Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.28, NO.4, Aug. 2018 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2018.28.4.869

사이버 공간 내 디지털 증거 수집 시스템에 관한 연구

정 효 정,† 최 종 현, 이 상 진‡
고려대학교 정보보호대학원

A Study on Digital Evidence Collection System in Cyberspace

Hyojeong Jeong,† Jong-hyun Choi, Sangjin Lee‡
Institute of Cyber Security & Privacy (ICSP), Korea University

요 약

사이버 공간 내 디지털 증거 데이터는 수정 및 삭제되기 쉬우며 실시간으로 변경사항이 반영되므로 사건 발생 시점 이후 증거 데이터의 빠른 획득이 필요하다. 클라이언트 측에서의 증거 수집은 별도의 행정절차로 인한 시간 지연 없이 데이터를 획득할 수 있다는 장점이 있지만, 대용량 데이터의 수집에 있어서는 마찬가지로 수집 시간 지연 문제에 취약하다. 따라서 본 논문에서는 사이버 공간 내 주요 웹 기반 서비스를 중심으로, 클라이언트 측면에서의 자동화 된 증거 수집 방식을 제안하여 대용량 데이터에 대한 효율적인 증거 수집이 가능하도록 한다. 나아가 제안한 방식을 사용하고 수집한 디지털 증거의 법정제출시점까지의 무결성을 보장하는 사이버 공간 내 디지털 증거 수집 시스템을 제안한다.

ABSTRACT

Digital Evidence Data in cyberspace is easy to modify or delete, and changes are reflected in real time, so it is necessary to acquire evidence data quickly. Collecting evidence on the client side is advantageous in that data can be acquired without time delay due to additional administrative procedures, but collection of large data is likewise vulnerable to collection time delay problem. Therefore, this paper proposes an automated evidence collection method on the client side, focusing on the major web-based services in cyberspace, and enables efficient evidence collection for large volumes of data. Furthermore, we propose a digital evidence collection system in cyberspace that guarantees the integrity of the collected digital evidence until the court submission.

Keywords: Digital Forensics, Digital Evidence, Data Collection

I. 서 론

사이버 공간에서 디지털 데이터는 영구적이지 않고 생성 시점 이후에도 삭제 및 수정되어 변경되기 쉬운 특성이 있다. 이러한 특성상 시간 요소는 사이버 공간 내 데이터를 증거로 사용하기 위해 고려되어야 할 중요한 요소로서 디지털 포렌식 조사 시 특정 시점에 특정 데이터가 특정 위치에서 존재했다는 것이 입증될 필요가 있다. 데이터의 생성부터 데이터가 증거로서 법정에 제출되기까지의 시간 진행을 Fig.1.과 같이 나타낼 수 있다. 사이버 공간 내 디지털 데이터를 특정 사건의 증거로서 사용하기 위해서는 Fig.1.의 사건 발생 시점에 존재하던 데이터를 수집하고, 수집된 데이터가 법정 제출 시점까지 변경되지 않았다는 것을 입증할 수 있어야 한다. 따라서 변경되기 쉬운 디지털 데이터의 특성상 사건 발생 시점(A)과 증거 수집 시점(B) 사이의 시간 지연을 최소화하고 증거 수집 이후에는 법정 제출 시점(C)까

Received(04. 27. 2018), Modified(07. 23. 2018), Accepted(08. 02. 2018)
† 주저자, jjeongh55@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr(Corresponding author)

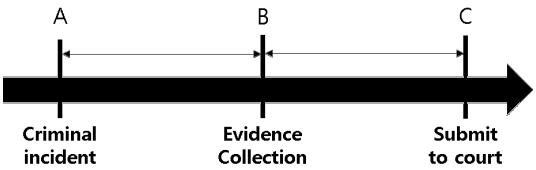

Fig. 1. The process of which data to be evidence

지 수집된 증거 데이터를 보존하는 것이 중요하다.

증거 수집 시점에서 사이버 공간 내 디지털 데이터의 수집은 데이터가 실제로 저장되어 있는 서버로부터 직접 수집하는 방식과 네트워크 통신 등을 사용하는 클라이언트를 통해 수집하는 방식이 있다. 이를 통해 수집한 디지털 데이터의 증거화를 위해 수집 데이터의 객관성 보장, 투명성 만족 등을 위한 수집 과정에서의 엄격한 절차가 필요하지만, 이로 인한 수집 시간의 지연 문제가 발생할 수 있다.

먼저 서버로부터 직접 데이터를 수집할 경우 행정 절차 상 서버를 소지하고 있는 서비스 제공자의 협조 요청이 필요하고, 더 나아가 서버가 해외에 위치한다면 국제 공조 요청이 필요할 수 있다. 이 과정에서 Fig.1.의 A~B와 같은 사건 발생 시점과 증거 수집 시점 사이의 시간 지연이 발생한다. 다만 이 문제는 최근 국외에 위치하는 이메일 서버 압수수색에 관한 대법원 판례[1]에 따라 확보된 접근 권한을 이용한 증거 데이터 수집 방식의 활용을 통해 일부 해결 가능할 것으로 보인다.

그러나 클라이언트 측에서의 증거 데이터 수집 방식을 사용하더라도, 조사자의 수집 과정 중 조작 가능성의 불식이 어려우며 꾸준히 증가하고 있는 사이버 공간 내 존재하는 데이터에 의한 대용량 데이터 수집 가능성을 염두에 둘 때 Fig.1.의 B에 해당하는 증거 수집 시점에서의 시간 지연을 피할 수 없다. 또한 증거 수집 이후 수집된 사이버 공간 내 디지털 데이터는 증거로서의 효력을 갖고 이를 유지하기 위해 Fig.1.의 B~C와 같은 법정 제출 시점까지 변경되지 않았음을 보일 수 있어야한다.

본 논문에서는 사이버 공간 내 웹 기반 서비스를 중심으로, 대용량 데이터에 대한 클라이언트 측면에서의 효율적인 증거 수집 방식을 제안한다. 나아가 제안한 방식을 사용하여 수집한 디지털 증거의 법정 제출시점까지의 무결성을 보장하는 사이버 공간 내 디지털 증거 수집 시스템을 제안한다.

II. 관련 연구

2.1 기존 디지털 증거 수집 시스템

국립과학수사연구원에서는 수사기관이 채증한 디지털 증거물에 대한 조작 논란을 차단할 수 있는 디지털 증거물 인증 서비스를 개발하여 활용하고 있다[2]. 현장에서 촬영 또는 녹음한 동영상, 사진, 음성 등의 전자 파일과 함께 전자 지문(hash), 시간 정보, 위치 정보, 사용자 ID를 함께 국과수 인증 서버로 전송하면, 인증 서버는 원본임을 증명하는 인증서를 발급해주어 제3자 인증을 제공한다.

강주영[3]은 웹 게시물 형태의 디지털 증거와 관련하여, 일반 사용자가 웹 사이트의 악의적인 게시물을 화면 캡처를 통해 신고할 수 있는 서비스를 제안하였다. 웹 브라우저에 플러그인(plug-in) 형태로 존재하는 Proxy Browser 프로그램을 통해 사용자가 직접 웹 화면을 캡처 후 신뢰할 수 있는 제 3의 기관에 해시 값, PC 정보 등이 포함된 화면 캡처 증거 팩 형태로 전달한다. 비공개 그룹 내 게시물을 수집 가능하다는 점에서 기존 시점 인증 및 공증 서비스들에 비해 증거 수집이 용이하다.

김아름 외 1명[4]은 웹 게시물에 대한 공증 요청을 통해 웹 게시물이 법적 효력을 갖기 위한 공증 시스템을 제안하였고, 이 시스템이 도입될 수 있는 법률적 제도적 요구사항을 확인하였다.

이러한 기존 디지털 증거 수집 시스템들은 대용량 데이터의 수집에는 특화되어 있지 않고, 특정 웹 게시물을 대상으로 한 증거화에 초점이 맞추어져 있다. 따라서 본 논문에서는 기존의 연구와는 디지털 증거 수집 방식을 달리한 시스템을 제안한다.

2.2 사이버 공간 내 데이터 수집 및 보존

사이버 공간 내 데이터를 수집하는 기술은 데이터의 유형 및 특성에 따라서 달라진다. 한국정보화진흥원에서는 '빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼'[5]을 통해 크롤링, FTP, Open API, RSS 등의 다양한 데이터 수집 기술을 소개하였다. 그 중 클라이언트 측에서 사이버 공간 내 데이터를 수집하기 위한 데이터 수집 방법으로 크롤링과 Open API를 선택할 수 있다. 크롤링은 주로 웹 문서를 수집하는 데에 이용되며 Open API의 경우 실시간 데이터 수집에 용이하다.

사이버 공간 내 데이터를 보존하는 기술은 기록물 보존의 입장에서 대표적으로 웹 아카이빙 기술을 통해 꾸준히 연구되어 왔다. 수집해야 하는 웹 사이트의 양이 방대함에 따라 아카이브 서비스들은 수집 단계에서 웹 크롤러와 같은 소프트웨어를 활용하고 있다. 이러한 웹 아카이빙 기술은 시간이 지남에 따라 삭제될 수 있는 사이버 공간 내 데이터를 보존한다는 점에서 특정 시점에 특정 데이터가 존재했음을 확인할 수 있는 데에 그 의의가 있다.

다만 지금까지의 사이버 공간 내 데이터의 수집 및 보존을 위한 기술은 빅데이터 활용을 위한 수집을 위해 사용되거나, 웹 기록 아카이브를 통한 데이터 보존에 목적이 있고 이를 통한 증거화에는 그 목적을 가지고 있지 않다. 따라서 디지털 포렌식 관점에서 사이버 공간 내 데이터의 수집과 보존에 대한 추가적인 연구가 필요하다.

III. 사이버 공간 내 데이터의 분류

3.1 공개 범위에 따른 데이터 종류

사이버 공간에서 클라이언트 측의 증거 수집을 진행할 경우 서버로부터 네트워크 통신을 통해 전달 받은 데이터만을 수집할 수 있다. 서버는 저장된 모든 데이터를 사용자에게 공개하지 않으며 사용자의 접근 권한을 확인하여 접근 가능한 데이터만을 사용자에게 전달한다. 따라서 데이터의 공개 범위에 따라 수집되는 증거 데이터를 모두에게 공개된 데이터와 특정 개인에게 속해있는 데이터로 구분할 수 있다. 특정 사용자의 웹 사용 흔적을 조사하기 위해서는 공개 데이터(public data)를 넘어서서 개인 데이터(private data)를 수집하여 증거화 할 필요가 있다.

3.1.1 공개 데이터 (public data)

특별한 접근 권한을 요구하지 않고 공개 되어 있는 데이터는 누구나 열람할 수 있다. 이러한 데이터들은 이미 많은 검색 엔진들에 의해 인덱싱 되어 있다. 사이버 공간의 큰 부분을 차지하는 월드 와이드 웹의 구분에서 표면 웹에 존재하는 데이터들을 공개 데이터라고 볼 수 있으며, 포털 사이트의 메인 페이지나 개인 SNS 계정의 전체 공개 게시물 등이 해당된다. 공개 데이터는 수집 대상의 계정 정보를 모르고, 접근 권한을 따로 획득할 수 없어도 데이터의 접근 및 획득이 가능하다.

3.1.2 개인 데이터 (private / semi-private data)

공개 데이터에 반하여 사용자가 적절한 접근 권한을 가지고 있을 때에만 열람이 가능한 데이터를 개인 데이터라고 정의한다. 이러한 개인 데이터는 표면 웹이 아닌 심층 웹의 범주에 해당하여 검색 엔진 등을 통해 확인할 수 없다.

사회 관계망 서비스(SNS)의 이용이 증가하면서 개인 데이터에 대한 세분화된 구분이 필요하다. 예를 들어 비공개 그룹 내 게시물은 기본적으로 공개되어 있지 않지만 사용자가 비공개 그룹에 소속되어 적절한 접근 권한을 가지고 있을 경우 그룹 내 게시물에 접근할 수 있다. 이 경우에는 수집 대상의 계정 정보를 모르더라도 관계망 내 제 3자의 계정 정보를 통해 데이터의 접근 및 획득이 가능하다.

그러나 작성자 본인만 볼 수 있도록 설정한 글이나, 개인 클라우드 계정 내 파일 목록 등 제 3자에게 공유되지 않은 데이터의 경우에는 수집 대상의 계정 정보를 알고 있을 경우에만 데이터의 접근 및 획득이 가능하다.

3.2 서비스 유형에 따른 데이터 종류

증거 수집 과정에서 존재하는 모든 데이터를 수집하기에는 그 양이 방대하고, 수집 후 분석 과정에서의 추가적인 데이터 정제가 필요하다. 또한 압수 수색 과정에서는 영장의 범위에 따라 수집 시 전체 수집이 아닌 선별 수집을 진행할 필요가 있다.

Table 1. Information and data type included by Web based Services

category	info(type)
email	- mail list (text) - email (text, image / eml) - attachment (file)
sns	- user profile (text, image) - user posts (text, image, video) - user interaction (text)
cloud storage	- file list (text) - file (file) - cloud document (file)
etc	- web posts (text + α)

그러나 사이버 공간 내에는 디지털 데이터를 생성하는 다양한 유형의 서비스들이 존재하고, 각 서비스별로 데이터의 유형 및 획득 가능한 정보가 상이하다. 서비스 별로 어떠한 데이터가 존재하며 수집의 필요성이 있는지 사전에 정의하는 과정은 증거 수집 과정을 용이하게 할 수 있다. Table 1.은 주요 웹 기반 서비스에서 획득할 수 있는 주요 정보 및 데이터 유형을 간단히 나열한 것이다.

3.3 사이버 공간 내 데이터의 증거능력

증거능력이란 증거가 엄격한 증명의 자료로 사용될 수 있는 법률상의 자격을 말한다[6]. 따라서 제출된 증거가 채택되어 법정에서 주장하는 내용을 뒷받침하기 위해서는 증거능력을 만족해야만 한다[3].

앞서 구분한 사이버 공간 내 데이터는 디지털 증거의 형태로 존재하며 이러한 증거가 증거능력을 갖기 위해서는 첫째, 증거 수집의 절차가 적법하게 이루어져야 하고 둘째, 진정성, 무결성, 신뢰성, 원본성이 보장되어야 하며 셋째, 전문법칙의 요건을 만족해야한다[7].

이중 전문법칙의 예외에 해당하는지 여부는 형사소송법 313조의 '성립의 진정을 부인하는 경우에는 과학적 분석결과에 기초한 디지털 포렌식 자료, 감정 등 객관적 방법으로 성립의 진정함이 증명되는 때에는 증거로 할 수 있다'는 개정안을 통해 다소 해결 가능하게 되었다.

본 논문에서는 첫 번째와 두 번째의 요구사항을 만족할 수 있는 디지털 증거 수집 시스템을 제안하여 사이버 공간 내 데이터가 증거능력을 가질 수 있도록 한다.

IV. 사이버 공간 내 증거 수집 시스템 설계

4.1 웹 크롤링과 Open API의 활용

사이버 공간 내 웹 사이트 정보를 수집하고 이를 보존하는 프로세스에 대한 연구는 대표적으로 웹 아카이빙 기술의 관점에서 진행되었다. The National Archive의 'Web archiving guidance'[8]에서는 이러한 웹 아카이빙의 유형을 client side, transaction based, server side의 세 가지로 구분하고 있는데, client side archiving은 가장 널리 사용되는 접근 방식으로 주로 웹 크롤러를 이용하여 서버에서 전송되는 응답을 수집한다. 웹 크롤러를 사용하여 데이터를 수집할 경우 대량의 데이터를 짧은 시간에, 적은 조작 횟수로 수집할 수 있다. 특별히 사전에 수집 대상의 구조분석을 통해 수집하고자 하는 데이터만을 찾아가 획득하는 래퍼(wrapper) 기반의 웹 크롤러를 활용하면 수집하고자 하는 데이터를 미리 정하여 정확하게 수집할 수 있다.

Open API란 사용자가 직접 응용프로그램과 서비스를 손쉽게 개발할 수 있도록 공개된 API로서, 주요 웹 기반 서비스들은 Open API를 통해 서비스 이용자로 하여금 서비스를 쉽게 활용할 수 있도록 한다. 각 서비스별 API 제공 범위에 따라 수집 범위는 달라질 수 있으나 API가 제공하는 데이터 읽기 기능을 이용한 응용프로그램 개발을 통해 간단하게 데이터 수집이 가능하다.

따라서 웹 크롤러 또는 Open API를 사용한 증거 수집 단계의 자동화를 통해 수집하고자 하는 증거 데이터가 대량이어도 조사자가 직접 수집하는 방식에 비해 빠르게 수집이 가능하여 증거 수집 단계에서의 시간 지연을 해소할 수 있으며, 조사자의 증거 수집 중 임의 조작 또는 실수에 의한 증거 데이터 훼손 가능성도 최소화 할 수 있다.

4.2 제안 시스템의 요구 사항

제안 시스템은 증거 수집 단계의 자동화를 통한 효율적인 증거 수집을 진행하고, 시스템을 통해 수집한 증거 데이터가 증거 능력을 가지기 위하여 다음과 같은 요구 사항을 만족해야 한다.

첫째, 시스템은 누구나 획득할 수 있는 공개 데이터를 포함하여 개인 데이터의 수집도 가능해야 한다. 조사과정에서 적법하게 획득한 사용자 계정 정보를 통해 접근 권한을 얻은 후 접근 권한이 존재할 때만 얻을 수 있는 개인 데이터를 수집할 수 있어야 한다. 따라서 시스템에서는 기본적으로 계정 정보를 입력하는 인증 단계를 수집 단계 이전에 포함하여 접근 권한을 얻은 후 증거를 수집한다.

둘째, 시스템은 증거 수집 단계에서의 시간 지연 문제를 해결하고 조사자 개인의 역량이나 외부 환경에 최소한의 영향을 받아야 한다. 기존 사이버 공간 데이터에 대한 증거 압수수색 과정의 경우 조사자가 직접 영장의 범위에 해당하는 데이터를 선택하여 획득함으로써 조작의 실수를 피하기 어렵고, 시간이 오래 걸리는 문제가 있었다. 따라서 시스템에서는 이러

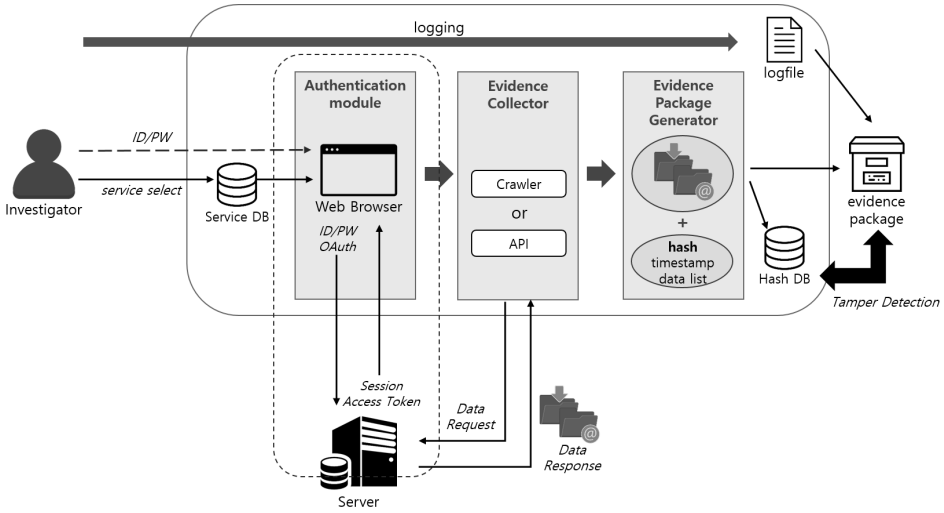

Fig. 2. Digital Evidence Collection System

한 영향을 최소화하기 위해 데이터 수집 과정을 자동화 한다. 또한 조사자의 증거 수집 요청부터 증거 팩 생성까지의 모든 단계는 로그 파일에 기록되어 적합한 절차를 수행하였는지 확인하기 위해 보관된다. 시스템이 안전하다면 수집된 증거 또한 훼손되지 않음을 보일 수 있다.

셋째, 시스템은 수집된 디지털 증거 데이터가 수집 시점과 법정 제출 시점 사이에 위변조 되지 않았음을 보여야 한다. 수집된 증거 데이터는 추후 증거 제출 시점까지 보관되는데, 이 때 수집 과정에서의 무결성 뿐만 아니라 보관 과정에서의 무결성 또한 보장되어야 한다. 따라서 시스템에서는 수집 완료 후 해시함수를 통해 생성한 해시 값을 시스템과, 수집한 디지털 증거 양쪽에 모두 저장 후 증거 제출 시점에서 비교할 수 있게 함으로써 보관 과정에서 증거 데이터가 위변조되지 않았음을 보일 수 있다.

넷째, 시스템은 제3의 인증기관을 통한 인증을 통해 시스템의 신뢰도를 인증 받아야 한다. 시스템의 모든 구성요소는 데이터의 위변조 가능성에 안전하며 각 요소 간 통신은 SSL과 같은 신뢰할 수 있는 프로토콜을 사용해야 한다. 또한 서비스 구조나 API의 변경에 따라 주기적으로 시스템 운용에 대한 관리 및 업데이트가 이루어져야 한다.

4.3 제안 시스템의 구성 요소

제안하는 증거 수집 시스템은 4.2절의 기술적 요구사항을 만족하도록 설계되었다. 시스템은 증거 수집을 진행하는 조사자, 증거 데이터에 대한 접근권한을 획득하는 인증 모듈, 웹 크롤러와 Open API를 통해 데이터를 수집하는 증거 수집기, 수집된 디지털 증거 데이터를 증거화하는 증거팩 생성기로 구성된다. 시스템의 전체 구성도는 Fig.2.와 같다.

4.3.1 조사자 (investigator)

조사자는 사이버 공간 내 특정 대상에 대한 데이터를 수집하고자 하는 시스템 사용자이다. 조사자는 수집하고자 하는 서비스를 목록에서 선택하거나 url 입력을 통해 시스템이 해당 서비스에서 증거 수집 과정을 진행하도록 요청한다.

4.3.2 Service DB

시스템은 사전에 정의된 각 서비스들에 대하여 인증 모듈과 증거 수집기를 가진다. 서비스별로 그 구조가 상이하기 때문이다. 조사자가 선택한 서비스에 대해 시스템에 사전 정의되어 있는지 확인하고 해당하는 모듈을 불러오기 위하여 시스템 내 Service DB에 접근한다. Service DB는 Fig.3.과 같이 서비스 유형별로 나눠진 테이블과 서비스 별 모듈을 로드하기 위한 모듈명을 저장하는 테이블이 존재한다. 조사자의 선택에 따라 시스템은 Service DB에 대한 쿼리를 수행하며 서비스별로 해당하는 인증 모듈

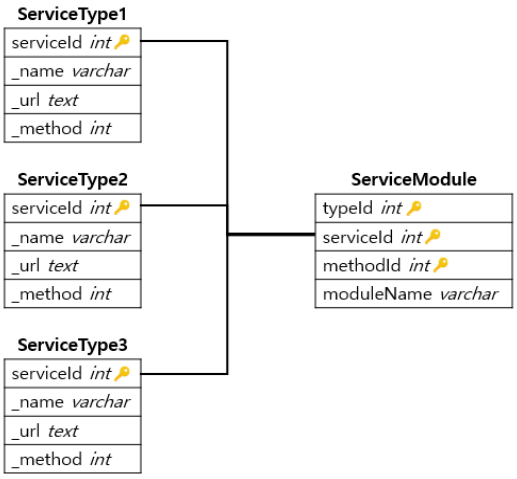

Fig. 3. Service DB - Schema Structure

과 증거 수집기를 로드한 뒤 다음 단계를 진행한다.

사전에 정의되어 있지 않은 서비스는 기존 방식과 동일하게 조사자가 직접 웹 브라우저를 통해 수집을 진행하되 안전한 환경이 보장되어야하며 본 논문에서 해당 범위는 논외로 한다.

4.3.3 인증 모듈 (authentication module)

인증 모듈은 개인 데이터 수집을 위한 접근 권한 획득을 위해 존재한다. 수사 과정에서 적법하게 획득한 계정 정보를 직접 입력하여 로그인 과정을 수행한 뒤 사용자가 서비스를 이용하는 것과 같은 환경을 얻거나, API를 사용하기 위해 OAuth 등의 서비스별로 제공하는 인증 방법을 통한 API 사용 권한을 획득한다. 계정 정보의 입력과 권한 획득은 서버와의 안전한 통신을 통해 이루어져야 한다.

4.3.4 증거 수집기 (evidence collector)

인증 모듈을 통해 접근 권한 획득 후 사용자가 선택한 수집 방법에 따라 증거를 수집한다. 시스템에서는 서비스 별로 가능한 수집 방식을 구분하여 제공한다. 수집 방식은 다음 두 가지로 나누어진다.

첫째는 웹 페이지 구조 분석을 통한 웹 크롤링 증거 수집 방식이다. 서비스 별로 수집할 수 있는 정보에 대한 웹 페이지 별 구조와 위치는 사전에 분석되어 서비스 모듈 내에 존재한다. 수집이 필요한 데이터 유형이 Table 1.에서 구분한 'text' 또는 'image'와 같이 실제 서비스 이용자가 웹 브라우저를 통해 확인하는 화면 속 내용이 중요한 데이터의 경우 보여지는 것과 동일한 형태의 pdf 파일을 생성하여 수집 당시의 웹 페이지를 수집한다. 크롤링 과정에서 웹 브라우저의 도움이 필요한 경우 웹 애플리케이션 UI 자동화 테스팅 도구로 주로 사용되는 'WebDriver'[9]를 활용한다. WebDriver의 Headless 옵션을 사용하면 GUI가 없는 형태로 웹 브라우저에서 원하는 동작을 수행할 수 있어 외부로부터의 데이터 위변조를 방지할 수 있다.

둘째는 서비스가 제공하는 Open API를 이용한 증거 수집 방식이다. 해당하는 서비스가 데이터 획득과 관련한 API를 제공할 경우에만 사용할 수 있다. 수집 데이터 역시 서비스 제공자가 제공하는 범위에 한정되어 있지만 서비스 이용자가 웹 브라우저를 통해 확인 가능한 내용 외의 추가적인 정보를 포함하기도 한다. 서버에 보낸 API 요청에 대한 응답의 형태는 주로 JSON 형태이므로 서비스 모듈에 이를 후처리하는 단계가 포함될 수 있다.

각 수집 방식에 대한 증거 수집기 구조는 Fig.4.와 같다.

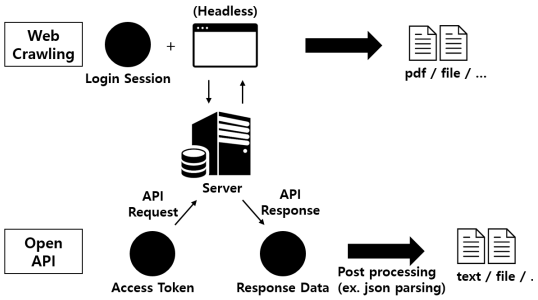

Fig. 4. Evidence Collector Structure

4.3.5 증거팩 생성기 (evidence package generator)

수집된 디지털 증거 데이터는 증거의 진정성을 보장하기 위해 수집 당시의 상태 및 환경을 기록할 수 있는 메타 데이터 정보와 함께 묶어 패키지의 형태로 보관한다. 기존 디지털 증거 확보체계에 대한 연구[10]에 따르면 사법기관에 제출할 수 있는 디지털증거데이터팩의 형태가 이미 표준으로 고시되어 있다[11]. 디지털증거데이터팩을 통해서 디지털 증거의 원본동일성, 무결성, 신뢰성 그리고 보관연속성을 확인하여 증거능력을 판단할 수 있다. 제안 시스템에서는 추후 증거팩을 표준화 된 디지털증거데이터팩의

형태로 발전시켜 나가기 위해 채증 정보로서 증거 수집 당시의 PC 정보, 타임스탬프 정보와, 압수물 목록, 해시 목록을 함께 저장한다.

또한 생성된 증거팩에 대한 해시는 시스템 내 데이터베이스에 저장되어 증거 수집 시점과 증거 제출 시점 사이에 변조되지 않았음을 보일 수 있다. 시스템 내 해시 정보는 디지털 증거 파기 시점에 함께 삭제한다. 해시 정보의 저장은 추후 더 나아가 공격자와 관리자에 의한 위변조 위험성을 낮추고, 시스템 해킹의 위험성을 줄이기 위해 블록체인을 활용한 방법[12]을 활용할 수 있을 것이다.

V. 사이버 공간 내 증거 수집 시스템의 적용

5.1 데이터 공개 범위에 따른 조사 대상 데이터 구분

사이버 공간 내에서 사건의 조사 대상 A가 가진 데이터에 대하여, 3.1절에서 언급한 것과 같이 대상이 가진 공개 데이터와 개인 데이터를 구분할 수 있다. 디지털 포렌식 조사를 통해 획득한 계정 정보는 조사 대상 A의 계정일수도 있지만, A가 아닌 A와 관련되어 있는 B의 계정일 수도 있다. 또한 획득한 계정 정보가 A와 관련 없는 C의 계정이거나 서비스에 접근할 수 있는 계정정보를 얻지 못할 수도 있다.

본 논문에서 제안한 시스템에서는 조사 과정 중 획득한 계정 정보를 입력하여 서비스에 접근 권한을 획득하고 접근 권한에 맞는 데이터들을 수집하는 방식을 취하므로 같은 서비스에 대한 데이터 수집을 진행하더라도 각 경우에 따라 획득할 수 있는 데이터의 범위가 상이하다. 따라서 데이터의 공개 범위를 중심으로 서비스 유형별 데이터의 수집 범위를 정리하였다. 이때 계정 정보를 얻지 못했거나 획득한 계정 정보가 C인 경우 공개 데이터만 수집이 가능하고, 획득한 계정 정보가 B의 계정일 경우 일부 개인 데이터 범주의 데이터 수집이 가능하다. 마지막으로 획득한 계정 정보가 A의 계정일 경우 앞선 공개 데이터, 일부 개인 데이터를 포함한 개인 데이터까지 수집이 가능하다.

5.2 서비스 유형별 데이터의 수집

5.2.1 이메일 (email)

디지털 포렌식 관점에서 이메일 서비스의 조사는 이메일의 출처와 내용을 증거로 보고, 메시지의 실제 보낸 사람과 받는 사람, 보낸 날짜와 시간 등을 확인하는 것을 말한다[13]. 따라서 이메일 서비스에서의 증거 데이터의 획득은 이메일 송수신 내역 및 이메일의 내용 확인이 중요시 된다.

Table 2.는 gmail 서비스에서 획득할 수 있는 데이터를 정리한 것이다. 이메일 서비스의 경우 개인과 개인 또는 개인과 그룹 사이에 정보를 전달할 뿐 공개 데이터는 존재하지 않는다.

Table 2. Gmail Data - Public scope based classification

Category	Data Description
public data	- not exist
semi-private data	- email list - email (eml/content) - email attachment *received from A
private data	- email list - email (eml/content) - email attachment

5.2.2 SNS (Social Network Service)

한국인터넷진흥원[14]에서는 2014년 SNS를 인터넷을 매개로 하며, 특정 목적을 위해 타인과 정보를 공유하거나 사회적 관계형성을 돕는, 쌍방향 소통 서비스라고 정의하였다. 이러한 SNS의 특성상 디지털 포렌식 관점에서 특별히 조사 대상의 개인 프로필, 게시물뿐 아니라 SNS 내에서 형성된 타 대상과의 관계 및 활동 내용들을 추가로 확인할 필요가 있다. 그러나 각 SNS 마다 주요 기능 및 목적이 다르

Table 3. Twitter Data - Public scope based classification

Category	Data Description
public data	- open user profile - open user Tweet - open user interaction with open user
semi-private data	- following user tweet - following user interaction with following user *include protected user
private data	- list (user created)
*interaction : 'retweet' and 'favorite'	

Table 4. Facebook Data - Public scope based classification

Category	Data Description
public data (no user auth)	- user profile - published page post* - open group post
public data (user auth)	- friends list - published page post - open group post - user public post
semi-private data (friends)	- user friends post - closed group post
private data (me)	- user only me post - closed group post
*post include 'post', 'comment', 'likes'	

기 때문에 서비스별 파악이 필요하다. 본 논문에서는 크게 사용자 프로필, 친구 목록, 사용자 게시물, 게시물에 대한 댓글, 다른 대상과의 상호 작용과 같이 범주를 구분하여 Twitter와 Facebook에서 획득 가능한 데이터를 각각 Table 3.과 Table 4.와 같이 나타내었다.

5.2.3 클라우드 스토리지 (cloud storage)

클라우드 스토리지 서비스는 사용자에게 가상 공간을 제공하여 저장 공간 뿐 아니라 문서 및 이미지 편집, 음악 및 비디오 플레이어, 이메일 전송 용량과 같은 다양한 추가 서비스를 제공한다[15]. 따라서 클라우드 스토리지 서비스에서의 증거 데이터의 획득은 저장 공간 내 파일에 대한 접근을 기본으로 각 클라우드 서비스가 제공하는 기능에 의한 추가적인 정보 유형을 파악하고 획득할 필요가 있다.

예를 들어 Google에서 제공하는 'Google Drive'는 파일 저장 공간과 더불어 'Google Docs' 형태의 문서 파일 편집 및 저장 기능을 가지고 있으므로, 저장 공간 내 업로드 또는 공유된 파일과 더불어 각 Google Docs 문서들에 대해 파일과 같이 취급하여 데이터를 획득한다. Google Drive에서 획득 가능한 데이터를 Table 5.와 같이 나타내었다.

Table 5. Google Drive Data - Public scope based classification

Category	Data Description
public data	- public shared file (if have a download links)
semi-private data	- shared folder/file list - shared folder's file (doc)
private data	- all owning folder/file list - all owning file (doc)

5.3 웹 크롤링과 Open API의 비교

본 장에서는 모두 Open API를 함께 제공하는 서비스들을 우선 살펴보았다. 따라서 기본적으로 수집할 수 있는 데이터의 종류는 유사하지만, 웹 크롤링을 통한 수집과 Open API를 통한 수집 결과를 서로 비교 가능하다.

이메일이나 클라우드 스토리지와는 달리 SNS는 보여지는 게시물의 순서가 명확하지 않은 경우가 많다. 5.2.2에서 살펴본 Facebook과 Twitter 모두 처음부터 전체 데이터를 표시하지 않고 스크롤을 내리는 등의 추가적인 요청에 따라 동적으로 데이터를 로드하는 형식을 사용하므로 보여지는 데이터를 수집하는 웹 크롤링 방식은 적절하지 않을 수 있다.

gmail API는 이메일 데이터의 기본이 되는 'message' 객체에 'historyId' 필드를 포함하고, history list를 출력하는 메소드를 제공하여 'messageAdded', 'messageDeleted', 'labelAdded', 'labelRemoved'와 관련한 메일함 내 변경 내역을 확인할 수 있다. 즉 API를 사용하여 웹 크롤링으로 획득할 수 없는 추가적인 정보를 확인할 수도 있다.

그러나 API를 사용할 경우 데이터 획득을 위한 요청 횟수가 API의 무료/유료 버전에 따라 다르게 제한되어 있는 Twitter나, 앱 검수 과정을 거친 앱을 사용할 경우에만 데이터 접근에 필요한 Permission을 포함하는 Access Token의 획득이 가능한 Facebook과 같이 각 서비스별 API 정책에 따라 제한 사항에 대한 대응이 필요할 수 있다.

따라서 시스템에서는 서비스 별 데이터의 특성과 API 정책을 사전에 확인하고 조사자의 시스템 사용 시 이에 적절한 방법을 권장하는 것이 필요하다.

VI. 결 론

사이버 공간 내 디지털 증거는 수정 및 삭제되기 쉬우며 실시간으로 변경사항이 반영되므로 사건 발생 시점 이후 증거 데이터의 빠른 획득이 필요하다. 클라이언트 측에서의 증거 수집은 별도의 행정 절차를 필요로 하지 않으므로 빠른 절차 수행이 가능하지만,

대용량의 증거 데이터를 임의로 수집하는 것은 비효율적이며 조사자에 의한 데이터 훼손 가능성을 해결할 수 없다.

따라서 본 논문에서는 알려진 데이터 수집 기술인 웹 크롤링과 Open API를 활용한 자동화된 증거 수집 방식을 제안하고 이를 포함하여 조사자의 증거 수집 요청부터 증거팩 생성까지의 일련의 단계를 수행하는 디지털 증거 수집 시스템을 제안하였다. 본 시스템을 통해 조사자는 알려진 서비스에 대한 디지털 증거의 효율적인 수집이 가능하고, 수집한 데이터에 대한 증거화를 통해 증거능력을 만족하는 증거팩을 획득할 수 있다. 또한 추후 법정 제출 시점까지 증거가 위변조되지 않았음을 증명할 수 있다.

향후에는 제안 시스템의 전체 구현을 진행하고, 더불어 신뢰할 수 있는 제 3의 기관과의 협력 및 제도적 관점에서의 시스템 도입을 위한 추가적인 연구가 필요하다. 또한 2단계 인증(Two-way Authentication) 등 서비스 자체 내 제한적 요소 해결 방안에 대한 연구가 진행되어야 할 것이다.

References

[1] Supreme Court Decision 2017Do9747 delivered on November 29, 2017.

[2] JoongAng Ilbo, "App to prevent forgery of evidence... NFS Korea launches Digital Evidence Certification Service", http://news.joins.com/article/19159954, Nov. 2015

[3] Ju Young Kang and Sang Jin Lee, "Screen capture authentication system for web postings to used as digital evidence," KIPS Transactions on Computer and Communication Systems, 6(1), pp. 9-16, Jan. 2017

[4] Ahreum Kim, Yeog Kim, and Sangjin Lee, "A study on notary system for web postings digital evidences," Journal of the Korea Institute of Information Security & Cryptology, 21(3), pp. 155-163, Jun. 2011

[5] NIA, "Big Data Use step by step procedure and technology manual (Version 1.0)," pp. 14, 2014.

[6] Yi Cheol Shin, The Criminal evidence law, Justinian, 2011.

[7] Sang Jin Lee, Information to Digital Forensics, eroon, 2015.

[8] The National Archives, "Web Archiving Guidance", UK, pp. 5, 2011.

[9] W3C Recommendation, "WebDriver", https://www.w3.org/TR/webdriver/, Jun. 2018

[10] Lee Insoo, "Digital Evidence Secure System," Journal of the Korea Institute of Information Security & Cryptology, 26(5), pp. 37-43, Oct. 2016

[11] "Digital Evidence Data Package," KS X 1220, Nov. 2014.

[12] Choi Bokyong and Hahm Youngwook, "A Study on Ways to Reinforce Integrity of Digital Evidence Using Blockchain," Police Science Institute, 30(3), pp. 295-318, Dec. 2016

[13] N. Meghanathan, S.R. Allam and L.A. Moore, "Tools and techniques for network forensics," International Journal of Network Security & Its Applications (IJNSA), vol. 1, no. 1, pp. 14-25, Apr. 2009.

[14] KISA, "Current status of use and Key issue analysis to domestic SNS," INTERNET& SECURITY FOCUS, pp. 56, 2014.

[15] Hyunji Chung, Jungheum Park, Sangjin Lee and Cheulhoon Kang, "Digital forensic investigation of cloud storage services," Digital investigation, vol. 9, no. 2, pp. 81-95, Nov. 2012.

〈저자소개〉

정 효 정 (Hyojeong Jeong) 학생회원
2016년 2월: 충남대학교 컴퓨터공학과 공학사
2016년 9월~현재: 고려대학교 정보보호대학원 정보보호학과 석사과정
〈관심분야〉 디지털 포렌식

최 종 현 (Jong-Hyun Choi) 학생회원
2012년 2월: 경희대학교 전자정보대학 컴퓨터공학과 공학사
2014년 8월: 고려대학교 정보보호대학원 정보보호학과 석사
2014년 9월~현재: 고려대학교 정보보호대학원 정보보호학과 박사과정
〈관심분야〉 디지털 포렌식

이 상 진 (Sangjin Lee) 종신회원
1989년 10월~1999년 2월: ETRI 선임 연구원
1999년 3월~2001년 8월: 고려대학교 자연과학대학 조교수
2001년 9월~현재: 고려대학교 정보보호대학원 교수
2008년 3월~현재: 고려대학교 디지털포렌식연구센터 센터장
2017년 3월~현재: 고려대학교 정보보호대학원 원장
〈관심분야〉 디지털 포렌식, 심층암호, 해쉬 함수