JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 24, No. 12: 1662~1669, Dec. 2020

디지털 포렌식 관점에서의 구글 클라우드 데이터 분석 연구

김도현[1*] · 김준기[2] · 이상진[3]

An Analysis of Google Cloud Data from a Digital Forensic Perspective

Dohyun Kim[1*] · Junki Kim[2] · Sangjin Lee[3]

[1*]Assistant Professor, Department of Computer Science, Catholic University of Pusan, Busan, 46252 Korea
[2]Ph.D Student, School of Cybersecurity, Korea University, Seoul, 02841 Korea
[3]Professor, School of Cybersecurity, Korea University, Seoul, 02841 Korea

요 약

구글 클라우드는 사용자가 업로드 및 동기화한 파일과 데이터뿐만 아니라 모든 클라우드 서비스들의 동기화 내역과 사용자의 스마트폰 사용 내역, 위치 정보 등도 포함하기 때문에 사용자 행위 분석 관점에서 디지털 포렌식 조사에 유용하게 사용할 수 있다. 우리는 본 논문을 통해 구글의 Takeout 서비스를 사용하여 수집 가능한 클라우드 데이터의 종류를 확인했고, 사용자 행위 분석에 필요한 데이터를 선별 및 분석하여 디지털 포렌식 연구와 조사에서 유용하게 활용할 수 있는 도구를 개발했다. 구글 클라우드 데이터는 컴퓨팅 기기의 종류와 상관없이 구글 계정을 통해 동기화되기 때문에 PC, 스마트폰, 태블릿 PC 등 다양한 기기에서 사용한 구글 서비스 데이터를 해당 기기가 없어도 구글 계정을 통해 수집할 수 있다. 따라서 본 논문의 연구 결과는 모바일 기기의 정보보호 기술의 발전으로 인해 데이터 수집이 어려워지고 있는 상황에서 디지털 포렌식 연구 및 조사에 매우 유용하게 활용할 수 있을 것으로 기대된다.

ABSTRACT

Google cloud includes data uploaded and synchronized by users, as well as synchronization history of all cloud services, users' smartphone usage, and location information. Therefore, Google cloud data can be useful for digital forensics from a user behavior analysis perspective. Through this paper, we have identified the types of cloud data that can be acquired using Google's Takeout service and developed a tool that can be usefully utilized in digital forensics research and investigation by screening and analyzing the data required for analyzing user behavior. Because Google cloud data is synchronized through Google accounts regardless of the type of computing device, Google service data used on various devices such as PCs, smartphones, and tablet PCs can be acquired through Google accounts without the device. Therefore, the results of this paper's research are expected to be very useful for digital forensics research and investigation in the current situation.

키워드 : 디지털 포렌식, 클라우드 포렌식, 구글 클라우드 데이터 분석, 구글 테이크아웃

Keywords : Digital forensics, Cloud forensics, Google cloud data analysis, Google takeout

Received 21 June 2020, **Revised** 23 June 2020, **Accepted** 29 June 2020

*** Corresponding Author** Dohyun Kim(E-mail:dohyun.daniel.kim@gmail.com, Tel:+82-51-510-0670)
Assistant Professor, Department of Computer Science, Catholic University of Pusan, Busan, 46252 Korea

Open Access http://doi.org/10.6109/jkiice.2020.24.12.1662 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

정보통신기술이 크게 발전하면서 PC, 노트북, 스마트폰, 태블릿 PC, 스마트 워치와 같은 웨어러블 기기 등 다양한 컴퓨팅 기기들이 출현함에 따라 이러한 다양한 하드웨어, 플랫폼의 다 기종 기기들을 많이 사용하는 사람들이 증가하고 있다. 최근 Cisco의 조사 결과에 따르면 2020년에는 전 세계적으로 1인당 평균적으로 사용하는 스마트 기기의 개수는 평균 3.6개로 예상되고, 그 중이 상위 3개 국가인 미국, 한국, 캐나다는 각 13.6개, 11.8개, 11개에 달할 것으로 예상된다[1].

다 기종의 컴퓨팅 기기들은 다양한 하드웨어와 플랫폼을 사용하는데, 사람들은 자신이 사용하는 여러 기기에서 업무 관련 데이터와 개인 정보의 연속적인 사용을 위해 이러한 다 기종의 컴퓨팅 기기의 데이터들을 동기화해주는 서비스를 사용한다. 이런 서비스 중 전 세계적으로 많이 사용되는 대표적인 제품들은 Amazon의 AWS, Microsoft의 AZURE, Google의 Google Cloud 등이 있다. 이 서비스 중에서 Google은 Windows, Android, iOS 등의 다양한 플랫폼에서 사용할 수 있는 클라우드 서비스를 제공하고 있다. 또한, 향후 공개용 개인 클라우드 기반의 서비스는 Google Cloud가 가장 많이 사용될 것으로 예상된다[2].

사람들은 일반적으로 클라우드 서비스를 통해 자신이 데이터를 선별적으로 클라우드 저장소인 서버로 업로드하기 때문에, 클라우드 서버에는 자신이 선택한 데이터와 자동으로 동기화를 허용한 데이터만 존재할 것으로 생각한다. 하지만 클라우드 서비스는 사용자의 서비스 사용 패턴, 성향 등을 분석하여 더욱 다양하고 편리한 서비스, 광고 등을 제공하기 위해 사용자에게 서비스 사용 내역, 위치 정보 등의 실시간 정보 동의를 통해 수집하고 있다.

따라서 클라우드 서버에는 사용자가 생성한 데이터와 클라우드 서비스가 자동으로 생성한 사용자 행위 데이터가 혼재하기 때문에, 이것은 디지털 포렌식 조사 관점에서 사용자 행위 분석에 유용하게 사용할 수 있다.

또한, 최근 업무의 효율성을 증대시키기 위한 BYOD (Bring Your Own Device)가 일상화되면서 많은 사람이 회사에서 자신의 스마트 기기를 업무에 활용하고 있고 [3, 4], COVID-19의 영향으로 전 세계적으로 재택근무가 증가함에 따라 개인의 컴퓨팅 기기 내부에 업무와 프라이버시에 관련된 정보가 혼재되고 있다. 즉 이런 데이터들은 클라우드 서버에도 같이 업로드되어 동기화될 가능성이 크다. 따라서 클라우드 데이터는 디지털 포렌식 조사 관점에서 매우 중요한 조사 대상이 되고 있다.

우리는 디지털 포렌식 관점에서 클라우드 데이터를 통해 사용자 행위를 분석하는 연구를 했다. 구글의 모든 클라우드 데이터는 계정을 통해 Takeout이라는 서비스를 사용하여 다운로드 받을 수 있다[5]. 우리는 Takeout으로 수집한 구글 클라우드 데이터의 종류와 내부 구조를 분석하여 디지털 포렌식 관점에서 사용자 정보와 행위 분석을 위한 연구와 포렌식 조사를 위한 도구를 개발하여 오픈소스로 공개했다. 우리의 연구 결과는 클라우드 서비스에 대한 디지털 포렌식 연구, 조사에 유용하게 활용될 수 있을 것이다.

Ⅱ. 관련 연구

2.1. 구글 클라우드

구글은 검색, 지도, 유튜브, 구글 드라이브 등의 다양한 서비스를 사용자에게 제공한다. 구글은 PC, 모바일, IoT 기기 등의 멀티 플랫폼에서 서비스를 제공하며, 멀티 플랫폼 간 데이터 동기화를 위해 사용자 계정 기반으로 서비스 사용 기록을 구글 클라우드에 저장한다.

즉, 구글은 일정, 주소록, 드라이브, 지메일, 행아웃, 지도, 유튜브 등의 다양한 서비스들을 계정을 통해 무료로 제공하고 이런 서비스들의 데이터와 사용자의 사용 내역 등을 서버로 동기화한다. 따라서 사용자가 어떤 하드웨어, 플랫폼이 탑재된 컴퓨팅 기기를 사용하더라도 구글 계정만 있다면 이것을 통해 자신의 사용 환경, 설정 환경에 맞는 구글의 서비스들을 연속적으로 사용할 수 있다. 사용자의 구글 계정을 기반으로 다양한 플랫폼에서 서비스를 사용한 기록이 동기화되므로 구글 클라우드에 저장된 데이터는 사용자 행위를 파악하는 데 유용하게 사용될 수 있다.

구글은 2020년 8월 기준 50개의 서비스를 제공하며, 사용자가 서비스를 사용하면 구글 클라우드에는 사용한 서비스의 종류, 사용 시간, 기기 정보가 저장된다. 구글 서비스 중 검색과 관련된 서비스는 검색 기록이 함께 저장되며, 위치와 관련된 서비스를 사용하면 위치 정보가 함께 저장된다. 또한, 사진 및 동영상, 도서 등 미디어

콘텐츠와 관련된 서비스는 콘텐츠에 접근한 시청 기록이 저장된다. 그리고 안드로이드 기반의 스마트폰을 사용하면 스마트폰의 기본 설정 정보가 저장되며, 스마트폰에서 애플리케이션의 실행 기록이 저장된다. 이렇듯 구글 클라우드는 사용자에게 편의를 제공하는 많은 서비스를 제공하는 동시에 사용자의 사용 내역들을 저장하기 때문에 디지털 포렌식 관점에서 매우 중요한 분석 대상이다.

구글 클라우드에 저장된 사용자 정보는 일반적인 연구에서도 활용되었는데, Xiaonan는 공기 오염과 사람들의 이동에 대한 연관성을 분석하기 위해 GMLH (Google Map Local History)라는 구글 지도와 관련된 사용자의 위치 정보 데이터를 사용했다[6]. 이 외에도 Markus는 유비쿼터스와 모바일 기기를 활용한 사람들의 이동 경로를 분석하기 위한 연구에서 GMLH의 데이터를 사용하기도 했다[7].

2.2. 클라우드 포렌식 연구

디지털 포렌식 분야에서 클라우드 포렌식은 지난 10여 년 동안 모바일 포렌식과 더불어 가장 활발히 연구되고 있는 분야다. 클라우드 포렌식과 관련하여 기존 연구들은 클라우드 서비스의 사용 흔적을 분석할 수 있는 아티팩트를 분석하거나 클라우드 포렌식 수행 절차를 제안했다.

Chung은 클라우드 저장소 서비스에 대한 전체적인 디지털 포렌식 조사 연구를 했고 그중에서 구글 문서(Google Docs)에 대한 데이터를 자세히 연구했다[8]. 또한, Darren Quick과 Corrado Federici 등은 클라우드 포렌식 연구를 위해 구글 드라이브 (Google Drive)에 대한 연구를 수행했다[9][10]. Williams 은 구글과 Facebook의 데이터 수집 범위에 대해 분석하였고, 그 결과 구글은 위치 서비스를 비활성화하여도 위치 정보가 클라우드에 저장되는 등 사용자의 설정보다 더 많은 데이터가 자동으로 동기화된다는 것을 확인했다[11].

Alex는 클라우드 서비스 제공자(Cloud Service Provider, CSP)에 대한 의존성을 완화한 클라우드 포렌식 절차(FMP)를 제시했다[12]. Brik은 클라우드 포렌식 과정에서 직면하는 기술적인 문제들과 그에 대한 솔루션을 제시하고, 근본적으로 클라우드 포렌식을 위해 클라우드 환경에서 저장되는 데이터의 국제적인 표준의 필요성을 주장하였다[13]. Manral는 기존 클라우드 포렌식 절차와 관련된 연구를 조사하여 클라우드 포렌식 솔루션을 네 가지(사고, 리소스, 공급자, 소비자)로 분류하였으며, 클라우드 포렌식과 관련된 과제를 제시하고 논의하여 클라우드 포렌식 조사 지침을 제안했다[14].

디지털 포렌식 조사를 위해 Takeout을 통한 구글 클라우드 데이터에 관한 연구도 있었다. Dongho Kim은 위치 정보 분석, 이메일 분석, 인터넷 사용 이력 분석을 위해 구글 클라우드의 다양한 데이터를 분석했다[15]. 이것은 Takeout에 대해 디지털 포렌식 분석을 한 유일한 연구로, 우리는 Takeout의 전체적인 구조 분석과 이 연구에서 다루지 않았던 사용자 데이터를 통해 전체적인 사용자 행위 분석을 위한 연구를 했고, 그 결과를 도구로 개발하여 구글의 Takeout 데이터를 디지털 포렌식 연구 및 조사에 유용하게 활용할 수 있도록 했다.

Ⅲ. 구글 클라우드 데이터 분석

구글 클라우드 데이터는 Takeout 서비스를 통해 피압수자와 같은 조사 대상 계정으로 로그인한 뒤 다운로드 받을 수 있다. 디지털 증거는 적법한 절차를 통해 수집되어야 법정에서 디지털 증거의 증거 능력을 인정받을 수 있다. 따라서 실제 디지털 포렌식 조사에서는 필수적으로 피압수자의 동의하에 클라우드 데이터를 수집해야 하며 본 연구는 이러한 방법을 통해 데이터를 수집한 경우를 가정하여 진행했다.

구글 Takeout에는 약 28여개의 구글 서비스의 데이터를 포함한다. 여기에는 구글 계정을 사용한 모든 기기에서 사용자가 사용한 구글 서비스들의 사용 내역, 동기화된 파일들뿐만 아니라 시간 정보를 포함하는 안드로이드에서 사용한 모든 앱의 구동 내역, 사용자의 위치 정보(위도, 경도 포함) 등도 포함된다. 우리는 이 중 디지털 포렌식 관점에서 사용자 행위 분석에 유용하게 사용할만한 정보를 포함하고 있는 표 1의 11개의 서비스를 선별하여 추출 및 분석했다.

이 서비스들은 사용자 사용 내역을 대부분 HTML과 JSON 파일 포맷으로 저장한다. HTML 포맷은 MDL (Material Design Lite)의 구조로 사용 내역과 관련된 로그의 종류, 제목, 내용 등을 저장하고, JSON 포맷은 일반적인 Key-Value 구조로 데이터를 저장한다[16].

Table. 1 Log information containing the usage history of Google services inside Takeout

Service	Parent Path	File Path
Android	My Activity\Android\	MyActivity.html
Android Device Configuration Service	\Android Device Configuration Service\	Device-[number].html
Assistant	\My Activity\Assistant\	MyActivity.html
Chrome	\My Activity\Chrome\	MyActivity.html
Contacts	\Contacts\All Contacts\	All Contacts.vcf
		[name].jpg
Drive	\Drive\	*.*
Gmail	\My Activity\Gmail\	MyActivity.html
Google Photos	\Google Photos\[YYYY-MM-DD]\	metadata.json
		*.jpg
Location History	\Location History\	Location History.json
	\Location History\Semantic Location History\YYYY\	YYYY_Month.json
Video Search	\My Activity\Video Search\	MyActivity.html
YouTube	\My Activity\YouTube\	MyActivity.html

3.1. 구글 서비스의 사용 내역 분석

사용자의 구글 서비스 사용 내역이 저장된 표 1의 11개 서비스들의 로그들을 분석했다. 우리는 각 단편적인 서비스들의 로그들을 디지털 포렌식 조사 관점에서 파싱하여 각 서비스들의 사용 내역을 쉽게 분석할 수 있도록 했다.

- **Android**: 이 서비스는 구글 계정으로 로그인한 안드로이드 기기에서 구동된 모든 앱들의 구동 내역을 저장한다. 이 서비스의 로그를 분석하면 사용자가 언제, 어느 기기에서, 어떤 앱을 구동했는지 알 수 있다. 앱에 대한 정보는 앱의 이름, 패키지 명 등을 포함한다. 다만, 그 앱을 통해 어떤 행위를 했는지, 얼마나 그 앱을 구동하고 있었는지에 대한 정보는 알 수 없다. 하지만, 시간 정보를 기준으로 구동된 앱들의 내역을 분석하면 사용자가 특정 시점에 어떤 목적을 위해 그와 관련된 앱들을 구동했으리라는 것을 유추할 수 있다. 예를 들어, 그림 1은 "\My Activity\Android\My Activity.html"의 내용 중 일부를 캡쳐한 것이다. 이것은 2020년 8월 1일 12:15:08 PM ~ 01:06:03 PM 까지 실행된 앱들의 내역으로 이것을 통해 구동한 시각, 앱 이름 등을 알 수 있다. 이러한 일련의 앱들의 구동 목적은 표 2와 같다.

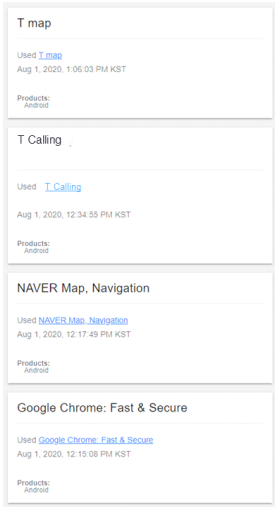

Fig. 1 Log of "Android" service (HTML format)

Table. 2 User's actual behavior history of the Fig. 1

Timestamp	App	Description
2020. 08. 01 12:15:08 PM	Google Chrome	Searching for restaurants using the Chrome app
2020. 08. 01 12:17:49 PM	Naver Map	Searching for the location of restaurants using the Naver Map
2020. 08. 01 12:34:55 PM	T Calling	Calling the restaurant using the phone app
2020. 08. 01 01:06:03 PM	T map	Using the navigation app to get to the restaurant

사용자(저자)는 점심 식사를 위한 음식점을 웹브라우저 앱을 통해 검색하고, 그 음식점의 위치를 지도 앱을 통해 확인했다. 그 후 그 음식점의 영업 여부를 확인하기 위해 전화 앱을 통해 전화 통화를 했고, 그 음식점으로 운전하여 가기 위해 내비게이션 앱을 구동했다.

- **Android Device Configuration Service**: 이 서비스는 구글 계정으로 로그인한 안드로이드 기기들의 정보를 저장한다. 이것의 로그를 분석하면 Serial Number, IMEI, Model, Brand, Manufacturer 등의 하드웨어 정보와 Firmware version, OS Fingerprint, Android SDK version과 같은 소프트웨어 정보, Timezone 및 네트워크 정보도 알 수 있다.
- **Assistant**: 이 서비스는 구글 어시스턴트를 사용한 내역을 저장한다. 이것의 로그를 분석함으로써 사용자의 질문 내용과 그에 대한 어시스턴트의 답변, 질문을 한 시간 정보, 이 서비스를 사용한 시점의 사용자 위치 정보를 분석할 수 있다. 또한, 만약 사용자가 음성 정보를 구글에 전송하도록 동의했다면 사용자가 질문하는 목소리가 MP3 파일 포맷으로 저장된다.
- **Chrome**: 이 서비스는 사용자가 구글 크롬 앱을 사용한 내역을 저장한다. 이 로그를 분석하면 사용자가 언제 어떤 사이트(웹 페이지 이름, URL)에 접속했는지 알 수 있다.
- **Contacts**: 이 서비스의 로그는 VCF 파일 포맷이며 이것을 분석하면 사용자가 구글로 동기화한 주소록 내역을 분석할 수 있다.
- **Drive**: 이 서비스는 사용자가 구글 드라이브에 업로드한 모든 파일을 포함한다. 따라서 파일들의 파일시스템 메타데이터를 분석하여 구글 드라이브를 통해 업로드된 파일들의 이름, 확장자, 수정 시각, 크기와 같은 정보를 알아낼 수 있다.
- **Gmail**: 이것의 로그는 사용자가 자신의 지메일 메일함에서 검색한 내역을 저장한다. 따라서 이것을 분석하면 검색을 수행한 시각, 키워드를 알아낼 수 있다.
- **Google Photos**: 이것은 사용자가 구글 포토로 업로드하거나 자동 동기화된 멀티미디어 파일에 대한 메타데이터와 실제 파일들을 저장한다. 따라서 이 로그를 분석하면 구글 포토에 업로드된 사진 또는 동영상 파일에 대한 앨범 명, 파일 명, 확장자, 파일 크기 정보와 다양한 시간 정보(앨범 생성 시각, 사진/동영상 파일의 촬영, 생성, 수정 시각)를 알 수 있다. 이뿐만 아니라 사진 파일에 EXIF 정보가 있는 경우 사진이 촬영된 위치 정보도 분석할 수 있다.
- **Location History**: 이 서비스는 사용자가 구글에 자신의 위치 정보 전송을 동의하면 구글 계정으로 로그인된 스마트 기기의 위치 정보를 20~30초 간격으로 수집한다. 이 로그는 위치 정보를 이동 경로와 방문 장소로 분류하며 이것을 분석하면 사용자가 이동한 경로의 출발/도착 위치에 대한 시간 정보와 위도, 경도 정보를 알 수 있고, 방문한 위치에 대한 시간, 위치 정보도 알 수 있다.
- **Video Search**: 이 서비스는 유튜브나 구글 크롬 앱 등을 통해 비디오 콘텐츠를 시청 또는 검색한 내역을 저장한다. 이 로그를 분석하면 사용자가 시청 및 검색에 사용한 서비스 종류, 콘텐츠 , URL, 검색 키워드 등을 알아낼 수 있다.
- **YouTube**: 이 서비스는 유튜브를 통해 사용자가 시청 또는 검색한 콘텐츠에 대한 내역을 저장한다. 이 로그를 분석하면 사용자가 언제, 어떤 채널의, 어떤 제목의 콘텐츠를 시청 또는 검색했는지 알 수 있다.

3.2. 사용자 행위 분석

각 서비스의 로그들을 분석한 결과를 활용하여 다양한 관점에서 디지털 포렌식 분석을 위한 다음과 같은 총 4가지 관점의 분석 방법을 연구했다.

3.2.1. 타임 라인 분석

사용자의 모든 행위를 시간 정보를 기준으로 분석하기 위해 시간 정보를 포함하는 로그의 데이터를 모두 통합, 정규화하여 분석한다. 이를 위해 Android, Assistant, Chrome, Drive(파일 수정 시각), Gmail, Google Photos (사진/동영상 촬영 시각), Video Search, YouTube 서비스들의 로그를 활용한다.

모든 이벤트를 일자, 시간대로 분석하면 사용자의 기상 시간과 취침 시간 등 생활 방식을 알 수 있다. 또한, 이벤트의 종류에 따라 어떤 앱을 많이 사용하는지, 어떤 시간대에 어떤 행위를 많이 하는지도 알아낼 수 있다. 이처럼 타임 라인 분석은 사용자에 대한 전체적인 정보를 빠르게 분석하는 데 유용하게 활용할 수 있다.

3.2.2. 검색 및 웹브라우징 내역 분석

사용자가 특정 키워드에 대해 검색을 하거나 그것과 관련된 콘텐츠를 시청한 내역을 분석하면, 사용자가 특

정 시점에 관심이 있었던 키워드를 알아낼 수 있다.

이를 위해 우리는 Assistant, Chrome, Gmail, Video Search, YouTube 서비스들의 로그를 분석했다. 이 로그들은 모두 시간 정보와 검색, 방문 또는 시청한 콘텐츠의 정보가 포함되기 때문에 기존의 웹브라우저에 대한 포렌식과 유사하고 이를 통해 분석한 내용은 정황 증거로 유용하게 사용할 수 있다.

3.2.3. 파일 정보 분석

사용자가 생성, 수정, 공유한 파일들의 정보를 분석하기 위해 Drive, Google Photos 서비스의 실제 파일들과 로그를 분석했다. 표 1.의 Takeout의 archive내부 경로에는 Drive와 Google Photos에 존재하는 사용자가 생성, 업로드, 동기화, 공유한 모든 실제 파일들이 존재한다. 따라서 우리가 "3.1 구글 서비스의 사용 내역 분석"에서 파싱한 각 파일의 메타데이터뿐만 아니라 실제 파일들도 확인할 수 있다.

Drive 분석 결과는 일반적인 클라우드, 공유 폴더에 대한 디지털 포렌식 조사처럼 활용할 수 있다. 즉, 파일의 존재 여부와 내부 데이터가 주요 쟁점인 기밀 유출 사고 조사에서 유용하다.

Google Photos는 자동 동기화를 설정해 놓으면 스마트폰에서 사진/동영상을 촬영하면 자동으로 서버로 업로드된다. 따라서 기기에서 촬영한 파일을 삭제하더라도 Takeout을 통해 구글 서버에 존재하는 원본과 같은 파일을 수집할 수 있다. 따라서 Google Photos 분석 결과는 사진과 동영상 파일이 쟁점이 되는 사건 조사에 유용하게 활용할 수 있다.

예를 들어, 도촬 사건 조사에서 용의자가 쟁점이 되는 파일을 삭제한 경우, Google Photos에 해당 파일이 동기화되어 있다면 쟁점이 되는 원본과 같은 파일을 확보할 수 있다. 특히 Google Photos의 로그에는 해당 파일들의 촬영 시각과 위치 정보가 포함되기 때문에 용의자가 특정 시점에 어디에서 해당 파일을 촬영하여 생성했는지 알아낼 수 있다.

3.2.4. 위치 정보 분석

구글은 사용자의 정보를 거의 실시간으로 동기화 및 수집하는 서비스가 많고 이것들 대부분을 Takeout으로 수집할 수 있다. 실시간으로 수집되는 정보 중 디지털 포렌식 관점에서 매우 유용하게 활용할 수 있는 것 중 하나가 위치 정보다. Takeout의 데이터 중에서 사용자의 위치 정보가 포함되는 서비스는 Assistant, Google Photos, Location History들이다. 이것들은 모두 위치 정보가 전송된 시점을 특정할 수 있는 시간 정보도 포함하고 있다.

Assistant는 해당 서비스를 사용한 시점의 사용자의 위치 정보가 해당 로그에 존재하고, Google Photos는 각 사진 파일이 촬영된 시점의 사용자의 위치 정보가 해당 로그와 각 파일의 EXIF 영역에 존재한다. Location History는 사용자의 위치 정보가 거의 실시간으로 저장되며 사용자가 이동하는 경우 사용자의 출발 위치와 도착 위치에 대한 위치 정보가 매우 높은 정확도로 저장된다.

따라서 우리는 이 3가지 서비스의 로그를 통합, 정규화하여 분석함으로써 타임 라인 분석과 비슷하게 사용자의 위치에 대한 정보를 빠르게 분석할 수 있다. 디지털 포렌식 관점에서 위치 정보는 매우 중요한 것으로 Takeout에서 수집되는 모든 정보 중 가장 유용하게 활용할 수 있다.

Ⅳ. 구글 클라우드 포렌식 도구

우리는 3장에서 사용자가 구글 클라우드 기반 서비스들의 사용 내역이 포함된 Takeout 데이터를 분석했다. 우리는 이것들을 디지털 포렌식 조사 관점에서 파싱 후 통합, 정규화하는 도구를 개발했고 이것들을 오픈소스로 공개했다.

이 도구는 수집된 Takeout 데이터를 입력하여 데이터를 파싱, 통합, 정규화하여 분석하는 gtForensics (https://github.com/dohyun-daniel-kim/gtForensics)와 gtForensics의 분석 결과를 시각화하는 뷰어인 gtForensics_viewer (https://github.com/JJun1207/gtForensics_viewer)로 구성된다.

이 도구를 사용하면 사용자의 모든 행위를 시간과 이벤트를 기준으로 재구성할 수 있고 생활 방식도 알아낼 수 있다. 그리고 다양한 검색 및 시청 기록과 사용한 프로그램 목록, 결제 내역 등을 통해 사용자의 관심사와 취향 등도 유추할 수 있다. 또한 다양한 데이터에 존재하는 시간 정보와 함께 저장된 위도, 경도와 같은 위치 정보와 IP 주소 등의 정보를 통해 사용자가 특정 시각에 실제 머물렀던 장소를 알아낼 수도 있다.

그림 2의 경우 사용자의 위치 정보를 분석한 결과를

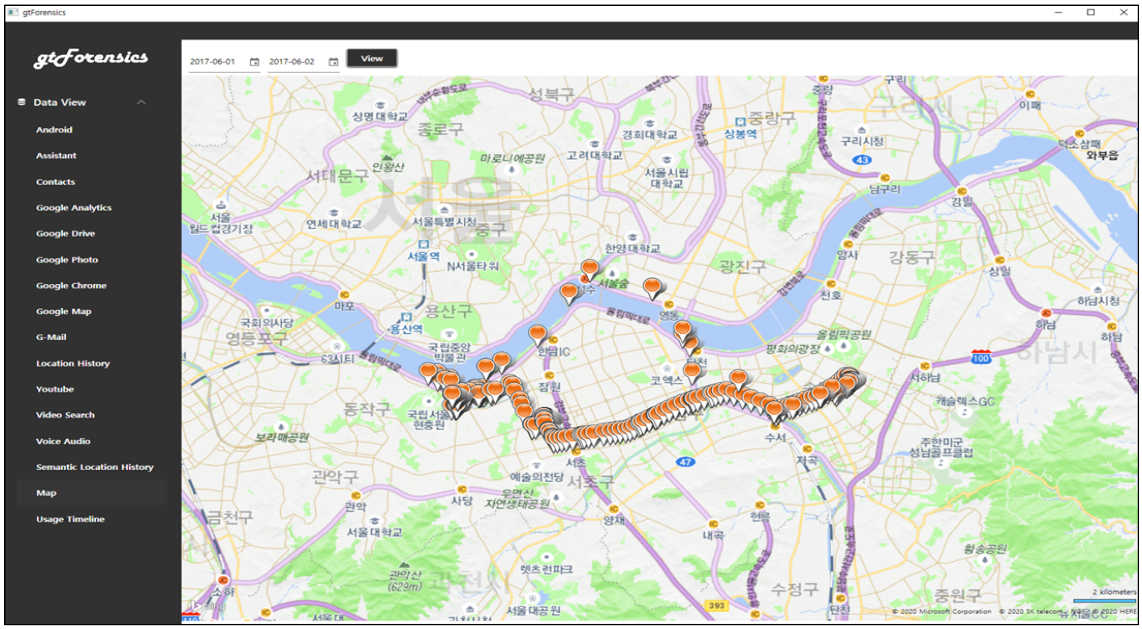

Fig. 2 The screen that visualizes the user's location in the gtForensics tool

Bing Maps API를 통해 시각화한 것이다[17]. 이처럼 사용자의 이동 경로를 시각화하면 특정 시점 또는 기간에 사용자의 위치 정보를 빠르게 분석하기 쉽다.

특히, Takeout은 사용자의 위치 정보를 거의 실시간으로 매우 상세하게 포함하고 있어서 최근 코로나19 팬데믹 상황에서 확진자의 동의하에 구글의 Takeout를 수집할 수 있다면, 우리의 위치 정보 분석을 통해 확진자의 이동 경로를 효과적으로 분석할 수 있을 것이다.

V. 결 론

본 논문은 구글의 클라우드 데이터에 대한 디지털 포렌식 분석과 관련된 것으로 이 연구를 통해 구글의 계정을 통해 동기화된 다 기종의 데이터를 효과적으로 분석할 수 있다. 우리는 약 11개의 구글 클라우드 서비스에 대한 데이터를 분석했고 이를 통합 분석하여 다양한 관점에서 복합적으로 사용자 행위를 분석했으며 이를 지원하는 도구를 개발 및 공개했다.

향후 우리는 Takeout내의 다른 데이터들도 더 분석하고 통합 분석하는 방안과 디지털 포렌식 조사에 유용하게 활용할 수 있는 시각화 방안에 대해서도 더 연구할 예정이다. 또한, 수많은 사용자 내역들에 딥러닝을 적용하여 다양한 관점에서 사용자에 대한 정보를 도출해내기 위한 연구도 진행할 계획이다.

최근 삼성 녹스와 같은 하드웨어와 결합한 보안 기능과 Bootloader, Kernel, OS 레벨에서의 보안 기능 향상으로 인해 스마트폰 데이터의 수집에 대한 어려움이 증가하고 있다. 이러한 시점에서 스마트폰 데이터를 다수 포함하고 있는 구글 클라우드 데이터에 관한 우리의 연구 결과는 디지털 포렌식 연구와 조사에 많은 도움이 될 수 있을 것으로 기대된다.

ACKNOWLEDGEMENT

This paper was supported by RESEARCH FUND offered from Catholic University of Pusan.

References

[1] Cisco Visual Networking Index: Forecast and Trends, 2017 – 2022 [Internet]. Available: https://davidellis.ca/wp-content/uploads/2019/05/cisco-vni-feb2019.pdf/.

[2] Current and planned usage of public cloud platform services running applications worldwide as of 2020 [Internet]. Available: https://www.statista.com/statistics/511467/worldwide-survey-public-coud-services-running-application/.
[3] Purposes for personal cloud service usage in South Korea in 2019 [Internet]. Available: https://www.statista.com/statistics/1013119/south-korea-cloud-service-use-personal-purpose/.
[4] Purposes for work-related cloud service usage in South Korea in 2019 [Internet]. Available: https://www.statista.com/statistics/1013121/south-korea-cloud-service-use-work-related-purpose/.
[5] Google Takeout [Internet]. Available: https://takeout.google.com/.
[6] X. Yu, A. L. Stuart, Y. Liu, C. E. Ivey, A. G. Russell, H. Kan, L. R. Henneman, S. E. Sarnat, S. Hasan, A. Sadmani, and X. Yang, "On the accuracy and potential of Google Maps location history data to characterize individual mobility for air pollution health studies," *Environmental Pollution*, vol. 252(A), pp. 924-930, 2019.
[7] M. Löchtefeld, "DetourNavigator - Using Google Location History to Generate Unfamiliar Personal Routes," *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*, Paper LBW1117, pp. 1-6, 2019.
[8] H. Chung, J. Park, S. Lee, and C. Kang, "Digital forensic investigation of cloud storage services," *Digital Investigation*, vol. 9, no. 2, pp. 81-94, 2012.
[9] D. Quick and K. K. R. Choo, "Google Drive: Forensic analysis of data remnants," *Journal of Network and Computer Applications*, vol. 40, pp. 179-193, 2014.
[10] C. Federici, "Cloud Data Imager: A unified answer to remote acquisition of cloud storage areas," *Digital Investigation*, vol. 11, no. 1, pp. 30-42, 2014.
[11] E. Williams and J. Yerby, "Google and Facebook Data Retention and Location Tracking through Forensic Cloud Analysis," *Southern Association for Information Systems (SAIS)*, 2019.
[12] M. E. Alex and R. Kishore, "Forensics framework for cloud computing," *Computers & Electrical Engineering*, vol. 60, pp. 193-205, 2017.
[13] D. Birk and C. Wegener, "Technical issues of forensic investigations in cloud computing environments," *2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*. IEEE, pp. 1-10, 2011.
[14] B. Manral, G. Somani, K. K. R. Choo, M. Conti, and M. S. Gaur, "A systematic survey on cloud forensics challenges, solutions, and future directions," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, pp. 1-38, 2019.
[15] D. Kim and S. Lee, "A Study on the Usage of Investigation of Google Cloud Data (Smartphone user-oriented)," *Journal of Digital Forensics*, vol. 12, no. 3, pp. 107-118, 2018.
[16] Material Design Lite [Internet]. Available: https://getmdl.io/.
[17] Bing maps [Internet]. Available: https://www.bing.com/api/maps/sdk/mapcontrol/isdk/loadmapasync/.

김도현(Dohyun Kim)
고려대학교 정보보호대학원 정보보호학과 공학박사
한국전자통신연구원 연구원
고려대학교 정보보호연구원 연구교수
부산가톨릭대학교 컴퓨터공학과 조교수
※관심분야 : 디지털 포렌식, 취약점 분석

김준기(Junki Kim)
경희대학교 전자정보대학 컴퓨터공학과 공학사
고려대학교 정보보호대학원 정보보호학과 공학석사
고려대학교 정보보호대학원 정보보호학과 박사과정
※관심분야 : 디지털 포렌식, 역공학, 시스템 보안

이상진(Sangjin Lee)
한국전자통신연구원 선임연구원
고려대학교 자연과학대학 조교수
고려대학교 정보보호대학원 교수
고려대학교 디지털포렌식연구센터 센터장
고려대학교 정보보호대학원 원장
※관심분야 : 디지털 포렌식, 심층암호, 해쉬함수