Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.25, NO.6, Dec. 2015 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2015.25.6.1421

윈도우 포렌식 도구의 검증용 데이터 세트의 개발*

김 민 서,† 이 상 진‡
고려대학교 정보보호대학원

Development of Windows forensic tool for verifying a set of data*

Min-Seo Kim,† Sang-jin Lee‡
Center for Information Security Technologies(CIST), Korea University

요 약

디지털 기기나 컴퓨터의 포렌식을 통한 정확한 분석을 위해 디지털 포렌식 도구의 신뢰성 검증이 매우 중요하다. 도구의 신뢰성을 검증하기 위해서는 도구에 입력할 데이터 세트에 대한 개발과 연구가 필요하다. 컴퓨터에서 많이 사용되고 있는 윈도우 운영체제에서는 시간과 관련된 데이터 기준과 사용자 행위 기준의 윈도우 포렌식 아티팩트가 있다. 본 논문에서는 이 두 개의 윈도우 아티팩트를 모두 분석할 수 있도록 윈도우 환경에서 데이터 세트를 개발한 후, 공개용 디지털 포렌식 도구들을 이용하여 테스트를 진행하였다. 따라서 개발한 데이터 세트는 행위 기준의 아티팩트를 분석할 수 있는 능력을 키우기 위한 교육용으로, 디지털 포렌식 신뢰성을 검증하기 위한 도구 테스트를 목적으로, 새로운 아티팩트 분석을 위한 재활용성의 특성을 들어 활용방안을 제시한다.

ABSTRACT

For an accurate analysis through the forensic of digital devices and computer, it is a very important validation of the reliability of digital forensic tools. To verify the reliability of the tool, it is necessary to research and development of the data set to be input to the tool. In many-used Windows operating system of the computer, there is a Window forensic artifacts associated with time and system behavior. In this paper, we developed a set of data in the Windows operating system to be able to analyze all of the two Windows artifacts and we conducted a test with published digital forensic tools. Therefore, the developed data set presents the use of the following method. First, artefacts education for growing ability can be analyzed acts standards. Secondly, the purpose of tool tests for verifying the reliability of digital forensics. Lastly, recyclability for new artifact analysis.

Keywords: Digital forensics, Data set, Corpus, Corpora, Digital forensics tool testing

I. 서 론

오늘날에는 컴퓨터와 디지털 기기가 전세계적으로 널리 사용되고 있으며, 이에 따라 컴퓨터와 디지털 기기를 이용한 범죄도 계속해서 늘어가고 있다. 이러한 범죄를 세밀하게 조사하기 위해서는 범죄에 악용된 컴퓨터와 디지털 기기를 압수 수색하여 디지털 증거의 분석이 필요하다. 그래서 디지털 기기에서 증거를 정확하게 수집하고, 분석할 수 있는 디지털 포렌식 도구의 신뢰성이 매우 중요하다. 디지털 포렌식 도구의 신뢰성을 검증하는 테스트는 두 가지가 있다.

첫 번째 방법은 이미 잘 알고 있는 샘플을 이용하는 것이다. 이 방법은 주로 도구의 업그레이드 또는 시스템의 업데이트가 이루어지는 등의 패치가 일어났

Received(08. 12. 2015), Modified(10. 27. 2015), Accepted(10. 27. 2015)
* 본 연구는 미래창조과학부 및 한국인터넷진흥원의 "2015년 고용계약형 정보보호 석사과정 지원사업"의 연구결과로 수행되었음
† 주저자, serenity@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr(Corresponding author)

을 때 이용하는 테스트이다. 알고 있는 특성을 갖는 샘플을 만들어 이 샘플을 테스트 하고자 하는 도구에 입력한 후, 나오는 결과가 '알고 있는 결과'와 일치하는지를 확인하는 것(Known test samples with known results)이다. 만약 결과가 '알고 있는 결과'와 다르다면, 도구에 문제가 있다고 판단할 수 있다[1].

두 번째 방법은 똑같은 기능을 제공하는 여러 도구들의 기능을 비교, 분석하는 방법이다. 동일한 샘플인 데이터 세트를 같은 기능을 제공하는 도구들에 입력한 후, 각각의 도구가 출력해주는 결과가 서로 같은지를 비교 분석하는 것이다. 만약 결과가 다른 도구가 있다면, 그 도구에 문제가 있다고 판단 할 수 있다.

이러한 방법들을 통해서 디지털 포렌식 도구의 신뢰성을 검증할 수 있다. 이 두 가지의 방법 모두 데이터 세트가 매우 중요한 역할을 하고 있다. 그러므로 도구의 신뢰성을 검증하기 위해서는 데이터 세트에 대한 연구와 개발이 필요하다.

한편, 데이터 세트는 corpora라고 하는데, corpora란 corpus의 복수로 사전적인 의미로는 총체, 말뭉치, 전집 등을 의미한다. 디지털 포렌식에서는 포렌식 기술과 도구에 대한 실험을 하고, 이를 검증할 수 있는 표준화된 데이터 세트를 의미한다[2]. 데이터 세트는 총 3가지의 종류로 분류 할 수 있다.

첫 번째, 특정 디지털 포렌식 이슈를 보여주기 위한 목적으로 구성된 데이터 세트가 있다. 주로 증거 분석 능력을 향상시키기 위한 교육 목적을 위해 만들어진 교육용 데이터 세트가 여기에 해당된다.

두 번째, 도구의 특정 기능을 검증하기 위해 의도적으로 생성하는 검증용 데이터 세트가 있다. 도구 테스트의 목적으로 개발된 테스트용 데이터 세트가 여기에 해당된다.

마지막으로는 인터넷과 같이 대용량의 데이터 자료에서 랜덤하게 선택하여 가져오는 샘플 데이터 세트가 있다. 예를 들어, 중고시장에서 구입한 하드웨어들에 대한 공개된 이미지 파일이 이에 해당된다.

현재 데이터 세트에 대한 연구는 국외에서 연구가 진행되고 있으나 국내에서는 디지털 포렌식 도구의 신뢰성을 검증할 수 있는 절차나 정책이 없는 상황이고, 데이터 세트에 대한 연구도 많이 미흡한 상태이다.

본 논문에서는 윈도우 운영체제 환경을 중심으로, 한국어로 된 데이터를 쌓아 대규모의 데이터 세트를 개발하였다. 개발한 데이터 세트로 공개용 디지털 포렌식 도구를 일부 선정하여 테스트를 하였다. 이렇게 개발한 데이터 세트의 개발 방법과 활용방안을 제시하고자 한다.

2장에서는 국외에서 공개된 데이터 세트를 소개하고, 3장에서는 데이터 세트 개발 방법을 제시하고, 데이터 세트 구축하는 과정을 설명한다. 4장에서는 개발한 데이터 세트에 대한 테스트 결과를 정리하고, 마지막 5장에서는 결론 및 향후 연구방향, 개발한 데이터 세트의 활용방안을 제시한다.

II. 관련 연구

미국표준기술연구소(National Institute of Standards and Technology, NIST)는 미국 상무성(Department of Commerce) 소속의 정부기관으로 디지털 포렌식 프로젝트를 국가적인 사업으로 수행하고 있다. NIST에서 수행하고 있는 CFTT(Computer Forensic Tool Testing)는 Jim Lyle이 주도하고 있는 대표적인 디지털 포레식 도구 테스트 프로젝트이다[3]. 이 프로젝트에서는 도구의 주요 기능을 검증 및 보장하기 위하여 각각의 도구들의 요구사항, 검증 방법 및 테스트 절차를 수립하는 것을 주된 목적으로 한다. CFTT 프로젝트를 소개하고 있는 홈페이지[24]에서는 디지털 증거의 수집과 관련된 기능인 디스크 이미징 기능과 증거 훼손을 방지하기 위한 쓰기 방지 기능, 증거 파일들을 복구하는 데이터 복구 기능과 증거 파일들을 검색할 수 있는 키워드 검색 기능 등이 소개되어 있고, 각각의 기능이 충족되어야 할 요구사항, 테스트 방법을 상세히 기술한 문서, 대표적인 도구들에 대한 테스트 결과 등을 공개하고 있다. 또한 CFTT 프로젝트를 통해서 개발한 데이터 세트들을 모아서 공개하고 있는데 이를 Computer Forensic Reference Data Sets(CFReDS)이라고 한다.

CFReDS 홈페이지[25]에서는 수사관의 디지털 증거 분석 능력을 키우기 위한 교육용 데이터 세트와 도구 테스트를 위한 데이터 세트가 있다. 교육용 데이터 세트는 데이터 유출 주제로 관련된 PC, USB, CD의 이미지 파일이 있다. 도구 테스트의 데이터 세트로는 검색 기능을 위한 러시아어와 영어 Unicode 문자열 이미지 파일, 메모리 분석 기능을 위한 라이브 메모리 캡처 파일, 모바일 기기 분석을 위한 모바일 디바이스 이미지 파일, 파일 복구 기능

을 위한 삭제된 파일이 있는 이미지 파일 등이 있다. 각각의 데이터 세트에는 모두 시나리오가 있고, 도구에 포함되어야 할 테스트 기능, 이 기능들을 테스트할 수 있도록 제작한 데이터 세트의 상세한 정보 등이 함께 공개되어 있다.

포렌식 도구 SleuthKit 개발자인 Brian Carrier는 디지털 포렌식 도구의 테스트를 위해 개발한 데이터 세트를 모아 공개하였는데, 이것을 Digital Forensic Tool Testing Images (DFTT)라고 한다[4]. DFTT는 CFTT 프로젝트에서 요구하는 광범위하고 복잡한 절차의 테스트의 간격을 줄이고, 신속하게 대중들에게 디지털 포렌식 도구의 신뢰성 정보를 제공하는 것을 목표로 하고 있다. DFTT는 국가기관이 아닌 민간 영역에서의 자발적으로 참여하는 구성원들이 개발한 데이터 세트 및 테스트 결과를 공유하는 오픈 프로젝트라 할 수 있다.

DFTT 홈페이지[26]에는 증거 파일의 검색 기능을 위하여 FAT, NTFS, EXT3 파일 시스템에서의 이미지 파일, FAT의 불륨 라벨 테스트를 위한 디스크 이미지 파일, NTFS 자동식별 테스트를 위한 raw 디스크 이미지 파일, 데이터 복구 기능을 위해 일부 문서 파일이 삭제된 이미지 파일, 윈도우 메모리 분석을 위한 메모리 이미지 파일 등의 데이터 세트가 공개되어 있다. 또한 각각의 이미지 파일에 테스트를 위해 인위적으로 조작한 특성들에 대한 정보도 함께 공개하고 있다.

NPS Corpus는 컴퓨터 포렌식 교육 연구의 목적으로 모아놓은 데이터 세트들의 집합이다[27]. 공개되어 있는 데이터 세트에는 교육용 데이터 세트, 도구 테스트를 위한 데이터 세트와 중고시장에서 구입한 샘플 데이터 세트 이 3가지 종류 모두 포함되어 있다. 교육용 데이터 세트로는 총 4개의 주제를 다루고 있고, 각각의 주제에는 시나리오가 있다. Nitroba 대학 내의 특정인을 괴롭히기 위한 목적으로 발송된 이메일에 대한 조사 주제로는 패킷을 캡처한 pcap 파일을 공개하고 있고, M57-특허 회사의 4주간의 행적 추적 주제로 디스크, USB, RAM 이미지 파일을 공개하고 있다[5]. 그리고 회사 기술 문서 유출의 주제로 관련 직원의 PC 이미지 파일, 국립 미술관 DC를 가상으로 공격하는 주제로 관련된 태블릿, 핸드폰, 이메일, 패킷 등의 이미지 파일을 공개하고 있다. 그리고 도구 테스트를 위해서 핸드폰 메모리 덤프 파일, 디스크 이미지 파일, Zip, JPEG 등의 각종 파일들, 패킷 덤프 파일을 공개하고 있다. 또한 세계 각국에서 얻은 실제 사용된 샘플 데이터 세트인 USB, 소니 메모리 스틱, SD카드 등 이미지 파일도 공개하고 있다. 이렇게 실제 사용된 샘플 이미지 파일들은 Real Data Corpus (RDC)라고도 하는데 RDC는 실제 사용된 기기들로부터 추출한 raw 데이터로 구성이 된다. NPS Corpus에서는 RDC를 2가지 종류로 분류하고 있다. 미국에서는 RDC를 미국 내에서 실제 사용된 샘플 이미지 파일들을 US Persons Disk Image Corpus로, 미국이 아닌 다른 나라에서 사용된 샘플 이미지 파일들은 Non-US Persons Disk Image Corpus로 분류하고 있다. RDC를 이용한 도구 테스트는 실제 사용되었던 데이터이기 때문에 예상하지 못한 결과가 나올 수 있다. 이를 이용하여 도구를 교정하는 등의 이점이 있다.

Lance Mueller는 디지털 증거 분석 능력을 향상시키고자 하는 교육 목적으로 개발한 데이터 세트를 ForensicKB 홈페이지[28]를 만들어 공개하였다. 교육용 데이터 세트가 총 3개가 존재하며 각각의 이미지 파일은 Encase 파일 형식이며, 시나리오가 존재한다. PC 이상 징후를 분석하는 주제로 관련 PC 이미지 파일과 패킷 덤프 파일이, 회사 기술 유출 분석 주제로 관련 PC 이미지 파일, 파일 복구 주제로 관련 USB 이미지 파일이 있다. 마지막으로 CCTV의 동영상 추출 주제의 시나리오가 있는데, 이 주제는 시나리오만 있고 관련 데이트 세트는 없다.

기존에 있는 데이터 세트들은 대부분 테스트만을 위한 목적으로 작은 규모로 구성되어 있다[6]. 또한 모두 도구 테스트 용 따로, 교육용 따로 분류되어 개발되어 있으며, 추가로 기존의 것에 덧붙여서 데이터 세트를 계속해서 재활용하여 사용할 수 없는 형태로 되어 있다. 예를 들어 윈도우의 IconCache.db와 같이 새롭게 발견된 아티팩트가 있을 경우, 이를 분석하는 도구의 테스트용 또는 이를 분석하는 능력을 키우기 위한 교육용 등의 각각의 목적에 맞는 새로운 데이터 세트를 개발해야만 하는 문제점이 있다.

본 논문에서는 이러한 문제점을 극복하기 위해 가상머신을 이용하여, 도구 테스트용과 교육용 2가지의 용도를 통합하고 계속해서 재활용하여 이용할 수 있는 데이터 세트를 개발하였다.

III. 윈도우 포렌식 도구 검증 및 교육용 데이터 세트의 개발

3.1 데이터 세트의 개발 방법

개발한 데이터 세트를 계속해서 재활용 할 수 있도록 만들기 위해 가상머신을 이용한다. 개인사용이나 연구 및 테스트용으로 사용한다면 라이센스가 무료인 가상 머신은 Oracle사의 VM VirtualBox와 Vmware사의 VMware Player가 있다[7].

가상 머신의 디스크 구성은 2가지의 방법이 있다. 첫 번째 방법은 Fig.1.과 같이 용량을 크게 설정한 1개의 디스크를 여러 개의 파티션으로 나누고 각각의 파티션에는 서로 다른 운영체제를 설치해서 이용하는 방법이다. 이 방법은 서로 다른 운영체제들 간에 쉽게 전환이 가능하다는 장점이 있지만, 처음에 설정해 놓은 파티션의 크기의 확장이 어려울 수 있다는 단점이 있다.

두 번째 방법은 일반적인 방법으로 하나의 디스크에 하나의 운영체제만을 설치하여 사용하는 방법이다. 이 방법으로는 디스크의 용량을 유동적으로 계속 확장시킬 수 있다는 장점이 있으나 가상머신의 이미지 파일의 크기가 커질 수 있다는 단점이 있다.

운영체제를 가상머신에 설치한 후, 아티팩트를 일련의 과정을 거쳐서 남겨야 한다. 아티팩트는 운영체제별로 종류가 다를 수 있으며, 본 논문에서는 윈도우 운영체제 중심으로 개발하였으므로 윈도우 운영체제 아티팩트에 초점을 맞추어 설명하겠다. 도구 테스트와 교육의 목적으로 활용할 수 있도록 데이터 세트를 개발하기 위한 윈도우 아티팩트를 총 2가지로 분류하였다.

시간과 관련된 데이터 기준의 아티팩트와 행위기준의 아티팩트로 나누었다[29]. 시간과 관련된 데이터 기준의 아티팩트는 Table 1.과 같다. 이 아티팩트는 디지털 포렌식 도구 테스트의 목적으로 활용하기 위한 아티팩트가 된다.

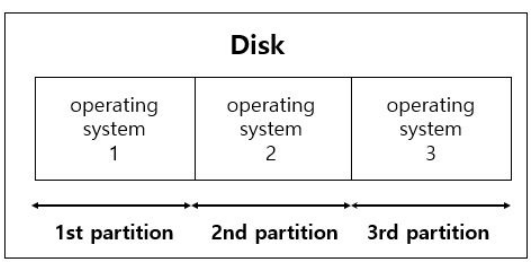

Fig. 1. Install different operating systems on multiple partitions

Table 1. Windows artifacts associated with time

No.	Windows artifacts
1	filesystem(NTFS) artifact
2	web usage
3	event log
4	superfetch
5	shortcut
6	jump list
7	restore point
8	recycle bin
9	Windows system logs
10	system temp files
11	thumbnail
12	Windows search DB

행위 기준의 아티팩트는 Table 2.와 같다. 이 아티팩트는 사용자의 행위가 그대로 반영되어 이를 분석하고자 하는 교육 목적으로 활용하기 위한 아티팩트가 된다.

데이터 기준의 아티팩트와 행위 기준의 아티팩트를 설치한 윈도우 운영체제에 일련의 과정을 거쳐서 흔적을 남겨서 데이터 세트를 개발해야 한다.

다만, 한 가지 주의해야 할 점이 있는데, 데이터 복구 도구를 위하여 이를 테스트하기 위한 별도의 데이터 세트를 개발해야 한다는 것이다. 그 이유는 운영체제가 설치된 가상머신에는 사용자 행위기준의 아티팩트가 생성되고, 이 과정에서 많은 어플리케이션과 파일들이 사용되면서 설치, 생성, 삭제 등의 과정

Table 2. Windows artifacts associated with system behavior

No.	Windows artifacts
1	reloading point
2	web usage
3	file download
4	program execution
5	network connection
6	USB or drive usage
7	suspicious file opening or creation
8	hidden file or process
9	account usage

을 거치기 때문에 이 가상머신 이미지 파일에서 데이터 복구 도구의 테스트를 한다면, 삭제된 데이터 관리가 어려워 도구의 정확한 검증이 어렵기 때문이다. 또한 파일시스템을 다르게 설정하여 데이터 복구 도구를 테스트해야 하므로 USB와 같이 저용량의 외부 저장 장치를 이용한 데이터 세트를 별도로 개발해야 한다.

3.2 데이터 세트의 개발

3.2.1 저용량의 이미지 파일 데이터 세트 개발

저용량의 이미지 파일은 총 6개의 파일로 개발하였다. 각각의 파일 시스템은 NTFS, FAT32, exFAT, EXT4, HFS+, NoFileSystem이다.

각 파일 시스템에는 Table 3.와 같은 확장자 별로 각각 정상파일 10개씩 파일을 구성하여 총 490개의 파일을 입력하였다.

입력 후 490개의 파일을 모두 삭제하였고, 이 상태로 이미지 파일을 생성하였다.

파일시스템의 존재 유무는 데이터 복구 도구가 파일시스템의 메타데이터를 활용하여 복구하는 것과 파일 시그니처와 복구하고자 하는 파일 카빙 성능을 테스트 할 수 있도록 데이터 세트를 개발한 것이다.

Table 3. The file extension for data recovery tools

Type	File extension
document	doc, docx, ppt, pptx, xls,xlsx, pdf, hwp, csv, txt, one
multimedia	jpg, gif, tif, bmp, png, psd, swf, flv, wmv, wma, mp3, mp4, mpg, wav, avi
compression	tar, alz, jar, gz, zip, 7z, rar, cab
web	htm, html, xml
executable file	exe, ico, sys, dll, lib
etc	lnk, pst, img
mac	dmg, key, pages, numbers

3.2.2 가상머신 이미지 파일 데이터 세트 개발

가상머신을 구성하기 위한 소프트웨어로는 VMware player를 사용하였으며 설치는 VMware player의 기본 설치방법으로 설치하였다.

가상머신 이미지 파일은 총 3개로, 1개의 디스크에 하나의 운영체제를 설치하였다. 설치한 운영체제는 윈도우 XP, 윈도우 7 32bit, 윈도우 7 64bit이다. 가상머신은 Fig.2.와 같은 절차로 아티팩트를 생성하여 데이터 세트를 개발하였다.

운영체제를 설치한 후, 사용자 계정과 비밀번호를 설정하였다.

그 다음으로는 문서 어플리케이션들을 설치하였다. 설치한 문서 어플리케이션은 MS Office 2007, MS Office 2013, 한글과 컴퓨터 2007, 한글과 컴퓨터 2010, Adobe Acrobat 9 Pro이다.

오피스 프로그램 설치 후, 웹 브라우저 환경을 구축하기 위해 인터넷 익스플로러, 크롬, 파이어폭스, 오페라, 사파리를 설치하였다. 각각의 웹브라우저로 국내 포털사이트의 실시간 검색어 순위를 이용하여 관련 기사와 블로그를 열람하였고, '오늘 날씨'의 키워드로 검색하여 날씨 정보와 관련 링크를 열람하였다. 그리고 '맛집', '지역명'의 키워드로 검색하여 지도 열람하였고, 자동 로그인 기능을 사용하여 메일을 전송하고 로그아웃하였다. 그 후에 pdf, doc, xls, ppt, hwp 파일을 다운로드 후 열람하였다. G마켓, 옥션에서 의류를 검색하고 로그인 하여 관심상품으로 추가하였고, Yes24, 교보문고에서 책 검색 및 열람하였다. 마지막으로 5개의 사이트를 북마크 하였다. 이 과정을 거치면서 Naver, Daum, Google을 이용하였다.

그 후, 가상머신에 4개의 외부 저장장치(USB)를 연결하여 외부 저장장치 안에 있는 문서 파일을 열람하였다. 4개의 외부저장장치에는 똑같은 문서파일이 doc, docx, xls, xlsx, ppt, pptx, hwp, pdf가 1개씩 있고, 이를 모두 한번씩 열람하였다.

가상머신에 이메일 아티팩트를 생성하기 위한 이메일 클라이언트로 Outlook Express, Windows Live Mail 2012, Microsoft Outlook, Eudora, Mozila Thunderbird, The Bat!을 이용하였다.

Windows 7 (64-bit)의 경우 디스크 용량 문제로 Windows Live Mail 2012, Eudora, Mozila Thunderbird 클라이언트를 설치하지 않았다. 설치한 이메일 클라이언트에 로그인하고, 메일 작성과 전송, 일부 메일을 삭제한 후, 로그아웃을 하였다.

그리고 각각의 가상머신에 복원지점을 생성하였다. 복원지점은 운영체제별로 5개의 지점을 1분 간격으로 연달아 생성하였다.

레지스트리에는 외부저장 장치 사용 흔적, 문서

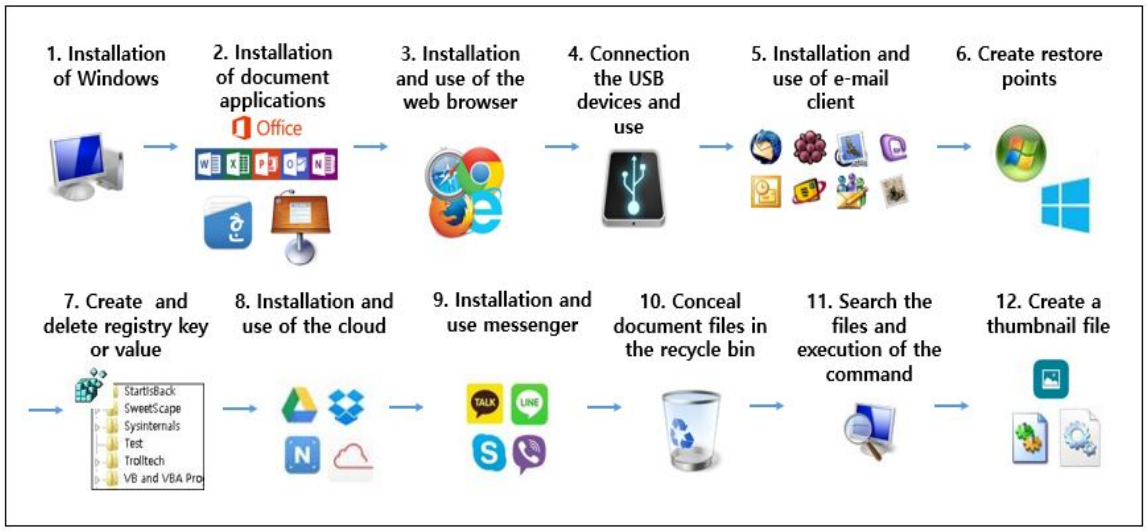

Fig. 2. Development procedure of the data set

파일 열람, 응용 프로그램 설치 등의 흔적이 남아 있고, 추가적으로 key와 value를 생성하고 삭제 하였다. 총 5개의 key를 생성하고, 그 key 안에는 value를 생성하였는데, 모두 똑같이 기본값, Dword값 0, 이진값 1, 문자열값 A,B를 생성하였다. 이 중 일부는 삭제하고 일부는 그대로 두었다.

클라우드를 사용하기 위해 PC 에이전트와 웹 접속 2가지 방법 모두 이용하였다. 클라우드 서비스는 N드라이브, 다음 클라우드, 구글 드라이브, 드롭박스를 이용하였다. 클라우드 PC 에이전트로 파일 다운로드, 파일 업로드, 파일 삭제를 하였고, 클라우드 웹에 접속하여 파일 다운로드, 파일 업로드, 파일 삭제를 하였다. 파일 업로드와 다운로드로는 모두 똑같은 11개 샘플 파일인 doc, docx, ppt, pptx, xls, xlsx, pdf, hwp, txt, jpg, rtf를 사용하였다.

매신저를 사용하기 위해 카카오톡, 네이버라인, Skype, Viber의 PC버전을 설치하였다. PC 버전의 매신저에 로그인하여 채팅을 하였고, 문서, 사진 파일을 전송하였으며, 전화통화를 하였다.

C드라이브에 있는 휴지통을 대상으로 휴지통에 문서 파일을 은닉하였다. 윈도우 7 환경에서는 $Recycle.Bin에 "Files"라는 이름으로 폴더를 생성한 후 그 안에 파일들을 은닉하였다. 은닉한 파일은 txt, hwp, docx, pptx, xlsx, jpg 파일이다.

윈도우 검색 기능을 이용하여 '그림', 'sample', 'internet', 'download', '문서', '2007', '버전', 'ABCD', '메모', 'Word', '표', '그림'의 키워드로 검색하였다. 명령어는 '서비스', 'cmd', 'msconfig'의 키워드를 입력하여 실행하였다.

그 다음으로는 'Thumbnail Sample TEST' 이름의 폴더를 생성하였고, 그 폴더 안에 10개의 그림 파일을 생성하였다. 10개의 그림 파일을 한번씩 열람한 후, 이 중 2개의 그림 파일은 삭제하였다.

이렇게 총 12단계를 거쳐 시간과 관련된 데이터 기준의 아티팩트와 행위 기준의 아티팩트를 생성하였다.

3.3 기존의 데이터 세트와의 비교

기존의 데이터 세트는 교육용과 도구 테스트용으로 성격이 명확히 구별되어 있어서 이용하고자 하는 목적에 따라 데이터 세트를 선별해야 한다. 하지만 본 논문에서 개발한 데이터 세트는 사용하고자 하는 목적의 구별 없이 교육용과 도구 테스트용 모두 사용이 가능하다.

NPS Corpus와 ForensicKB의 교육용 데이터 세트의 경우에는 각각의 데이터 세트에는 어떤 아티팩트가 있는지에 대한 자세한 정보가 공개되어 있지 않다. 이와 대조적으로 본 논문에서 새롭게 개발한 데이터 세트는 어떤 아티팩트가 있는지에 대한 정보도 공개하고 있어서 기존의 데이터 세트와는 명확한 차별성을 갖는다.

CFTT, CFReDS, DFTT의 도구 테스트용 데이터 세트는 디지털 포렌식의 도구의 기본적인 기

능[1]들에 대해서만 다루고 있으며, 이 외의 레지스트리 분석, 웹브라우저 사용 흔적 분석 등 다양한 도구들은 물론, 이 도구들이 제공하는 세부적이고 포괄적인 기능 테스트에 대해서도 다루고 있지 않다. 또한 하나의 데이터 세트에는 하나의 도구 기능 테스트만이 가능하여 테스트 하고자하는 도구와 그 기능에 맞는 데이터 세트를 선택해서 이용해야한다. 그리고 dd, img, EnCase 등의 이미지 파일로 공개되어 있어서 사용자가 해당 데이터 세트를 재활용하거나 제작자의 지속적인 업데이트가 불가능하다. 그래서 너무 오래된 데이터 세트나 환경이 바뀌어 다시 제작해야하는 데이터 세트들이 많이 있다. 또, 시스템 활성 상태에서만 동작하는 도구들에 대한 도구 테스트를 할 수가 없다. 또한 기존의 데이터 세트는 모두 영문, 러시아어 등 외국어로된 데이터만이 있고, 한국어로 된 데이터가 없어서 도구의 한국어 정상 출력 여부를 테스트 할 수 없다.

이와 반대로 본 논문에서 제작한 데이터 세트로는 레지스트리 분석, 웹브라우저 사용흔적 분석, 링크 파일 분석, 프리패치 분석 도구 등 현재 존재하는 폭넓은 도구들을 테스트 할 수 있다. 게다가 하나의 데이터 세트로 다양한 기능들을 한꺼번에 테스트를 할 수 있다. 그리고 가상머신 vmdk파일로 공개되어 있어서 사용자는 언제든지 추가적으로 데이터를 쌓아 재사용이 가능하며, 제작자의 지속적인 업데이트도 가능하다. 또한 시스템 활성 상태에만 테스트 할 수 있는 도구들을 모두 테스트 할 수 있다. 그리고 데이터 세트에는 한국어로 된 데이터가 있으므로, 도구가 한국어를 제대로 출력하는지 여부도 검증할 수 있다는 장점이 있다.

IV. 포렌식 도구 테스트 사례

4.1 테스트 할 디지털 포렌식 도구의 소개

개발한 데이터 세트를 가지고 디지털 포렌식 도구 중에서 레지스트리 분석 도구와 웹브라우저 분석 도구만 선정하여 테스트를 하였다. 제조사와 도구 이름은 Table 4.와 같다.

이 도구들은 모두 공개용 도구로 각각의 도구의 홈페이지에서 무료로 다운로드 및 이용이 가능하다.

1) 디지털포렌식 도구의 기본적인 기능은 디스크 이미징, 쓰기방지, 파일시스템 탐지, 문자열 검색, 데이터 복구 기능 등이 있다.

Table 4. List of published digital forensic tools

No.	Classification	Manufacturer	Tool Name
1	registry analysis	Harlan Carvey	RegRipper
		Didier Stevens	UserAssist
		woanware	RegRipperRunner
			ForensicUserInfo
			USBDeviceForensics
		MiTeC	Windows Registry Recovery
2	web browser artifacts analysis	NirSoft	BrowsingHistoryView
			ChromeCacheView
			ChromeHistoryView
		woanware	ChromeForensics
			FireFoxForensics
			Firefoxsessionstoreextractor
		NirSoft	IECacheView
			IECookiesView
			IEHistoryView
			MozillaCacheView
			MozillaCookiesView
			MozillaHistoryView
			MyLastSearch
			OperaCacheView
			SafariCacheView
			SafariHistoryView
			WebBrowserPassView

4.2 테스트 결과

4.2.1 레지스트리 분석 도구 테스트

레지스트리 분석 도구는 윈도우 운영체제의 시스템 구성 정보를 저장한 데이터베이스인 레지스트리 정보를 추출 및 해석해주는 도구이다. 레지스트리 분석 도구의 테스트 결과는 Table 5.와 같다.

RegRipperRunner는 윈도우 XP에서 동작하지 않는다. RegRipper와 RegRipperRunner의 경우 레지스트리를 기반으로 많은 정보들이 출력해주고 있지만, 기능별로 따로 정리되어 출력되는 것이 아니라 Sam, Software, Security, System 에서 출력될 수 있는 정보가 모두 한꺼번에 텍스트 파일로 출력되어 보기가 불편하다는 문제점이 있었다.

Table 5. The test results of the registry analysis tools

	RegRipper	UserAssist	RegRipperRunner	ForensicUserInfo	USBDeviceForensics	Windows Registry Recovery
key, value information	O	O	O	O	O	O
deleted value, key information	X	X	X	X	X	X
search by keyword	X	O	X	X	X	X
time search	X	X	X	X	X	X
installed Windows information	O	X	O	X	X	O
user activity information[2)]	O	X	O	△	X	O
system configuration information[3)]	O	X	O	X	X	O
application Information[4)]	O	O	O	X	X	O
network information	O	X	O	X	X	O
hardware information[5)]	O	X	O	X	△	O
comparison of the hive file	X	X	X	X	X	X
timeline	X	X	X	X	X	X

UserAssist는 lnk파일과 exe파일에 대한 정보만을 출력해 준다. ForensicUserInfo는 사용자 계정정보만을, USBDeviceForensics는 외부저장 장치에 대한 정보만을 출력해준다. 또한 자동 파싱 기능이 있어서 분석할 레지스트리의 하이브 파일을 따로 도구에 입력할 필요가 없다.

Windows Registry Recovery는 입력한 하이브 파일별로 출력되는 정보만 화면에 보여준다.

UserAssist를 제외한 나머지 도구들은 자체적으로 Local PC의 레지스트리 하이브 파일을 수집하는 기능은 제공하고 있지 않아서 사용자가 따로 레지스트리 하이브 파일을 수집하여 도구에 입력해야 하는 불편함이 있었다.

4.2.2 웹브라우저 사용 흔적 분석 도구 테스트

웹브라우저 사용흔적 분석 도구란 인터넷 익스플로러, 크롬, 파이어폭스, 오페라, 사파리 등의 웹브라우저를 사용하여 접속한 웹 사이트 기록, 검색기록, 다운로드한 쿠키, 임시 파일과 같은 사용흔적을 분석하는 도구를 말한다. 선정한 도구마다 지원하는 웹브라우저와 출력해주는 아티팩트가 다르다. 이를 정리하면 Table 6.과 같다.

테스트 결과는 Table 7.과 같다.

Table 6. Information of supported web browsers and artifacts

Tool name	Targeted Web Browser	Artifacts
BrowsingHistoryView	Internet Explorer, Chrome, Firefox, Opera Safari	history
ChromeCacheView	Chrome	cache
ChromeHistoryView	Chrome	history
ChromeForensics	Chrome	history keywordsearch autofill cookies thumnails
FireFoxForensics	X	X
Firefoxsessionstoreextractor	X	X
IECacheView	Internet Explorer	cache
IECookiesView	Internet Explorer	cookies

2) user activity information : 사용자 계정 정보, protected storage, 실행명령, IE-열어본페이지, 원격데스크톱 연결 등

3) system configuration information : 서비스 정보, 설치된 드라이버, 자동실행 등

4) application Information : 설치된 응용프로그램, 응용프로그램 사용 로그, 한글 및 MS Office 문서 최근 실행 파일 등

5) hardware Information : 장치 관리자, 저장 장치 등

IEHistoryView	Internet Explorer	cache
MozillaCacheView	Firefox	cache
MozillaCookiesView	Firefox	cookies
MozillaHistoryView	Firefox	history
MyLastSearch	Internet Explorer, Chrome, Firefox	keyword search
OperaCacheView	X	X
SafariCacheView	Safari	cache
SafariHistoryView	Safari	history
WebBrowserPassView	Internet Explorer, Chrome, Firefox, Opera	logins

BrowsingHistoryView는 지원하는 웹브라우저가 5가지 모두 지원을 하기 때문에 한꺼번에 여러종류의 웹브라우저 아티팩트를 분석해야 할 때 이용하기에 적합하다.

FirefoxForensics,Firefoxsessionstoreextractor,OperaCacheView는 정상적으로 설치 및 실행이 가능하지만 최신 버전의 Firefox와 Opera 웹브라우저의 아티팩트는 분석하지 못한다.

ChromeForensics은 다운로드 목록이 제대로 출력되지 않는 문제점이 있었지만, 그 외에 5가지의 아티팩트들에 대해 정보가 정상적으로 출력되므로 크롬 웹브라우저를 분석할 때에 이용하면 한꺼번에 많은 아티팩트의 정보를 분석 할 수 있다.

WebBrowserPassView는 일부 로그인시 입력한 비밀번호만 출력되고 사용자 계정 이름이 출력되지 않는 문제점이 있었다.

이 도구들 중에서 삭제된 아티팩트를 출력해주는 도구는 없었으며, 타임라인을 출력해주는 도구도 없었다.

V. 결론 및 향후 연구

디지털 포렌식 도구의 신뢰성을 검증하기 위해서 사용되는 데이터 세트는 매우 중요한 역할을 한다. 현존하는 데이터 세트는 교육 목적으로 제작된 데이

Table 7. The test results of the Web browser analysis tools

	Portable	Normal artifacts parsing	Deleted artifacts parsing	Hangul output	Support updates	Search Filtering	Support output	Support timeline
BrowsingHistoryView	O	O	X	O	O	O	O	X
ChromeCacheView	O	O	X	O	O	O	O	X
ChromeHistoryView	O	O	X	O	O	O	O	X
ChromeForensics	X	△	X	-	X	X	O	X
FireFoxForensics	X	X	-	-	X	-	-	X
Firefoxsessionstoreextractor	O	X	-	-	X	-	-	X
IECacheView	O	O	X	O	O	O	O	X
IECookiesView	O	O	X	O	O	O	O	X
IEHistoryView	O	O	X	O	X	O	O	X
MozillaCacheView	O	O	X	O	O	O	O	X
MozillaCookiesView	O	O	X	O	O	O	O	X
MozillaHistoryView	O	O	X	O	O	O	O	X
MyLastSearch	O	O	X	O	O	O	O	X
OperaCacheView	O	X	-	-	O	-	-	X
SafariCacheView	O	O	X	O	X	O	O	X
SafariHistoryView	O	O	X	O	X	O	O	X
WebBrowserPassView	O	△	X	O	O	O	O	X

터 세트와 도구 테스트를 위해 제작된 데이터 세트로 이 두 가지의 특성이 명확하게 구별되어 있다. 또한 dd, iso, EnCase 등의 이미지 파일로 배포된 기존의 데이터 세트는 새로운 아티팩트를 분석하기 위해 추가로 데이터 세트를 더 쌓는 등과 같이 재활용을 할 수가 없다.

본 논문에서 개발한 데이터 세트는 다음과 같이 활용할 수 있다. 첫째, 교육 목적으로 사용할 수 있다. 개발한 데이터 세트에는 사용자 행위 기반 아티팩트가 저장되어 있어서, 사용자의 특정 행위를 분석하기 위한 증거 분석 능력을 향상시키고자 하는 교육 목적으로 사용할 수 있다. 단순히 행위 기반 아티팩트만 있는 것이 아니라 일련의 절차에 따라 아티팩트를 생성하였기에 행위 기반 아티팩트의 타임라인을 그릴 수 있다.

둘째, 디지털 포렌식의 신뢰성을 검증하기 위한 도구 테스트 목적으로도 사용할 수 있다. 개발한 데이터 세트를 가지고 이미 알고 있는 아티팩트가 존재하기 때문에 이를 통해서 테스트 결과가 알고 있는 결과가 출력되는지, 동일한 기능을 제공하고 있는 도구의 테스트 결과가 동일한지 모두 검증이 가능하다.

마지막으로 계속해서 데이터 세트를 재활용하여 사용할 수 있다. 새로운 아티팩트가 발견된다면, 이 데이터 세트에서 데이터 세트를 추가로 생성하여 본 논문에서 개발한 데이터 세트를 재활용 할 수 있다. 또한 발견된 아티팩트가 만약 기존에 있었던 것이라면 이미 개발한 데이터 세트에 포함되어 있을 해당 아티팩트를 분석할 수 있다.

이러한 활용 방안을 통해 본 논문에서 개발한 데이터 세트는 디지털 증거 분석 능력을 향상시키고자 하는 사용자에게 도움이 될 수 있으며 또한 앞으로 꾸준히 새롭게 개발될 디지털 포렌식 도구들의 신뢰성을 테스트 하는데에도 도움이 될 것이다. 계속해서 재활용을 통해 앞으로의 디지털 포렌식 도구의 개발 및 데이터 세트 연구에도 도움을 줄 것이다. 6개의 저용량 이미지 파일은 1개의 파일〔53〕로 압축하였고, 윈도우 XP 데이터 세트는 총 3개의 파일〔54〕〔55〕〔56〕로 분할 압축을 하였다. 그리고 윈도우 7 32bit 데이터 세트는 총 4개의 파일〔57〕〔58〕〔59〕〔60〕로 분할 압축하였으며, 윈도우 7 64bit 데이터 세트는 총 3개의 파일〔61〕〔62〕〔63〕로 분할 압축하였다. 이렇게 압축한 파일들은 모두 구글 드라이브에 업로드를 하여 공개하고 있다.

향후 연구로는 윈도우 8과 10, 리눅스 그리고 맥 OS 환경의 데이터 세트와 스마트폰인 안드로이드, 아이폰에 대한 데이터 세트 연구 및 개발을 진행할 예정이다.

References

〔1〕 F.Cohen, Digital Forensic Evidence Examination, 4th Ed, Fred Cohen & Associates, 2009-2012

〔2〕 C.Altheide, H.Carvey, Digital Forensics with Open Source Tools : Using Open Source Platform Tools for Performing Computer Forensics on TargetSystems: Windows, Mac, Linux, Unix, etc, 1st Edition, Syngress Media Inc, 2011

〔3〕 H.Carvey, Windows Registry forensics: advanced digital forensic analysis of the Windows Registry, Syngress Publishing, 2011

〔4〕 S.Garfinkel, P.Farrel, V.Roussev and G.Dinolt, "Bringing science to digital forensics with standardized forensic corpora," Digital Investigation 6, 2009.

〔5〕 JR Lyle, DR White and RP Ayers, "Digital forensics at the national institute of standards and technology," National Institute of Standards and Technology, 2008

〔6〕 Lei Pan, "Robust performance testing for digital forensic tools," Digital Investigation, vol. 6, pp. 71-81, Sept 2009

〔7〕 K.Woods, C.Lee, S.Garfinkel, D.Dittrich, A.Russell and K.Kearton, "Creating realistic corpora for security and forensic education," Proceedings of the ADFSL Conference on Digital Forensics, Security and Law, 2011.

〔8〕 S.Carfinkel, "Forensic corpora, a challenge for forensic research," unpublished manuscript, 2007.

〔9〕 Peng Li, "Selecting and using virtualization solutions: our experiences with VMware and VirtualBox," Journal of Computing Sciences in Colleges, Vol. 2,

pp 11-17, Jan. 2010.
[10] F.Buchholz, E.Spafford, "On the role of file system metadata in digital forensics," Digital Investigation, Vol. 1, pp 298-309, Dec. 2004.
[11] B.Carrier, EH.Spafford, "An event-based digital forensic investigation framework," Digital forensic research workshop, 2004.
[12] M.Geiger, "Evaluating commercial counter-forensic tools," 2005 Digital Forensic Workshop, 2005.
[13] H.Carvey, "The Windows Registry as a forensic resource," Digital Investigation, Vol. 2, pp. 201-205, Sept. 2005.
[14] V.Mee, T.Tryfonas and I.Sutherland, "The Windows Registry as a forensic artefact: Illustrating evidence collection for Internet usage," Digital Investigation, Vol. 3, pp. 166-173, Sept. 2006.
[15] E.Huebner, D.Bem and CK.Wee, "Data hiding in the NTFS file system," Digital Investigation, Vol. 3, pp. 211-226, Dess. 2006.
[16] A.Castiglione, A.De Santis and C.Soriente, "Taking advantages of a disadvantage: Digital forensics and steganography using document metadata," Journal of Systems and Software, Vol. 80, pp. 750-764, May. 2007.
[17] TD.Morgan, "Recovering deleted data from the Windows registry," Digital Investigation, Vol. 5, pp. S33-S41, Sept. 2008.
[18] B.Park, J.Park and S.Lee, "Data concealment and detection in Microsoft Office 2007 files," Digital Investigation, Vol. 5, pp. 104-114, Mar. 2009.
[19] H.Chung, J.Park, S.Lee and C.Kang, "Digital forensic investigation of cloud storage services," Digital Investigation, Vol. 9, pp. 81-95, Nov. 2012.
[20] J.Collie, "The windows IconCache.db: A resource for forensic artifacts from USB connectable devices," Digital Investigation, Vol. 9, pp. 200-201, Feb. 2013.
[21] MG.Meshram, D.Kapgate, "Investigating the Artifacts Using Windows Registry and Log Files," IJCSMC, Vol. 4, pp. 625-631, Jun. 2015.
[22] NK.Shashidhar, D.Novak, "Digital Forensic Analysis on Prefetch Files," International Journal of Information Security Science, Vol. 4, no. 2, 2015.
[23] SK.Khode, VN.Pahune and MR.Sayankar, "Digital Forensic Tool for Decision Making in Computer Security Domain," International Journal for Research in Emerging Science and Technology, Vol. 2, Apr. 2015.
[24] CFTT, http://www.cftt.nist.gov/
[25] Computer Forensic Reference Data Sets, http://www.cfreds.nist.gov/
[26] Digital Forensic Tool Testing Images, http://dftt.sourceforge.net/
[27] NPS Corpus, http://digitalcorpora.org/
[28] Lance Mueller's Homepage, http://www.forensickb.com/
[29] Portable Forensics, Windows Artifact Analysis,http://portable-forensics.blogspot.kr/2014/10/windows-artifact-analysis.html
[30] Harlan Carvey, RegRipper, v2.8, https://code.google.com/p/regripper/downloads/list
[31] Didier Stevens, UserAssist, v2.6.0, http://blog.didierstevens.com/programs/userassist/
[32] woanware, RegRipperRunner, v1.0.4, http://www.woanware.co.uk/forensics/regripperrunner.html
[33] woanware, ForensicUserInfo, v1.0.5, http://www.woanware.co.uk/forensics/forensicuserinfo.html
[34] woanware, USBDeviceForensics, v1.0.14, http://www.woanware.co.uk/forensics/usbdeviceforensics.html

〔35〕 MiTeC, Windows Registry Recovery, v1.5.3, http://www.mitec.cz/Data/XML/data_downloads.xml
〔36〕 NirSoft, BrowsingHistoryView, v1.69, http://www.nirsoft.net/utils/browsing_history_view.html
〔37〕 NirSoft, ChromeCacheView, v1.66, http://www.nirsoft.net/utils/chrome_cache_view.html
〔38〕 NirSoft, ChromeHistoryView, v1.22, http://www.nirsoft.net/utils/chrome_history_view.html
〔39〕 woanware, ChromeForensics, v1.0.5, http://www.woanware.co.uk/forensics/chromeforensics.html
〔40〕 woanware, FireFoxForensics, v1.0.5, http://www.woanware.co.uk/forensics/firefoxforensics.html
〔41〕 woanware, Firefoxsessionstoreextractor, v1.0.2, http://www.woanware.co.uk/forensics/firefoxsessionstoreextractor.html
〔42〕 NirSoft, IECacheView, v1.53, http://www.nirsoft.net/utils/ie_cache_viewer.html
〔43〕 NirSoft, IECookiesView, v1.77, http://www.nirsoft.net/utils/iecookies.html
〔44〕 NirSoft, IEHistoryView, v1.70, http://www.nirsoft.net/utils/iehv.html
〔45〕 NirSoft, MozillaCacheView, v1.6, http://www.nirsoft.net/utils/mozilla_cache_viewer.html
〔46〕 NirSoft, MozilaCookieView, v1.50, http://www.nirsoft.net/utils/mzcv.html
〔47〕 NirSoft, MozilaHistoryView, v1.56, http://www.nirsoft.net/utils/mozilla_history_view.html
〔48〕 "NirSoft, MyLastSearch, v1.63, http://www.nirsoft.net/utils/my_last_search.html
〔49〕 NirSoft, OperaCacheView, v1.06, http://www.nirsoft.net/utils/opera_cache_view.html
〔50〕 NirSoft, SafariCacheView, v1.11, http://www.nirsoft.net/utils/safari_cache_view.html
〔51〕 NirSoft, SafariHistoryView, v1.01, http://www.nirsoft.net/utils/safari_history_view.html
〔52〕 NirSoft, WebBrowserPassView, v1.60, http://www.nirsoft.net/utils/web_browser_password.html
〔53〕 https://drive.google.com/open?id=0ByMck91GiIuqNUVNN2pzcjkyT1E
〔54〕 https://drive.google.com/open?id=0Byhj6HV8ySUyMTV6Q2FETF9kQ2s
〔55〕 https://drive.google.com/open?id=0Byhj6HV8ySUycHRoVHI3SEJLbVk
〔56〕 https://drive.google.com/open?id=0Byhj6HV8ySUyNDdydHlERm01TE0
〔57〕 https://drive.google.com/open?id=0B9Sfk3oZxm9IUTJiMjRPWHltRjA
〔58〕 https://drive.google.com/open?id=0B9Sfk3oZxm9IdW9XZHdINTRtMUE
〔59〕 https://drive.google.com/open?id=0B9Sfk3oZxm9IQjJtaWNIVU9qOEU
〔60〕 https://drive.google.com/open?id=0B9Sfk3oZxm9IQUZNZjYycVMxZTg
〔61〕 https://drive.google.com/open?id=0B9Xaut-MwPuJTGVnSC16MzJvbGc
〔62〕 https://drive.google.com/open?id=0B9Xaut-MwPuJTVhldWhqRDVGM2s
〔63〕 https://drive.google.com/open?id=0B9Xaut-MwPuJN1gxeVMwVXNVV1E

〈저자소개〉

김 민 서 (Min-Seo Kim) 학생회원
2014년 3월: 숙명여자대학교 이과대학 컴퓨터과학과 이학사
2014년 3월~현재: 고려대학교 정보보호대학원 금융보안학과 석사과정
〈관심분야〉 디지털 포렌식, 정보보호, 역공학

이 상 진 (Sang-jin Lee) 종신회원
1987년 2월: 고려대학교 수학과 학사
1989년 2월: 고려대학교 수학과 석사
1994년 8월: 고려대학교 수학과 박사
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
〈관심분야〉 디지털 포렌식, 심층 암호, 해쉬 함수