9] Wanju Kim, Changwook Park, Soojin Lee, and Jaesung Lim, "Methods for Classification and Attack Prediction of Attack Groups based on Framework of Cyber Defense Operations", Journal of KIISE, Vol. 20, No. 6, pp. 317-328, June 2014.

[10] Kyuyong Shin, Jincheol Yoo, Changhee Han, Sungrok Kang, Jongkwan Lee, Minam Moon, Kyoung Min Kim, "A Study on Cyber Threat Analysis based on Open Source Intelligence", Technical Report, ROK Cyber Command, Dec. 2018.

[11] ISVG(Institute for the Study of Violent Groups), ISVG Relational Database Codebook, Sam Houston University, Huntswille, TX. 2005.

[12] Yeongdo Jung and Jeonggi Seog, "A Study on Countermeasures against North Korea's Cyber Attack", Journal of Information and Security, Vol. 16, No. 6, pp 43-50, Oct. 2016.

〔저자소개〕



신 규 용 (Kyuyong Shin)
1996년 3월 육군사관학교 학사
2000년 2월 한국과학기술원 석사
2009년 12월 (美)노스캐롤라이나 주립대학교(NCSU) 박사
email : kyshin@kma.ac.kr

유 진 철 (Jincheol Yoo)
1989년 3월 육군사관학교 학사
1993년 미국 아이오와 주립대학교 (Iowa State Univ.) 석사
2003년 미국 펜실베니아 주립대학교 (Pennsylvania State Univ.) 박사
email : jyoo@kma.ac.kr

한 창 희 (Changhee Han)
1990년 3월 육군사관학교 이학사
1994년 6월 (美)Syracuse Univ. 석사
2004년 6월 (美)Univ. of Southern California 박사
email : chhan@kma.ac.kr

김 경 민 (Kyoung Min Kim)
2003년 2월 육군사관학교 학사
2008년 8월 Auburn대 전산학 석사
2016년 12월 ~ 현재
육군사관학교 컴퓨터과학과 조교수
2018년 6월 ~ 현재 육군사관학교 사이버전 연구센터 연구원
email : kimkyou@kma.ac.kr

강 성 록 (Sungrok Kang)
1996년 3월 육군사관학교 학사
2001년 2월 연세대학교 석사
2010년 8월 (美)오리건주립대(OSU) 박사
email : ksr6452@mnd.go.kr

문 미 남 (Minam Moon)
2001년 3월 육군사관학교 학사
2006년 2월 고려대학교 수학석사
2015년 8월 텍사스 A&M 대학교 수학박사
email : hereandnow@kma.ac.kr

이 종 관 (Jongkwan Lee)
2000년 3월 육군사관학교 학사
2004년 3월 한국과학기술원 석사
2011년 3월 아주대학교 박사
email : jklee64@kma.ac.kr