ISSN(Print) : 1976-5304, ISSN(Online) : 2713-931X
https://kdfs.jams.or.kr, DOI : 10.22798/kdfs.2022.16.2.34

메모리 포렌식 기반 키워드 매칭을 통한 다크 웹 사용자 행위 분석

김 예 은*, 도 예 진*, 한 상 연*, 김 성 민**
성신여자대학교 융합보안공학과 (대학생)*, (교수)**

Dark Web User Behavior Analysis through Keyword Matching Based on Memory Forensics

Yeeun Kim*, Yejin Do*, Sangyeon Han*, Seongmin Kim**
Dept. of Convergence Security Engineering, Sungshin Women's University (Undergraduate Student)*, (Professor)**

요 약

최근 토어 네트워크가 보장하는 익명성을 기반으로 한 다크 웹을 통해 불법적인 거래가 다수 발생함에 따라, 다크 웹 내의 범죄 행위 추적을 위한 디지털 포렌식 수사의 중요성이 강조되고 있다. 토어 브라우저의 경우 범용 웹 브라우저와 달리 웹 아티팩트 흔적을 남기지 않아 분석이 제한적이다. 그러나 메모리 덤프 기반 포렌식을 수행할 경우 사용자의 행위에 관한 평문이 그대로 저장되어 포렌식 아티팩트를 획득할 수 있다. 이에 본 논문에서는 메모리 포렌식을 통해 토어 브라우저를 통한 다크 웹 내 사용자 행위를 분석하는 프레임워크를 제안한다. 호스트 내 토어 설치 여부 및 접속 여부와 같은 포렌식 수사 과정에서 필요한 아티팩트를 수집한 뒤, 평문 형태로 저장되는 사용자의 다크 웹 검색 내역을 메모리 덤프로부터 추출한다. 제안한 프레임워크는 사용자의 불법적 행위를 효과적으로 탐지하기 위해 토어 히든 서비스들로부터 웹 크롤링을 통해 자동으로 수집한 키워드들을 바탕으로 패턴 매칭을 수행함으로써 수사의 효율성을 보장한다. 또한 실제 범죄에서 수사망을 피하기 위해 사용하는 은어들도 수집해 패턴 매칭을 함으로써 실질적인 범죄 행위 탐지가 가능함을 보여준다. 효율성 검증을 위해 다크 웹을 통한 불법적인 거래 시나리오를 설정한 뒤 이를 시뮬레이션한 결과, 제안한 프레임워크가 성공적으로 범죄 행위를 탐지하는 것을 확인하였다.

주제어 : 다크 웹, 토어 네트워크, 메모리 포렌식, 디지털 포렌식

ABSTRACT

With the recent occurrence of a number of illegal transactions through the Dark Web based on anonymity guaranteed by the Tor network, the importance of digital forensics investigations to track criminal acts in the Dark Web is being emphasized. Unlike generic Web browsers, Tor browser does not provide sufficient clues for analysis because it does not leave traces of Web artifacts. However, in the case of memory dump forensic, the plaintext about the user's behavior is stored as it is, so we can acquire forensics artifacts. To address this challenges, we propose a framework for analyzing user behavior on the Dark Web using Tor browser through memory forensics. After collecting Tor artifacts, such as whether to install and access Tor in the host, the user's Dark Web search history stored as plaintext is extracted from the memory dump. The proposed framework ensures the efficiency of the investigation by performing pattern matching based on keywords automatically collected through Web crawling from Tor Hidden services to effectively detect illegal behavior of users. In addition, it shows that it is possible to detect practical criminal behavior by collecting slang used in actual crimes and matching patterns. While simulating an illegal transaction scenario through the Dark Web to verify its efficiency, we confirm that the proposed framework successfully detects criminal behavior.

Key Words : Dark Web, Tor Network, Memory Forensic, Digital Forensic

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2021R1G1A100632611)과 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0008703, 2022년 산업혁신인재성장지원사업)을 받아 수행된 연구임.
- Received 28 February 2022, Revised 04 March 2022, Accepted 14 June 2022
- 제1저자(First Author) : Yeeun Kim (Email : yeuna119@kakao.com)
- 교신저자(Corresponding Author) : Seongmin Kim (Email : sm.kim@sungshin.ac.kr)

Ⅰ. 서론

다크 웹(Dark Web)을 악용한 불법 금융 거래 및 마약, 아동 음란물 사이트 등에 접속하는 사례가 증가함에 따라 사이버 범죄의 심각성이 증대되고 있다. 실제로 토어(Tor) 네트워크 기반 다크 웹 브라우저인 토어 브라우저에서 운영되고 있는 5,205개의 사이트를 무작위로 추출하였을 때, 약 30%인 1,547개의 사이트가 마약, 불법 금융 거래 등을 위해 사용되었음이 확인되었다[1]. 이러한 범죄 행위에 대응하기 위해 다크 웹상에서 행해지는 범죄 행위를 탐지하기 위한 비 익명화 기법 및 다크 웹 포렌식을 바탕으로 한 수사 기법의 연구 필요성이 대두되고 있다. 3-hop Onion routing 기반의 토어 네트워크가 보장하는 익명성으로 인해[2], 특정 토어 노드나 서버 종단에서의 네트워크 패킷 캡처를 수행하더라도, 토어 클라이언트를 통한 범죄 행위를 추적하기가 어렵다는 근본적인 한계점이 존재한다.

또한, 웹 브라우저를 대상으로 하는 기존 웹 아티팩트 및 포렌식 기법들의 활용이 토어 브라우저를 이용한 경우 제약된다는 점은 다크 웹을 대상으로 한 디지털 포렌식을 어렵게 하는 요소이다[3]. 게다가, 토어 브라우저는 브라우저 핑거프린팅(fingerprinting)에 대한 기본적인 방어 기능을 제공한다[4]. 네트워크 단에서 클라이언트의 토어 브라우저를 통한 다크 웹 접속 탐지가 어려우므로, 명확한 다크 웹 접속 기록 등을 추적하기 위해서는 호스트 내에서 범죄 행위의 흔적을 찾아야 한다. 하지만 사용자가 토어 브라우저를 통해 웹에 접속할 경우, 웹 캐시, 웹 히스토리(history) 등의 웹 아티팩트 흔적을 남기지 않는다. 이에 전통적인 디지털 포렌식 기법인 메모리 덤프 기반 활성 데이터 수집 및 분석이 하나의 옵션이 될 수 있다.

메모리 덤프를 통한 포렌식 기법은 토어 브라우저 관련 데이터를 포함한 호스트 내 애플리케이션 실행 등의 사용자 행위를 평문 데이터 형태로 접근 가능케 한다[5]. 즉, 획득한 활성 데이터에 대해 범죄 행위로 의심할 수 있는 키워드를 이용하여 패턴 매칭을 수행하면 호스트 내 사용자의 다크 웹을 통한 범죄 행위를 판단할 수 있는 포렌식 아티팩트를 획득할 수 있다. 그러나, 크기가 큰 메모리 덤프 파일에 대해 단순히 사전에 기반한 패턴 매칭을 수행할 경우, 키워드 매칭의 오버헤드가 크고 EntryPoint 분석을 위한 데이터의 양이 방대해질수록 비효율성은 커질 수밖에 없기에 이를 해결하기 위한 효율적인 수사 방안이 필요하다.

본 논문에서는 입력으로 주어진 메모리 덤프 형태의 활성 데이터로부터 효율적인 메모리 포렌식을 통해 다크 웹 사용자 행위를 자동으로 분석하는 프레임워크를 제안한다. 먼저 호스트 내 토어 설치 여부 및 접속 여부와 같은 포렌식 수사 과정에서 필요한 정보를 수집한 뒤, 평문 형태로 저장되는 사용자의 다크 웹 검색 내역을 메모리 덤프로부터 추출한다. 이 과정에서, 사용자 행위 분석의 효율성 향상을 위해 실제 토어 네트워크상에서 운용되고 있는 토어 히든 서비스로부터 웹 크롤링을 통해 키워드를 수집한다. 이렇게 자동으로 수집한 키워드들을 바탕으로 활성 데이터에 대해 패턴 매칭을 수행함으로써 수사의 효율성을 보장한다. 키워드에는 범죄와 관련된 직접적인 단어(e.g.drug) 뿐만 아니라 불법 거래 시 은밀하게 사용되는 은어도 수집되는 실질적 포렌식 아티팩트 탐지가 가능하며, 실제 다크 웹상의 Onion 도메인에서 수집된 키워드를 선정하기 때문에 범죄 행위 패턴 매칭 수사의 효율성을 올릴 수 있음을 보여준다. 제안한 프레임워크의 효율성 검증을 위해 다크 웹을 통한 불법적인 거래 시나리오를 설정한 뒤 이를 시뮬레이션한 결과, 제안한 프레임워크가 성공적으로 범죄 행위를 탐지하는 것을 확인하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 관련 연구를 서술하고, 3장에서는 토어 네트워크 및 히든 서비스와 관련된 배경지식을 설명한다. 4장에서는 다크 웹 사용자 행위 분석을 위한 프레임워크 구조 및 동작 흐름도를 다루고, 5장과 6장에서는 제안한 프레임워크의 포렌식 성능 평가를 위해 가상의 범죄 행위 시나리오를 설정하고 이에 대한 탐지 여부를 검증한다. 마지막으로 7장에서는 결론 및 향후 연구 방향에 관해 서술한다.

Ⅱ. 관련 연구

토어 내 불법 거래 및 서비스에 대한 실태 파악 및 수사 기반 마련을 위해 다양한 연구가 이루어지고 있는데, 본 장에서는 이와 관련된 연구를 크게 다크 웹 수집기 및 토어 포렌식으로 나누어 분석한 뒤, 해당 연구들과 제안한 연구의 차별성을 설명한다. 이후, 본 연구에서 다크 웹 사용자 행위 분석을 위해 활용할 메모리 덤프 기반 포렌식을 활용한 연구들을 살펴본다.

2.1. 다크 웹 수집기

토어 브라우저는 일반 웹 브라우저와 다르게 익명성을 위한 별도의 애플리케이션 프로토콜을 사용하고 소켓 연결이 있어야 하는 등 일반 웹 크롤러와 같은 방식으로 서비스 분석을 원활하게 수행할 수 없다.

따라서, 토어 분석을 위해서는 Proxy와 소켓 통신을 사용하는 등 일반 웹 브라우저와는 다른 방식을 사용해야 한다. Jonghyeon Park 외 1명은 불법적인 히든 서비스를 탐색하기 위해 토어 전용 크롤러를 제안했다[6]. 해당 연구에서는 토어 내에서뿐만 아니라 일반 네트워크의 히든 서비스 탐색도 가능한 대역 외 탐색의 주소 수집기를 구현했다. 약 한 달간의 실험을 통해 고유 히든 서비스 주소 2,439개와 그중 정상 운영 중인 1,345개의 서비스를 확인하였다. 해당 연구를 통해 발견된 히든 서비스는 전체의 2%에 불과했으나 일반 네트워크에서의 히든 서비스 탐지에 관한 연구 가능성을 보여주었고, 이를 바탕으로 심층 분석을 수행한다면 추가적인 히든 서비스를 발견할 수 있을 것으로 예측하였다.

Hyunsu Mun 외 2명은 토어 Socks5 프록시 서버를 활용해 렌더링 및 스크립팅 작업 제거를 통한 토어 기반 다크 웹 수집 성능 개선 방법을 제안하였다[7]. 이 연구에서는 다크 웹 주소 612개를 대상으로 실험하여 기존 수집기와 비교해 평균 약 1.8배를 개선한 실험 결과를 보였고, 렌더링과 스크립팅을 제거하여 수집 성능을 최대 10배가량까지 개선하였다. 본 논문의 실험에서도 Onion 사이트의 크롤링을 위해 SOCKS와 Socket을 통한 프로토콜을 연결하여 실험을 진행하였다. 이처럼 다크 웹 수집기에 관한 연구는 다양한 방법으로 이루어지고 있으며, 본 연구에서 제안하고자 하는 빈도수 기반 키워드 매칭을 위해서는 범죄 관련 Onion 주소의 크롤링 과정이 선행적으로 요구된다. 이러한 다크 웹 수집기 관련 기법들과 본 연구에서 제안한 방법을 연계할 경우, 다크 웹 사용자 행위 분석의 성능 향상을 위한 양질의 키워드 수집이 가능할 것으로 볼 수 있다.

2.2. 토어 포렌식

다크 웹 내 사용자 및 사이트 운영자의 개인정보 및 IP 주소를 비 익명화하기 위한 관련 연구로, Gyungbin Lee는 토어 네트워크에서 운영된 히든 서비스 사이트의 운영방식 사례를 분석하여 특징을 파악하고 디지털 포렌식을 수행하였다[1]. 해당 논문에서는 전송 속도가 중요한 불법 콘텐츠의 업로드 및 다운로드를 하는 서버를 대상으로 포렌식 사례 연구를 수행하였다. 구체적으로, 토어를 사용할 때 여러 노드를 거친 암호화 통신으로 인해 불가피하게 발생하는 속도 저하를 개선하기 위해 히든 서비스가 외부 프록시를 사용하는 경우들이 존재하며, 이 경우 트래픽이 노출되기 때문에 추적 방법론에 적용할 수 있다고 보았다. 나아가, 저자는 다크 웹의 추적은 불가능할 것이라는 인식으로 인해 다크 웹의 기술을 지나치게 고도화하여 생각하는 것에 대해 논하며 현재까지 연구된 추적 기법과 더불어 다크 웹 히든 서비스상에서 발생 가능한 취약점에 관한 연구 필요성을 제기하였다.

Priya P Sajan 외 3인은 토어 브라우저가 생성하는 아티팩트를 추출한 뒤 이를 분석하고자 메모리 덤프 파일과 HxD를 이용하여 방문한 웹사이트의 URL을 확인할 수 있음을 보여주었다[8]. 본 연구에서는 메모리 덤프 기반 포렌식을 활용하여 토어 사용자가 방문한 웹사이트 URL을 획득하는 수준에 그치지 않고, 구체적인 다크 웹 사용자의 행위 분석을 위한 효율적인 키워드 기반 매칭 기법을 제안한다. 또한, 앞서 언급한 기존 연구와 달리 다크 웹에서 운용되는 히든 서비스 자체의 취약성이나 토어 브라우저에 존재하는 취약성을 활용하지 않기 때문에 포렌식을 위한 조건 측면에서 별도의 제약이 존재하지 않는다.

2.3. 메모리 덤프 기반 및 키워드 분석 웹 포렌식

전통적인 메모리 포렌식 기법을 활용한 웹 포렌식 연구는 지속적으로 이루어지고 있는데, 대표적인 연구로는 Dongwon Shin와 Taeshik Shon이 수행한 크롬(Chrome) 브라우저에서 SNS, 온라인 결제 및 금융 서비스 등 총 5가지의 서비스 활용에 대한 아티팩트를 메모리 덤프를 바탕으로 분석한 연구가 있다[5]. 해당 연구에서는 덤프 파일에서 문자열 검색을 수행해 사용자의 정보를 서비스별로 분석하였고, 개인정보와 관련된 키워드를 확인하여 덤프 파일에서 사용자 정보가 있는 위치를 찾는 방안을 제시하였다. 또한 키워드 추출 연구와 관련한 연구로 W.J.C. van Staden과 E. van der Poel은 디지털 포렌식에서 발생하는 주요 문제로 방대한 양에 대한 데이터의 패턴을 발견하기 위한 검색, 필터링을 언급하며 용의자 후보를 식별하기 위한 상대적 중요도의 순위를 매기는 자동화된 키워드 추출 결과를 제시했다[9]. Nicole Lang Beebe와 Jan Guynes Clark은 현재 디지털 포렌식 수행 시 인덱싱 알고리즘 기반 텍스트 문자열 검색 도구를 사용하는 것을 언급하며 검색 결과를 그룹화하지 못하는 한계를 지적하였다. 이에 디지털 포렌식 텍스트 검색 후 클러스터링의 타당성 및 유용성을 제안하였다[10]. S. M. Hejazi 외 2인은 스트링 매치를 활용해 응용 프로그램과 프로토콜 흔적을 분석해 사용자 이름, 암호화 키, 방문한 URL과 같은 데이터 조각인 민감 정보를 추출하는 기술과 보안에 민감한 API 분석을 통해 문자열 매칭 기반 기법으로는 추출할 수 없는 민감 데이터를 추출하는 연구를

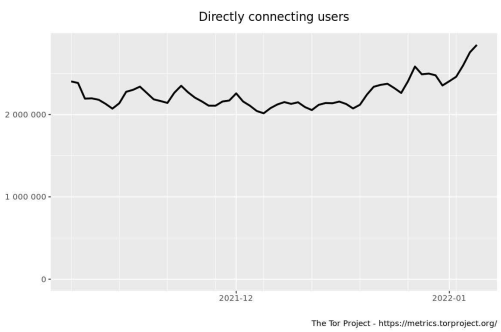
수행하였다[11]. 본 연구에서는 토어 브라우저에서 메모리 덤프를 통한 사용자 행위 분석을 수행하였는데, 이 과정에서 사용자 행위를 추적할 방안을 제시하기 위해 우선 범죄 사이트의 카테고리를 범주화하였다. 이후 선행 연구처럼 키워드를 선정하였는데, 제안 프레임워크에서는 실제 다크 웹 내 Onion 도메인에서 수집된 단어를 바탕으로 각 사이트의 범죄 관련 빈도수 기반 키워드를 추출해 키워드 매칭을 보다 효율적으로 수행하는 방안을 제시하고자 한다.

Ⅲ. 배경 지식

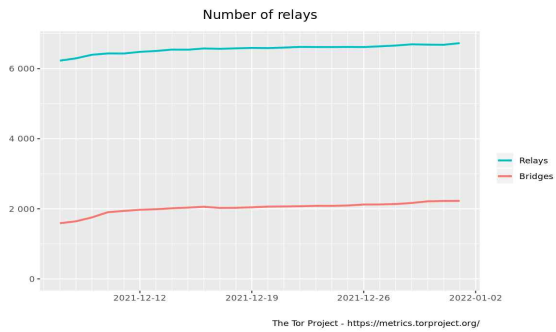
3.1. 토어 네트워크

토어 네트워크는 Onion Proxy 기반의 통신을 제공해 웹에 접속하는 클라이언트 및 서버의 익명성을 보장하는 네트워크이다[18]. 토어는 클라이언트, 노드 또는 릴레이(Relay), Directory Authorities(DA)로 크게 3가지 구성요소를 지니며, 이중 Relay의 경우 입구 노드(Guard Relay), 중계 노드(Middle Relay), 출구 노드(Exit Relay) 중 하나로 선택될 수 있다. DA는 토어 네트워크에 참여하는 모든 릴레이의 정보를 가지며 이들에게 역할을 부여한다.

토어 네트워크를 사용할 경우, 우선 사용자가 토어 브라우저를 실행하면 토어 디렉터리 서버(DA)로부터 토어 라우터로 등록된 서버 목록을 전달받는다. 이 중 임의로 선택한 3개의 서버를 입구 노드, 중계 노드, 출구 노드로 설정하여 토어 서킷(Circuit)을 형성한다[1]. 여기서 토어 서킷이란 토어 릴레이 3개(입구 노드, 중계 노드, 출구 노드)가 선택되고, 선택된 세 개의 노드로 이루어진 통신 회로가 구축된 것을 의미한다. 구축된 릴레이를 통해 데이터는 입구 노드부터 차례대로 송신되면서 암호화가 되고, 서버가 요청에 대해 응답을 하면 암호화를 했던 반대 순서로 복호화 된다. 이렇게 순서대로 3중 암호화가 진행되고, 다시 라우터를 거쳐 양파가 한 겹씩 벗겨지듯이 복호화를 수행하기 때문에 이를 Onion routing이라고 부른다. 이러한 3-hop 구성을 통해, 토어 회로를 구성하는 어떠한 노드도 클라이언트와 서버의 IP 주소를 동시에 알 수 없으므로 토어 네트워크를 사용할 경우 익명성이 보장된다. 토어 네트워크에서는 이를 구성하는 노드의 수가 많을수록 익명성이 강화된다고 볼 수 있다. 그림 1,2는 2022년 1월 기준 토어 브라우저 사용자와 릴레이 서버, 브릿지 서버의 통계 자료이다. 사용자는 약 2,500,000명, 릴레이 서버는 약 7,000개, 브릿지는 약 2,000개인 것으로 확인되었다[13]. 이는 2021년 12월과 비교해 모두 증가하였음을 보여준다. 브릿지는 토어가 보장하는 익명성 향상을 위해 운용되는 일종의 토어 노드로[14], DA 또한 브릿지에 대한 정보를 알 수 없다. 이와 같이 많은 수의 토어 노드 및 브릿지가 운용되고 있기 때문에, 토어 네트워크 기반의 다크 웹을 통해 이루어지는 범죄 행위를 네트워크 상에서 추적하는 것은 어려운 문제이다.



〈Figure 1〉 Directly connecting Tor users



〈Figure 2〉 Number of Tor relays and Bridges

또한 웹 사용자가 손쉽게 토어 클라이언트를 통해 네트워크에 접속하게 할 수 있도록 별도의 토어 브라우저를 제공한다. 그림 3은 토어 브라우저와 일반 브라우저에서의 IP를 비교한 것이다. 왼쪽은 토어 브라우저, 오른쪽은 Firefox 브라우저를 나타낸다.

같은 호스트 PC임에도 불구하고, 토어 브라우저를 통해 접속하였을 때 IP가 스웨덴에 있는 것으로 나타난다. 이처럼 토어 브라우저를 사용하면 웹 서버에 찍히는 클라이언트의 IP가 실제 접속 PC와 다른 주소로 나오기 때문에, 사용자가 어느 국가 및 지역에서 어떻게 사용하는지 확인하기 어렵다는 특징을 갖는다.

〈Figure 3〉 IP Comparison between Tor browser and Firebfox browser

3.2. 토어 히든 서비스(Tor hidden service)

토어 히든 서비스는 토어 네트워크의 Onion 라우팅을 통해 웹 서버의 IP 주소를 노출하지 않고 웹서비스를 제공하는 목적으로 활용된다. Chrome, Firefox와 같은 일반 브라우저에서는 접근이 불가능하고, 토어 브라우저에서 .onion 형태로 끝나는 도메인 주소를 통해 접근할 수 있다[2]. 이렇게 토어 브라우저로 특정 주소를 사용해야만 접근할 수 있어 토어 히든 서비스로 불리며, 사실상 이러한 히든 서비스들이 다크 웹을 구성하는 핵심 요소로 볼 수 있다[1]. 히든 서비스의 도메인은 서비스 제공자가 원하는 값이 아닌 랜덤 해시값으로 생성되어 외우거나 알아내는 것이 쉽지 않고, 주소가 비정기적으로 변경되기도 한다. 이러한 Onion 주소의 특징을 악용해 불법 서비스 운영자들이 마약 거래, 음란물 사이트, 아동 성 착취물 등 사이버 범죄의 수단으로 활용하고 있다.

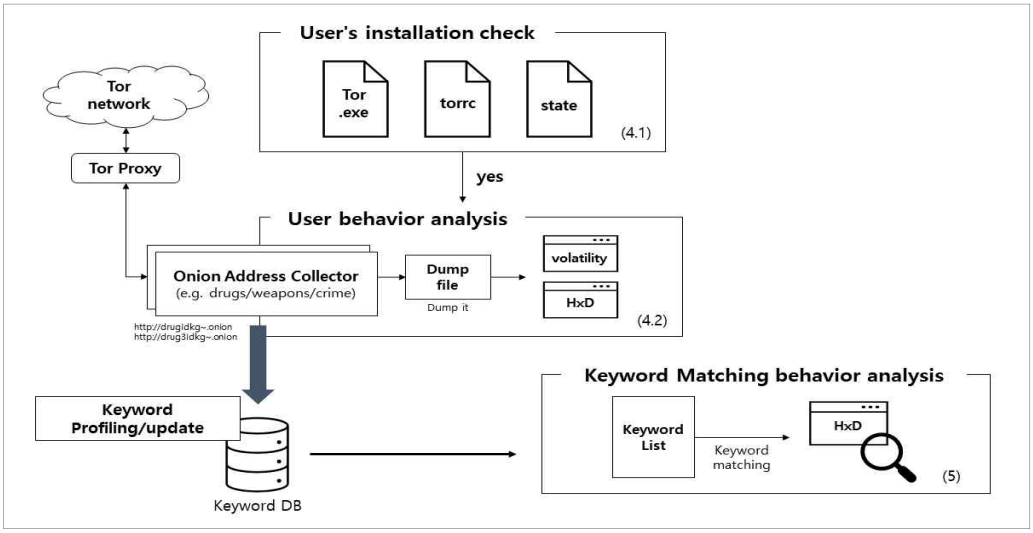
Ⅳ. 메모리 덤프 기반 토어 아티팩트 분석

본 장에서는 메모리 덤프로부터 추출한 토어 아티팩트를 바탕으로 사용자 행위 분석을 수행하기 위한 프레임워크를 제안한다. 그림 4는 제안하고자 하는 전체 프레임워크의 시스템 개요도 및 데이터 흐름도를 나타낸다. 제안한 프레임워크는 크게 3가지 모듈로 구성되는데, 이는 각각 토어 아티팩트 분석, 사용자 토어 행위 분석, 키워드 매칭 기반 행위 분석으로 구분된다. 토어 아티팩트 분석 모듈은 호스트의 토어 설치 여부 및 접속 여부를 확인하는 등 토어 브라우저 사용에 대한 아티팩트 분석을 통해 포렌식 수사 과정에 기본적으로 필요한 정보를 수집한다. 설치 여부가 확인되면 사용자 토어 행위 분석 모듈에서 Onion 주소 수집기로 수집된 범죄 사이트에서 활동한 사용자의 평문 검색 내역을 메모리 덤프 형태의 활성 데이터로부터 추출한다. 이때, 초기 덤프 파일 분석 시 어떤 키워드를 바탕으로 패턴 매칭을 통한 효율적인 행위 분석을 할 수 있는지 모르기 때문에 drug 등과 같은 넓은 범위의 단어를 수집한 사이트들에서 검색하여 얻은 데이터를 처리함으로써 경량성을 갖춘 키워드 DB를 생성하도록 한다. 이를 통해 키워드 매칭 기반 행위 분석 모듈은 효율적인 패턴 매칭을 통해 사용자의 다크 웹을 통한 범죄 행위를 효과적으로 탐지한다.

4.1 사용자 토어 설치 및 사용 분석

본 절에서는 제안하는 프레임워크의 첫 수행 단계인 토어 아티팩트 분석 모듈을 통해 사용자의 토어 브라우저 설치 여부 및 사용 분석 과정을 기술한다. 구체적으로, 익명 브라우저인 토어를 사용할 시 웹 설치 및 활동에 대한 정보가 어느 경로에 어떤 형태로 저장하는지 등의 아티팩트를 분석한다.

분석 대상 OS는 Windows 10이며 호스트의 하드 디스크 드라이브에 남겨지는 아티팩트를 확인하고자 관련 선행 연구의 방법을 바탕으로[3], 토어 브라우저를 설치한 후 웹서핑 및 북마크 설정과 같은 활동을 한 뒤 브라우저를 종료하는 작업을 선행하였다. 표 1은 사용자의 토어 브라우저 설치 여부 및 실행 여부, 북마크 정보를 수집할 수 있는 디스크 내 파일 명 및 해당 파일들의 경로를 나타낸 것이다.



〈Figure 4〉 System overview and its work flow

〈Table 1〉 Tor browser trace file

File name	Path
torrc	C:\Users\User\Desktop\Tor Browser\Browser\TorBrowser\Data\Tor
state	C:\Users\User\Desktop\Tor Browser\Browser\TorBrowser\Data\Tor
TOR.EXE	C:\Windows\Prefetch
Place.sqlite	C:\Users\User\Desktop\Tor Browser\Browser\TorBrowser\Data\Browser\profile.default

4.1.1 torrc

토어 브라우저 설치 여부를 파악할 수 있는 기본적인 아티팩트로는 토어 네트워크에서 설정을 관리하기 위해 사용되는 텍스트 파일인 torrc 파일이 있다. 이는 다양한 항목에 대한 구성을 유지하는 역할을 하는데, 출구 노드를 선택할 국가, 특정 입구 노드 또는 출구 노드의 사용 등 다양한 항목에 대한 환경 설정을 할 수 있다[15]. 포렌식 관점에서 C:\Users\User\Desktop\Tor Browser\Browser\TorBrowser\Data\Tor 경로 속 torrc 파일을 확인하면 토어 브라우저가 설치된 위치와 처음에 토어가 설치되었던 파일의 위치에 대해 알 수 있다. 실제 그림 5의 윗부분(line 1-7)을 보면 토어 브라우저가 설치된 경로를 확인 가능하다. 이를 통해 범죄자의 토어 설치 여부를 확인하여 수사 과정에 메모리 덤프 기반 분석이 필요한 대상인지를 초기에 판단할 수 있다.

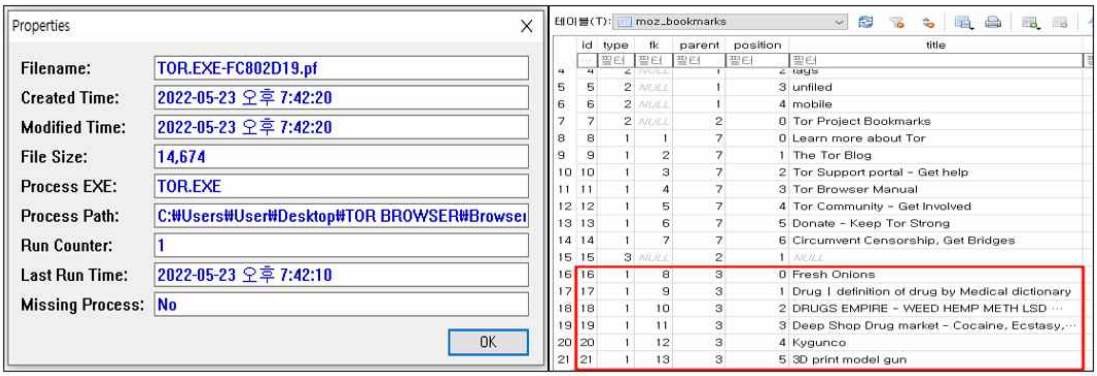
```
# This file was generated by Tor; if you edit it, comments will not be preserved
# The old torrc file was renamed to torrc.orig.1, and Tor will ignore it

ClientOnionAuthDir C:₩Users₩User₩Desktop₩Tor Browser₩Browser₩TorBrowser₩Data₩Tor₩onion-auth
DataDirectory C:₩Users₩User₩Desktop₩Tor Browser₩Browser₩TorBrowser₩Data₩Tor
GeoIPFile C:₩Users₩User₩Desktop₩Tor Browser₩Browser₩TorBrowser₩Data₩Tor₩geoip
GeoIPv6File C:₩Users₩User₩Desktop₩Tor Browser₩Browser₩TorBrowser₩Data₩Tor₩geoip6

# Tor state file last generated on 2022-05-23 11:02:17 local time
# Other times below are in UTC
# You *do not* need to edit this file.
```

〈Figure 5〉 Torrc file and State file

4.1.2 state

사용자가 토어 브라우저를 실행한 시각 정보는 state 파일 분석을 통해 파악할 수 있다. state 파일은 torrc와 거의 유사한 규칙으로 구성되는데, 파일을 작성한 토어 프로그램의 버전과 마지막으로 작성된 시간을 포함한다. C:\Users\User\Desktop\Tor Browser\Browser\TorBrowser\Data\Tor 경로 속의 state 파일을 확인하면 마지막으로 토어 브라우저를 열람했던 시각을 확인할 수 있다. 그림 5의 아랫부분(line 9-11)을 보면 마지막 토어 실행 시각이 협정 세계시(UTC) 형식으로 기록된 것을 확인할 수 있다. 이를 통해 용의자가 토어 네트워크에 언제 접속했는지에 대한 정보 획득이 가능하여 범죄 행위 시간 분석을 진행할 수 있게 된다.

4.1.3 TOR.EXE & place.sqlite



〈Figure 6〉 TOR.EXE and place.sqlite (bookmarks)

WinPrefetchView와 같은 유틸리티를 통해 tor.exe 실행 프로그램을 분석함으로써, 파일 속성 내 프로그램이 설치되고 수정된 시각, 마지막으로 실행된 시각 등의 정보를 파악할 수 있다. 그림 6의 왼쪽을 보면 현재 Tor.exe의 마지막 실행 시각은 2022년 05월 23일 오후 7시 42분 10초로 나타나는데, 해당 정보를 바탕으로 사용자의 마지막 토어 실행 시각을 유추해 볼 수 있다. 또한 그림 6의 오른쪽은 place.sqlite 파일의 스크린샷으로, 토어 브라우저의 북마크 목록을 보여주며, C:\Users\User\Desktop\Tor Browser\Browser\TorBrowser\Data\Browser\profile.default 경로에 존재한다. 북마크는 다음 접속 시 편의를 위해 관심 웹 사이트의 주소를 등록해 놓는 것으로[16], 만약 범죄자가 관심 토어 브라우저 상에서 웹사이트를 북마크로 등록했을 경우 관심 목록과 접속 웹 사이트를 확인함으로써 행위 분석 수사가 가능하다. 그림 6 오른쪽의 북마크 목록을 확인해 보면 해당 사용자는 마약과 무기에 관심이 있다는 것을 예측할 수 있다. 위 두 아티팩트를 통해 사용자의 접속 시간과 관심 목록 확인이 가능했으며 추가적인 사용자 행위 분석 필요성에 대한 여부를 확인할 수 있다.

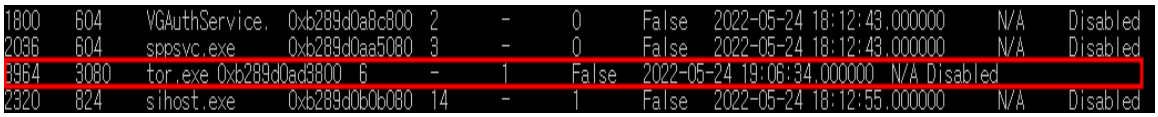
4.2 사용자 토어 행위 분석

본 절에서는 사용자 토어 행위 분석 모듈의 동작에 관해 설명한다. 해당 모듈의 동작은 크게 메모리 덤프 내 활성 데이터로부터 토어 브라우저 접속 여부 확인과 토어 내 검색 기록 추출, 이 두 가지로 구분된다.

4.2.1 Volatility를 활용한 토어 접속 여부 확인

메모리 덤프 기반 사용자 행위 분석의 첫 단계로서 토어 프로세스의 종료 여부를 판단하기 위해, 메모리 포렌식 도구를 통한 활성 데이터 분석을 수행한다. 메모리 덤프를 통해 수사관은 실행 중인 프로세스, 종료 프로세스, 휘발성 조각 및 네트워크 연결과 같은 중요한 정보 분석이 가능해 사용자의 행동에 대한 심층 분석을 하는데 용이하다[12]. 제안한 프레임워크에서는 오픈 소스 메모리 포렌식 도구인 Volatility 툴체인 내 프로세스 확인 기능을 활용한다. Volatility는 Python 스크립트로 작성되었고 Windows, Mac, Linux에서도 실행할 수 있는 대표적인 메모리 포렌식 도구로, 프로세스에서 사용되고 있는 파일, 네트워크 세션 등 프로세스 관련 세부 정보를 확인할 수 있다[17]. 이때, 메모리 분석을 위해서는 덤프(Dump) 과정이 요구되며, 이를 입력으로 활용하여 메모리 포렌식 수행이 가능

하다. 본 연구에서는 활성 데이터 수집을 위해 DumpIt 툴체인을 사용하여 메모리 덤프를 획득하였으며, 이를 Volatility의 입력으로 활용하였다. DumpIt은 메모리 수집 기법에 활용되는 도구로 디지털 포렌식, 침해 사고 대응에 활용되며 Windows 환경에서 물리적인 메모리를 덤프 하는데 사용되는 유틸리티이다. 시스템을 덤프 하는 수사는 오랜 시간이 걸리지만 증거 보존에 있어 효과적인 수사 방식이다. 이후 6장에서 진행될 시나리오 기반 사용자 행위 분석에 불법 행위를 한 용의자가 토어 브라우저를 종료한 상태에서 수사관이 메모리 덤프를 수행하는 것으로 가정하기 때문에 사용한 토어 브라우저가 열려있지 않은 채로 덤프를 수행하였다. 그림 7을 보면 Volatility의 psscan 명령어를 통해 종료된 시스템 프로세스 정보인 토어 프로세스(tor.exe)가 있음을 확인할 수 있다. 이를 통해 용의자가 tor.exe를 실행 후 종료했음을 알 수 있다.

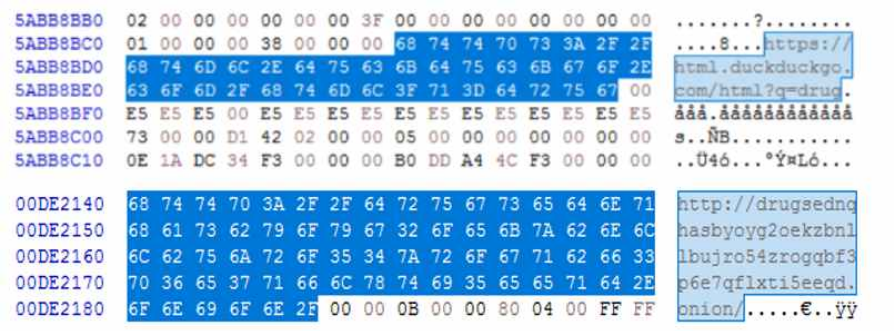


〈Figure 7〉 Confirm that Tor was terminated

4.2.2 HxD를 활용한 토어 내 검색 기록 추출

메모리 덤프 기반 사용자 행위 분석의 두 번째 단계로는 메모리 포렌식 도구를 활용한 활성 데이터 분석 및 사용자 행위 추적이 수행된다. 토어 접속 여부 확인에서와 마찬가지로, DumpIt으로 획득한 메모리 덤프 파일에 대해 패턴 매칭을 수행하여 토어 사용자 행위 분석을 진행한다. 이렇게 메모리에 저장된 데이터는 휘발성임에도 불구하고 평문 데이터에 대한 접근을 가능케 하므로, 웹 기반 포렌식에서 패턴 매칭을 통해 효과적으로 활용될 수 있다. 본 논문에서도 DumpIt을 활용하여 실제 토어 브라우저에서 검색했던 키워드 결과와 접속한 URL을 찾을 수 있음을 확인하였다.

문자열 검색을 통해 토어 사용자 행위를 분석하는 방법은 크게 두 가지로 나뉠 수 있는데, 키워드 기반으로 사용자의 검색 기록을 추적하는 것과 URL 상에 특정 키워드가 포함되었는지로 구분할 수 있다. 전자의 경우, 특정 웹사이트에서 사용자가 검색했던 키워드 검색을 통해 범죄 행위 분석을 진행하는 것이다. 그림 8의 윗부분과 같이 검색엔진 DuckDuckGo에서 키워드 'drug'를 검색했던 행위에 대한 패턴 매칭을 수행해 행위 분석 결과가 나온 것이 그 예로 볼 수 있다. 후자는 URL 자체에 특정 범죄 행위와 관련된 키워드가 포함되어 패턴 매칭에서 발견되는 경우이다. 실제 기존 토어 히든 서비스 URL 탐색 방안 연구에서도 상당수의 URL이 서비스와 관련된 키워드를 포함해 URL을 구성하고 있음을 확인하였다[18]. 예를 들면, 그림 8의 아랫부분과 같이 마약 사이트의 URL에 'drug'가 포함이 되는 경우 이를 기반으로 키워드 매칭을 수행하여 사용자가 접속한 URL을 찾는 행위 분석 수사가 가능하다. 이처럼, 획득한 메모리 덤프에 대해 특정 키워드에 기반한 패턴 매칭을 수행하여 특정 웹사이트 내에서 사용자의 검색 키워드와 접속 URL 탐색이 가능한 것을 확인할 수 있다.



〈Figure 8〉 Search for keyword 'drug' / 'drug' keyword URL

Ⅴ. 키워드 매칭 기반 사용자 행위 분석

덤프 된 메모리 내 활성 데이터에서 어떤 키워드를 검색할 것인가는 포렌식 수사 과정에서 범죄 행위 탐지의 효율성 측면에서 핵심적인 부분이다. 패턴 매칭 수행 과정에서 어떠한 단어들을 키워드로 선정하여 패턴 매칭을 수행하느냐에 따라, 성능 및 시간복잡도 측면에서 비효율적인 수사가 행해질 수 있다. 또한 수사의 시작점인 EntryPoint가 불분명하고, 이를 바탕으로 어떤 범죄 행위를 분석해야 하는지에 대한 어려움이 있는데 패턴 매칭을 위한 키워드 데이터 프레임이 있다면 어떤 형태로 범죄자가 다크 웹을 활용하는지 사용자 행위 분석에 기초하는 데이터로 활용이 가능할 것이다. 또한 키워드 추출을 통해 잠재적 사용자의 식별 및 행위를 판단하여 디지털 증거에 대한 효율적인 분석도 가능해진다[9]. 본 논문에서는 분석 대상 사이트들에서 웹 크롤링을 통해 빈도수가 높은 키워드와 은어를 추출하여 범죄에 활용되는 단어에 대한 통계 분석을 진행함으로써 키워드 매칭 기반 포렌식 수사의 효율성을 향상하고자 한다.

5.1 Onion 사이트 html 크롤링

범죄 행위 탐지에 사용될 데이터 프레임 구축을 위해, 제안한 프레임워크에서는 활성 Onion 사이트로부터 웹 크롤링을 통해 키워드를 수집한 뒤 분석을 수행한다. 토어 브라우저 크롤링의 경우 토어 네트워크의 익명성이라는 특징으로 인해 일반 웹 크롤링 방식이 다르다. 구체적으로, 제시하는 키워드 매칭 기반 행위 분석 모듈은 SOCKS을 통해 프로토콜을 연결해 실행하였다. 여기서 SOCKS란 Socket Secure의 약자로 서버-클라이언트 간의 연결을 직접적인 IP 접근 없이 성공적으로 접근할 수 있도록 하는 TCP 기반 프로토콜이다[19]. SOCKS는 암호화가 없어 빠르게 적용되며 프록시의 방화벽 주위를 회피하여 IP 주소를 숨긴다. 이때, IP 주소를 변경하거나 특정 지역에 접속할 수 없는 콘텐츠에 접근할 수 있으므로 토어 브라우저 크롤링을 위해서는 SOCKS가 필수적이다. 따라서 그림 9에서도 볼 수 있듯이 소켓 객체를 생성하여 진행하였음을 확인할 수 있고, 9150 포트로 프록시를 설정하여 토어 네트워크를 사용해 통신할 수 있도록 하였다.

키워드 매칭 기반 행위 분석 모듈에서는 웹사이트의 html을 크롤링하기 위해 Python 내 Beautiful Soup 패키지를 활용한다. 그림 9는 drug라는 키워드를 통해 마약 관련 사이트 html을 크롤링하기 위해 Beautiful Soup이 사용한 예시 코드를 나타낸다. Beautiful Soup은 Python으로 작성된 도구로 html 혹은 xml 포맷에서 원하는 데이터를 추출하기 위해 일반적으로 사용되는 패키지이다[20]. 코드 내 URL에 html을 추출하고 싶은 Onion 사이트 주소를 입력 후 실행하면, .txt 형식으로 태그를 제외한 html 텍스트 추출 파일이 저장된다. 생성된 파일은 이후 단계인 빈도수 추출 및 데이터 베이스 생성 과정의 입력으로 사용되는 파일이 된다.

```
socks.set_default_proxy(socks.SOCKS5, "localhost", 9150)
socket.socket = socks.socksocket

def getaddrinfo(*args):
    return [(socket.AF_INET, socket.SOCK_STREAM, 6, '', (args[0], args[1]))]

socket.getaddrinfo = getaddrinfo

url = 'http://aapd2gkc7azk5q1w3r4ugy3hm41ahc5k7p761tkllzolwwt7ncc6oyad.onion/'
req = requests.get(url)
html = req.text
soup = BeautifulSoup(html, 'html.parser')
```

〈Figure 9〉 Part of the html crawling code on the drug site

5.2 데이터베이스 생성 및 데이터 시각화

키워드 매칭 기반 사용자 행위 분석을 위한 두 번째 과정은 크롤링한 텍스트 파일에서 키워드 데이터 프레임을 생성하고 통계 데이터를 시각화하는 단계이다. 해당 기능의 구현을 위해, 개발 모듈에서는 Python 내 Pandas와 Matplotlib 라이브러리를 활용하였다. Pandas는 Python 프로그래밍 언어로 작성된 데이터 프레임 생성 및 분석을 위한 도구로써 다양한 데이터 처리 및 가공을 가능케 한다[21].

아래 그림 10은 Pandas 라이브러리를 활용하여 텍스트로 추출된 파일에서 단어를 추출해 빈도수를 확인하여 데이터 프레임을 구축하는 부분의 코드를 나타낸다. 데이터 프레임 구축과 함께, 통계 데이터 결과를 시각적으로 나타내어 사용자의 행위 분석을 위해 필요한 데이터 프레임을 한눈에 파악하기 위해

Matplotlib 라이브러리를 사용한다. Matplotlib란 Python을 활용해 차트나 플롯 형태로 데이터를 시각화해주는 라이브러리로 가장 많이 사용되는 데이터 시각화 패키지이다[22].

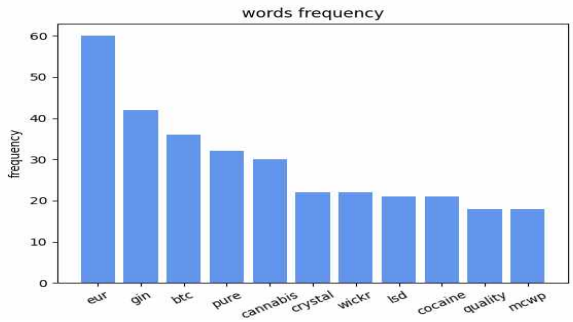
```
df = pd.DataFrame(columns=['words','frequency'])
for words in frequency_list:
    df = df.append(pd.DataFrame([[words, str(freq[words])]], columns=['words', 'frequency']),ignore_index = True)
```

〈Figure 10〉 Top 10 Keyword extraction code part

그림 11과 그림 12는 총 6개의 drug 관련 Onion 사이트에서 단어 빈도수를 분석해 상위 10개의 키워드가 추출된 데이터 프레임 및 이에 대한 시각화 그래프를 보여준다. gin, cannabis, LSD, cocaine 등 대표적인 drug 관련 키워드 결과로 분석된 것을 확인할 수 있다. 특히 마약 판매에서는 일반어를 은어로 사용해 수사망을 피하려는 경우가 많은데, gin, pure, crystal과 같이 각각 Cocain, Heroin, Methamphetamine의 은어로 사용되는 단어들이 나오는 것을 볼 수 있다. 실제로 표 2는 마약 단속국인 DEA(Drug Enforcement Administration)인텔리전스 보고서에서 밝힌 마약 관련 은어들을 정리한 것이다[23]. 다크 웹상에서 마약이 거래될 때는 정상적인 단어가 아닌 은어를 사용해 거래에 활용되는 경우가 많다[24]. 따라서 직접적인 마약 관련 단어뿐만 아니라 은어 수집 데이터도 활용해 실제 Onion 페이지에 존재하는 크롤링 데이터 기반으로 패턴 매칭 수행이 가능해 보다 구체적인 범죄 행위를 탐지할 수 있고, 이를 기반으로 수사 과정에서 빈도수가 높은 키워드 중심으로 사용자의 행위를 추적할 수 있다. 또한, 덤프한 메모리 내에서 키워드 매칭으로 사용자의 흔적을 찾고, 해당 메모리 주소 근처의 메모리 영역들까지 탐색함으로써 분석 영역의 범위를 넓히는 수사도 진행할 수 있다. 이를 통해 토어 브라우저 사용자의 활동 정보 및 html 페이지에 대해 분석이 가능하게 된다.

	words	frequency
14	eur	60
19	gin	42
6	btc	36
39	pure	32
7	cannabis	30
9	crystal	22
50	wickr	22
27	lsd	21
8	cocaine	21
40	quality	18
28	mcwp	18

〈Figure 11〉 Top 10 keywords



〈Figure 12〉 Visualizing top 10 keywords

〈Table 2〉 Drug Slang Code Words

Drug	Drug Slang Code Words
Cocaine	Gin, Clear Kind, Brisa, Cement, 777, Love Affair, Old Lady, Girl, Girlfriend, Happy Powder, T-Shirts, Turkey
Heroine	Pure, Bad seed, Coco, Junk, White Girl, White Junk, White Lady, White Nurse, Antifreeze, Big Bag, Beast
Methamphetamine	Crystal, Glass, Salt, Water, Windows, Aqua
LSD	Chinese Dragons, Coffee, Blue Chair, White Lightening

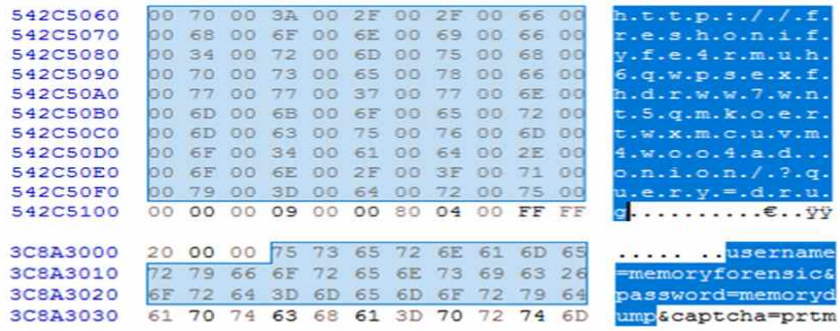
Ⅵ. 범죄 시나리오 시뮬레이션을 통한 키워드 매칭 기반 행위 분석 실효성 검증

제안한 데이터 프레임 실효성 검증을 위해 범죄 행위를 추적하는 시나리오를 설정하고, 직접 범죄 행위를 시뮬레이션했을 때 키워드 매칭 기반 행위 분석을 활용한 가상의 수사가 이를 올바르게 탐지하는지에 대한 평가를 수행하고자 한다. 표 3은 사용자가 토어 브라우저에 접속하여 행한 범죄 시나리오를 가상으로 작성한 것이다. 우선 사용자가 토어 네트워크에 접속하여 웹사이트에서 'drug'를 검색하고, 이후 로그인과 장바구니 품목 담기를 진행한 순으로 시나리오를 가정하였다. 여기서 용의자는 시나리오 행위를 완료한 뒤 증거 인멸을 위해 토어 브라우저를 닫았다고 가정한다. 이 시점에서 수사관의 입장으로 메모리 덤프가 수행되는

것이며 앞서 메모리 덤프 기반 스트링 매치에서 보았던 것과 같이 4.1과 4.2장의 툴 체인이 구비된 환경에서 메모리 덤프 형태의 활성 데이터로부터 패턴 매칭을 수행하는 가상의 범죄 행위 추적 포렌식 수사를 진행한다.

〈Table 3〉 The procedure of crime scenario using Tor

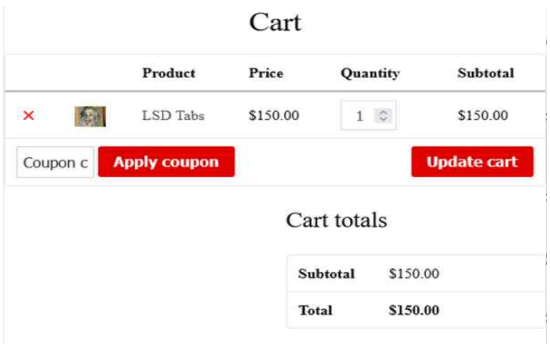
1. Connect to Tor
2. Search 'drug' on Freshonion website.
3. Log in on the site (http://7b···uod.onion/)
4. Search keywords on the site (cocain, LSD)
5. Putting in the shopping cart.
6. Close Tor Tabs. → **(Criminal investigator) Dump Memory**



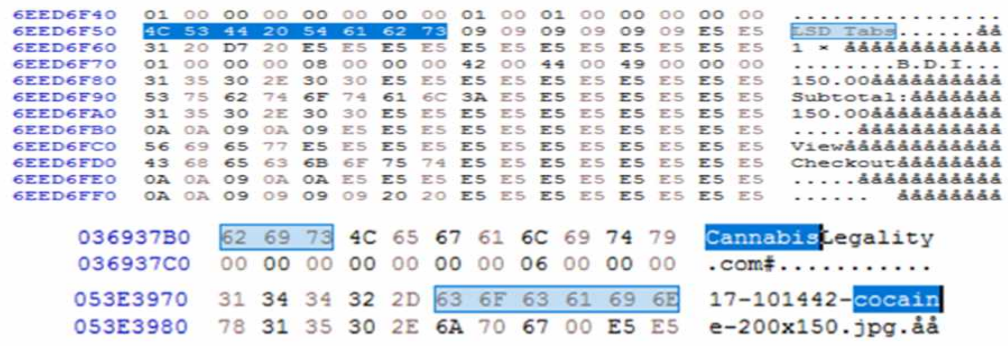
〈Figure 13〉 The drug site url / User ID and password

제안한 프레임워크 내 사용자 토어 행위 분석 모듈을 통해, 사용자가 마약 웹사이트에서 검색한 'drug' 키워드가 탐색 된 것을 그림 13의 윗부분에서 확인할 수 있다. 이를 통해 사용자가 마약 사이트를 접속하였다는 행위 추적이 가능하다. 또한, 그림 13의 아랫부분에서는 사용자가 접속한 사이트의 ID와 비밀번호 평문 데이터를 확인할 수 있다. 비밀번호 값은 password 키워드와 함께 key-value pair 형태로 데이터가 관리되기 때문에[5], 범죄 행위를 행한 용의자의 추가적인 행동 추적에도 해당 정보를 활용할 수 있다.

다음으로, 제안한 키워드 매칭 기반 행위 분석 모듈이 가상 시나리오에서 토어 브라우저 사용자가 마약 사이트에 접속해 어떠한 검색을 수행했는지 특정하기 위한 행위 추적을 효과적으로 수행하는지에 대해 검증한다. 그림 14에 나타내는 바와 같이, 가정한 범죄 시나리오에서 이용자는 'LSD Tabs'라는 마약의 종류를 장바구니에 담는 행위를 토어 브라우저상에서 수행하였다. 이를 제안한 프레임워크를 통해 패턴 매칭 수사를 진행한 결과, 크롤링을 통해 구축한 마약 관련 데이터 프레임(그림 11, 12 참고) 내에 'LSD' 키워드가 존재하며, 이에 따라 'LSD Tabs'가 그림 15의 윗부분과 같이 패턴 매칭 과정에서 탐지되었음을 확인할 수 있다. 또한, 그림 15의 아랫부분에서도 분석된 상위 10개의 키워드 중 하나인 'Cannabis', 'cocain' 이 검색되면서 토어 사용자의 행위 분석 및 범죄 행위가 탐지될 수 있었음을 보여준다. 다시 말해, 제안한 프레임워크를 활용함으로써 사용자가 마약 사이트를 이용한 범죄자임을 특정할 수 있다.



〈Figure 14〉 Putting items in the shopping cart



〈Figure 15〉 Searching shopping cart items / Top 10 keyword-based string matching

Ⅶ. 결론

본 연구에서는 토어 설치 및 접속 여부 확인 및 메모리 덤프 기반 포렌식을 통해 토어 브라우저 내 사용자의 행위 정보 및 아티팩트를 획득할 수 있는 프레임워크를 개발하였다. 기존 연구와 다르게 다크 웹 내 판매자들의 Onion 도메인 바탕으로 키워드가 선정된 것이기 때문에 이는 실질적인 범죄 행위의 패턴 매칭 수사 효율성을 올릴 수 있다. 이렇게 키워드 빈도수를 기반으로 만든 데이터 프레임을 활용해 패턴 매칭을 수행함으로써 가상 시나리오의 범죄자 행위 탐지가 가능함을 확인하였다.

분석 사이트 대상으로 빈도수가 높은 키워드를 추출해 만든 데이터 프레임은 포렌식 수사에 필요한 데이터를 범주화한다는 측면에서 3가지 의의가 있다. 첫째로, 데이터 범주화는 향후 실제 발생한 범죄와 관련해 어떤 키워드들이 범죄에 많이 활용되었는지 통계자료를 생성할 수 있다. 또한, 수집한 데이터베이스를 기반으로 수사를 진행하여 ISO와 메모리 덤프에서 빈도수가 높게 추출된 키워드를 바로 검색해 사용자 행위를 추적할 수 있다. 마지막으로, 크기가 큰 메모리의 모든 키워드에 대해 패턴 매칭 수행하는 것은 오버헤드가 크기 때문에 빈도수 기반의 키워드 패턴 매칭을 수행하여 효율적으로 범죄 행위를 도출해 낼 수 있다. 이러한 데이터 범주화의 의의는 다크 웹 내 사이버 범죄가 증가하고 있는 상황에서 사용자의 범죄 행위 패턴 분석을 하는데 중요한 요점이 된다. 본 논문에서는 마약 카테고리에 대해서만 데이터 프레임을 구축해 분석을 수행하였다. 그러나 제안한 수집기의 자동화로 메모리 덤프 이미지만 확보되면 사용자 행위 분석을 위한 크롤링 기반으로 다른 카테고리 내의 검색어를 매칭시켜 어떤 불법 행위를 했는지를 알 수 있는 대규모 수사도 가능할 것이다. 따라서 향후 연구로는 마약 카테고리에서만 한정된 것이 아닌, 여러 범죄 관련 카테고리 내에서 키워드 매칭 기반 수사에 관한 사례연구를 수행하고자 하며 수사망을 피하기 위해 주기적으로 변경되는 은어 분석도 함께 진행해 제안한 프레임워크의 고도화를 통해 수집 성능 및 정확도 개선을 진행하고자 한다.

참 고 문 헌 (References)

[1] Gyungbin Lee, "A Study on Tor-Hidden Service Server Forensic Case and Tracking Technique", Digital Forensic Research, 12(1), Jun.2018, pp.37-47.

[2] Dingledine, Roger and Mathewson, Nick and Syverson, Paul, "Tor: The Second-generation Onion Router", Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13, pp 1-17, 2004.

[3] Mina Seo and Sangjun Park, "Web browser forensic analysis technique.", Department of Mathematical Information Science, Graduate School of Convergence Science and Technology, Seoul National University, Feb.2018.

[4] https://blog.torproject.org/browser-fingerprinting-introduction-and-challenges-ahead/

[5] Dongwon Shin and Taeshik Shon, "Analysis of personal identification information exposure using memory dump-based forensics.", Digital Forensics Research, 15(2), Jun.2021, pp. 35-49.

[6] Jonghyeon Park, Youngseok Lee, "Tor Hidden Service Collector Using Out-of-band Exploration", The Korean Institute of Information Scientists and Engineers, June.2017, pp.1190-1192(3 pages).

[7] Hyunsu Mun, Soohyun Kim, Youngseok Lee, "Improvement in Tor-based Dark Web Crawling Performance by Eliminating Web Browser Rendering and Scripting Tasks", Journal of KIISE 47(10), Oct.2020, pp.1008-1013 (6 pages).

[8] Priya P Sajana, C. Balanb, M.J. Devi Priyac, A.L. Sreedeep, "TOR Browser Forensics", Turkish Journal of Computer and Mathematics Education Vol.12 No.11,2021.

[9] W.J.C. van Staden, E. van der Poel, "Using Automated Keyword Extraction to Facilitate Team Discovery in a Digital Forensic Investigation of Electronic Communications", SAIEE Africa Research Journal, Volume 108, Issue 2, Jun.2017.

[10] Nicole Lang Beebe, Jan Guynes Clark, "Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results", Digital Investigation, Volume 4, Supplement, September 2007, Pages 49-54.

[11] S. M. HejaziC. TalhiM. Debbabi, "Extraction of forensically sensitive information from windows physical memory", Digital Investigation, Volume 6, Supplement, Sept. 2009, Pages S121-S131.

[12] MalaK Alfosail, Peter Norris, "Tor forensics: Proposed workflow for client memory artefacts", Computers & Securuty, Volume 106, July.2021.

[13] https://metrics.torproject.org/userstats-relay-country.html

[14] https://bridges.torproject.org/

[15] Konstantinos Stefanidis - Vozikis, "A Distributed Performance Measurement Tool for Tor Browser", KU LEUVEN, 2020-2021.

[16] https://terms.naver.com/entry.naver?docId=816615&cid=42344&categoryId=42344

[17] Jisung Han and Sangjin Lee, "Windows Physical Memory Analysis Tools for Live Forensics", Journal of the Korea Institute of Information Security & Cryptology 21(2), Apr.2011, pp.71-82.

[18] Sungha Park, "An Approach to Discover Hidden Services in Tor Network", The Korean Institute of Information Scientists and Engineers, Dec.2018, pp.1006-1008(3 pages).

[19] Seok-Chul Baek and Seung-Min Park, "Windows NT Firewall Construction with SOCKS V5", Proceedings of Symposium of the Korean Institute of communications and Information Sciences, Jan.1997, pp.358-366(9 pages).

[20] JuWon Seong ,Sung Hyun Kim, Sang-Chul Kim, "A Study on the Efficient Crawling Techniques using Python Libraries", Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp.988-989(2 pages), 2019.

〔21〕 Hyo-Kwan Kim and Won-Yong Hwang, “Proposal For Improving Data Processing Performance Using Python”, Journal of Korea Institute of Information, Electronics, and Communication Technology 13(4), Aug.2020, pp.306-311(6 pages).
〔22〕 Hwan-Soo Kang and Hee-Chern Kim, “A Design on Deep Learning Lecture for Computer Programming Education”, Journal of Digital Contents Society 21(10),Oct.2020, pp. 1801-1808(8 pages).
〔23〕 Slang Terms and Code Words: A Reference for Law Enforcement Personnel, July.2018.
〔24〕 J. Tuomas Harviainen, Ari Haasio, Lasse Hämäläinen. “Drug Traders on a Local Dark Web Marketplace”, Academic MindTrek, Jan.2020.

저 자 소 개

김 예 은 (Yeeun Kim)
준회원

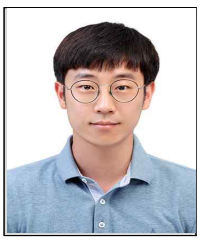
2019년 3월~현재 : 성신여자대학교 융합보안공학과 학사과정
관심분야 : 디지털 포렌식, 다크 웹, 시스템 보안

도 예 진 (Yejin Do)
준회원

2019년 3월~현재 : 성신여자대학교 융합보안공학과 학사과정
관심분야 : 디지털 포렌식, 시스템 보안, 소프트웨어 보안

한 상 연 (SangYeon Han)
준회원

2019년 3월~현재 : 성신여자대학교 융합보안공학과 학사과정
관심분야 : 디지털 포렌식, 시스템 보안, 정보보호



김 성 민 (Seongmin Kim)
준회원

2012년 2월: 한국과학기술원 전기 및 전자공학과 졸업
2014년 2월: 한국과학기술원 전기 및 전자공학과 석사
2019년 2월: 한국과학기술원 정보보호대학원 박사
2019년 9월~2020년 8월: 삼성전자 삼성리서치 Staff Engineer
2020년 9월~현재: 성신여자대학교 융합보안공학과 조교수
관심분야 : 신뢰 실행 환경, 클라우드 컴퓨팅, 시스템 보안