DOI: 10.3745/KIPSTC.2010.17C.5.391

디지털 포렌식 관점의 데이터 복구 및 분석 프레임워크

김 진 국[†] · 박 정 흠[††] · 이 상 진[†††]

요 약

대부분의 디지털 포렌식 관련 도구들은 저장매체의 할당된 영역에 존재하는 파일을 분석하는데 초점을 맞추고 있기 때문에 용의자가 의도적으로 삭제하거나 이전에 사용되었던 파일들을 복구하는 기능이 부족하다. 따라서 삭제된 파일을 효과적으로 분석하기 위해서는 데이터 복구 도구로 얻은 결과물을 다시 디지털 포렌식 도구로 분석해야한다. 이러한 작업은 사건을 빠르고 효과적으로 대응하거나 데이터의 무결성을 보존하는데 적절하지 않다. 따라서 본 논문에서는 기존 디지털 포렌식 관련 도구들의 한계점을 벗어나 디지털 포렌식 관점에서 데이터 복구 및 분석 도구를 통합 구현하기 위한 프레임워크를 제시하고 구현하였다.

키워드 : 디지털 포렌식, 프레임워크, 데이터복구, 데이터카빙

A Framework for Data Recovery and Analysis from Digital Forensics Point of View

Jinkook Kim[†] · Jungheum Park[††] · Sangjin Lee[†††]

ABSTRACT

Most of digital forensics tools focus on file analysis of allocated area on storage. So, there is a lack of recovery methods for deleted files by suspects or previously used files. To efficiently analyze deleted files, digital forensic tools depend on data recovery tools. These process not appropriate for quick and efficient responses the incident or integrity preservation. This paper suggests the framework for data recovery and analysis tools from digital forensics point of view and presents implementation results.

Keywords : Digital Forensics, Framework, Data Recovery, Data Carving

1. 서 론

디지털 증거는 초기 사이버 범죄를 수사하는데 결정적인 역할을 하였다. 최근에는 강도 및 살인, 사기, 도박 등의 범죄에서도 디지털 증거는 필수 조사 대상이 되었다. 이것은 현대인의 생활방식에 디지털 기기가 깊숙이 자리 잡고 있다는 것을 나타낸다.

디지털 데이터는 디지털 기기에서 사용되는 0과 1로 표현된 정보이다. 디지털 데이터는 디지털 기기가 동작하는 중에만 사용되는 휘발성 데이터와 디지털 기기의 전원이 차단된 이후에도 유지되는 비휘발성 데이터로 나눠볼 수 있다. 전원이 차단된 이후에도 사용되기 위해서는 하드디스크, 플래시메모리, CD(Compact Disk) 등의 저장매체를 이용해 저장해야 한다. 일반적으로 저장매체를 이용해 저장되는 데이터는 파일형태로 저장된다. 이러한 이유로 디지털 기기로부터 데이터 복구는 저장매체에서 파일을 복구하는 것으로 설명될 수 있다.

파일 복구는 삭제된 데이터를 삭제되기 이전 상태로 복원시키는 것이다. 파일이 삭제되는 원인은 파일이 더 이상 필요하지 않아서 삭제되는 경우, 응용프로그램의 필요에 의해 삭제되는 경우, 그리고 용의자가 의도적으로 삭제하는 경우로 나눠볼 수 있다. 디지털 포렌식 관점에서 데이터 복구는 용의자가 의도적으로 삭제하였거나 이전에 사용되었던 데이터들을 대상으로 한다. 따라서 이를 효과적으로 분석한다면 수사에 단서로 사용될 가능성이 매우 높다. 하지만 현재 널리 사용되는 디지털 포렌식 도구들은 삭제되지 않은 파일에 중점을 두어 분석을 수행한다. 일부 복구 기능을 제공하는 도구들도 존재하지만 그 기능이 미약하다. 이러한 상황에서 삭제된 파일을 효과적으로 분석하기 위해서는 전문 데이터 복구를 통해 얻은 결과를 디지털 포렌식 분석 도구로 다시

※ 본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업의 일환으로 수행하였음[10035157, 실시간 분석을 위한 디지털 포렌식 기술 개발].
† 준 회 원 : 고려대학교 정보경영공학전문대학원 석사과정
†† 준 회 원 : 고려대학교 정보경영공학전문대학원 박사과정
††† 종신회원 : 고려대학교 정보경영공학전문대학원 교수
논문접수 : 2010년 3월 18일
수 정 일 : 1차 2010년 5월 28일, 2차 2010년 7월 13일
심사완료 : 2010년 8월 3일

분석해야 한다. 하지만 이 경우 분석에 번거로움이 생길뿐만 아니라 디지털 포렌식 분석에서 가장 중요한 데이터 무결성이 훼손될 우려가 있다. 결국 삭제된 파일을 디지털 포렌식 분야에 효과적으로 활용하기 위해서는 데이터 복구에 기반한 디지털 포렌식 분석 도구가 필요하다. 따라서 본 논문에서는 기존 디지털 포렌식 관련 도구들의 한계점을 벗어나 디지털 포렌식 관점에서 데이터 복구 및 분석을 통합적으로 할 수 있는 프레임워크를 제시하고 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존에 연구된 디지털 포렌식의 프레임워크와 도구를 살펴보고, 3장에서는 데이터 복구 기법에 대해서 살펴본다. 5장에서는 디지털 포렌식 관점에서 데이터 복구를 위한 기능 요구사항을 살펴보고, 6장에서는 디지털 포렌식 관점에서 효과적인 데이터 분석을 위한 기능 요구사항에 대해 살펴본다. 그리고 7장에서 결론을 맺는다.

2. 관련 연구

2.1 기존 디지털 포렌식 프레임워크 연구

기존에 디지털 포렌식 관련 프레임워크는 수사 절차에 관한 내용이 대부분이었다[1-3]. 그리고 네트워크 환경에서의 분석에 초점을 맞춘 프레임워크가 있다[4, 5]. Nick L. Petroni, Jr et al. 2006년에 발표한 논문은 휘발성 메모리만을 대상으로 하고 있고, Mark Reith et al. 의해 발표된 논문은 추상적인 모델만 언급하고 있다[6, 7]. Jeroen van den Bos et al.이 발표한 논문에서 전자 장치에서의 데이터 획득 및 디코딩에 대해서 다루지만 모바일 장치만을 대상으로 하고 있다[8]. 이렇듯 아직까지 데이터 복구를 기반으로 한 프레임워크는 연구된 바가 없다.

2.2 전문 디지털 포렌식 도구

2.2.1 EnCase Forensic Edition〔9〕

EnCase는 GuidanceSoftware社에서 개발한 디지털 포렌식 도구로 NIST(National Institute of Standards and Technology)의 CFTT(Computer Forensics Tool Testing)을 통과하였고, 현재 전 세계적으로 사용되고 있다. 다양한 파일시스템 볼륨을 비롯하여 디지털 포렌식 이미지 또한 분석할 수 있다. 그리고 인식된 파일시스템의 할당 영역 파일, 삭제된 파일, 파일 슬랙, 비할당 영역을 효과적으로 분석할 수 있는 기능을 제공한다. 하지만 데이터 카빙 복구 기능은 별도의 EnScript를 통해 지원하지만 헤더와 푸터만 입력받아 카빙하므로 많은 오탐이 발생한다.

2.2.2 FTK(Forensic Toolkit)〔10〕

FTK는 AccessData社에서 개발한 디지털 포렌식 도구로 EnCase와 함께 전 세계적으로 널리 사용되고 있다. 분석에는 오라클 데이터베이스가 사용되어 분산 처리나 인덱스 검색을 효과적으로 사용할 수 있도록 지원한다. 게다가 이미지 생성, 레지스트리 분석, 해쉬 분석을 비롯하여 복호화 기능, 패스워드 크랙, 심층암호 분석 기능을 지원한다. 하지만 FTK 역시 할당되어 있는 데이터 분석에 초점을 맞추고 있다. 데이터 카빙 복구 기능은 헤더와 푸터만을 기반한 카빙 기법을 사용하므로 많은 오탐이 발생한다.

2.2.3 FINALForensics〔11〕

FINALForensics는 데이터 복구 서비스로 시작하여 현재는 다양한 디지털 포렌식 관련 제품을 서비스하고 있는 FINALDATA社에서 개발한 디지털 증거 분석 솔루션이다. FINALForensics는 기존에 데이터 복구 솔루션인 FINALDATA 제품에 기반한 제품이기 때문에 데이터 복구 기능은 많이 갖추고 있다.

하지만 디지털 포렌식 이미지를 지원하지 않기 때문에 별도의 가상마운트 프로그램이 필요하다. 게다가 스캔 단위의 제약으로 인해 용의자가 저장매체 포맷 시 클러스터나 블록 크기를 변경할 경우 이전 데이터를 복구하기 어렵다. 또한 메타 정보를 이용한 파일 복구와 데이터 카빙이 분리되어 있지 않아 사건의 성격에 따라 능동적으로 데이터를 복구하기 어렵다.

2.3 전문 데이터 복구 도구

데이터 복구 도구는 TopTenREVIEWS에서 발표한 2010 Data Recovery Software Review를 대상으로 살펴보았다[12]. 데이터 복구 전문 도구들은 메타 정보를 이용한 파일 복구를 비롯하여 데이터 카빙, 파티션 복구 등의 기능을 지원한다. 하지만 단순히 데이터를 복구하기 위한 도구이기 때문에 디지털 포렌식 분석을 위해서는 새로운 저장매체에 다시 저장한 후 저장된 파일을 대상으로 분석해야 한다. 이 경우 파일의 시간 정보와 데이터가 변경되어 무결성이 훼손될 염려가 있다. 무결성이 훼손된다면 디지털 포렌식 분석에서 필수적인 파일 간의 시간 관계 분석이나 해쉬 분석을 수행하기 어렵게 되며, 증거로서 인정받을 수 없다. 결과적으로 무결성을 보존하기 위한 추가적인 연구가 필요하다.

이와 같이 기존에 널리 사용되는 디지털 포렌식 도구들은 데이터 복구 기능이 부족한 반면, 데이터 복구 도구들은 디지털 포렌식 분석을 하기에 적합하지 않다. 디지털 포렌식 도구들은 실제 수사 환경에서 사용되고 법률적인 효력이 필요한 만큼 그 기능이 명확하게 정의되어야 한다. 하지만 대부분의 디지털 포렌식 연구들이 수사 절차나 법률적인 문제에 초점을 맞추고 있어 개발된 도구들이 저마다의 서로 다른 장단점을 가지고 있는 경우가 많다. 따라서 디지털 포렌식 분석에 적합한 데이터 복구 모델이 필요하다.

본 논문은 디지털 포렌식 관점에서 데이터 복구 및 분석에 관한 도구가 갖추어야할 기능들에 관한 프레임워크를 제시하는데 초점을 맞추고자 한다. 특히 데이터분석에 관한 기능 요구사항은 복구된 데이터뿐만아니라 다양한 디지털 포렌식 데이터를 분석하는데 필수 기능으로 적용될 수 있을 것이다.

3. 데이터 복구

3.1 저장매체 물리적 손상에 의한 복구

물리적 데이터 복구는 저장매체를 구성하는 하드웨어가 손상된 경우의 복구 기법이다. 예를 들어 하드디스크의 경우 내부적으로 액추에이터, 암, 플래터, 스핀들 모터, PCB(Printed Circuit Board) 등의 요소로 구성된다. 데이터는 플래터 위에 자기장을 사용해 인코딩되어 있는데 플래터를 제외한 나머지 구성 요소들이 손상된 경우 해당 요소를 교체하여 플래터의 내용을 복구할 수 있다. 또한 플래터가 손상된 경우에도 손상 정도에 따라 복구가 가능한 경우도 있다. 한편 플래시 메모리의 경우 데이터를 저장한 칩을 제외한 나머지 부품이 손상된 경우 칩을 제거한 후 리볼링(Re-Balling) 방법을 통해 복구가 가능하다.

3.2 논리적 데이터 복구

논리적 데이터 복구는 저장매체의 물리적인 하드웨어에 기반하여 복구하는 기법이 아닌 논리적인 파일시스템이나 파일에 기반 하여 복구하는 기법이다. 논리적 데이터 복구는 크게 파일시스템 메타정보를 이용한 복구, 데이터 카빙 복구, 덮어써진 파일 복구로 나눌 수 있다.

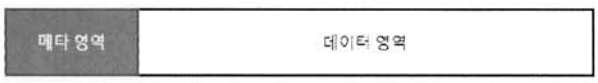
3.2.1 메타 정보를 이용한 파일 복구

메타 정보를 이용한 파일 복구는 복구를 위해 파일시스템의 메타 정보를 이용한다. (그림 1)은 파일시스템의 추상화된 구조이다. 다수의 파일을 효과적으로 관리하기 위해 파일명, 크기, 위치, 특성 등은 메타 영역에 구조적으로 저장된다. 그리고 실제 파일 데이터는 데이터 영역에 저장된다.

일반적인 파일시스템은 파일 삭제 작업이 발생하면 파일의 실제 데이터를 지우는 것이 아니라 메타 영역의 일부 정보만 변경하여 파일이 삭제되었다는 것을 표시한다. 파일시스템의 이러한 특성은 삭제된 파일의 메타 영역이 새로운 파일로 덮어써지지 않는다면 메타 영역의 파일 위치 및 크기 정보를 이용해 데이터 영역에 저장된 파일의 실제 데이터를 비교적 쉽게 획득할 수 있다. 이렇듯 파일시스템의 메타 정보를 이용하여 파일을 복구하는 기법을 메타 정보를 이요한 파일 복구라 한다.

3.2.2 데이터 카빙 복구

데이터 카빙 복구는 파일시스템의 특성을 이용하는 것이 아니라 파일의 특성을 이용하는 것이다. (그림 1)을 살펴보면 파일이 삭제된 후 메타 영역에 존재하는 파일 정보까지 삭제된 경우 파일의 위치, 크기 등을 알 수 없기 때문에 더 이상 파일시스템이 제공해 주는 정보는 사용하기 어렵다.



(그림 1) 추상화된 파일시스템 구조

이 경우에는 데이터 영역에서 현재 사용되지 않는 영역(비할당 영역)을 대상으로 복구하고자 하는 파일의 특성을 이용해 복구할 수 있는데 이러한 기법을 데이터 카빙 복구라 한다. 일반적으로 데이터 카빙 복구는 파일이 조각나지 않은 경우에 복구하는 기법과 조각난 경우에 복구하는 기법으로 나눌 수 있다.

조각나지 않은 파일 카빙 기법은 헤더/푸터 시그니처를 이용하거나 램슬랙(Ram Slack)을 이용하여 복구하는 기법이다[13]. 조각난 파일 카빙 기법은 파일 구조체, 내용 및 엔트로피에 기반하여 복구하는 방법으로 아직까지 가능한 포맷이 제한적이고 많은 시간이 요구된다[14].

3.2.3 덮어써진 파일 복구

덮어써진 파일 복구는 (그림 1)에서 메타 영역과 데이터 영역이 모두 새로운 데이터로 덮어써진 경우에 복구하는 방법이다. 현실적으로 불가능해 보이지만 1996년 Peter Gutmann에 의해 복구 가능성이 드러났다[15]. 하드디스크의 경우 데이터는 자기장을 사용해 인코딩 된다. 이러한 상황에서 데이터가 덮어써질 경우 STM(Scanning Tunneling Microscopy)을 통해 자기장의 변이를 관찰하면 덮어쓰기 이전의 데이터 흐름을 알 수 있다는 것이다. 하지만 이는 매우 제한적인 상황에서 일부 비트들의 복구가 가능한 것으로 현실적으로 원본의 데이터를 완벽하게 복구하는 것은 매우 힘들다.

본 논문에서는 소프트웨어로 복구가 가능한 논리적 복구 중 메타 정보를 이용한 파일 복구와 데이터 카빙을 대상으로 디지털 포렌식 관점에서 복구 및 분석하기 위한 방법을 살펴본다.

4. 디지털 포렌식 관점의 데이터 복구를 위한 기능 요구사항

4.1 복구 대상 지정

일반적인 데이터 복구 도구들은 운영체제에 마운트 되어 있는 저장매체를 대상으로 한다. 하지만 디지털 포렌식 관점에서 볼 때 저장매체 전체를 압수하거나 복제할 경우 외에도 저장매체의 이미지만 획득하는 경우도 있기 때문에 이를 고려해야 한다. 저장매체 이미지는 물리적인 저장매체의 전체 영역을 비트스트림 복제하거나 논리적인 파일시스템 영역만 이미징하는 경우가 있다.

저장매체의 이미지는 내부적으로 파일시스템을 가지고 있기 때문에 이미지를 가상 드라이브로 마운트 하거나 도구상에서 이미지를 직접 해석해야 한다. 게다가 파일시스템 자체가 손상된 경우에도 복구가 가능해야 하기 때문에 RAW 데이터도 추가할 수 있어야 한다.

4.2 복구 영역 지정

최근에 대용량의 저장매체가 보편화됨에 따라 데이터를

복구하는 시간이 매우 많이 소요된다. 따라서 저장매체의 일부 영역을 대상으로 복구하는 것도 의미가 있다. 앞서 메타 정보를 이용한 파일 복구와 데이터 카빙에 대해 살펴보았는데, 메타 정보를 이용한 파일 복구는 파일시스템 정보를 이용하므로 데이터 카빙에 비해 빠른 시간 내에 복구가 가능하다. 또한 이 파일은 다른 파일로 덮어써지기 이전이므로 비교적 최근 파일이라고 볼 수 있다. 반면 데이터 카빙 복구는 파일시스템 정보 없이 파일의 특성만을 이용해 복구해야 하기 때문에 많은 시간이 소요된다. 따라서 사건의 성격에 따라 메타 정보를 이용한 파일 복구와 데이터 카빙을 적절히 활용할 수 있도록 독립적으로 동작할 수 있어야 한다. 또한 저장매체 전체영역에서 메타 정보를 이용한 파일 복구와 데이터 카빙 복구가 완료된 영역을 제외한 나머지 영역을 대상으로 텍스트 추출과 같은 추가적인 분석이 가능하도록 해야 한다.

4.3 복구 단위 지정

대부분의 데이터 복구 도구들은 복구하기 위해 저장매체를 검색하는 단위로 현재 파일시스템이 사용하는 클러스터나 블록 단위를 사용한다. 하지만 고정된 단위의 검색은 용의자가 파일시스템 포맷 시 클러스터나 블록 단위를 변경할 경우 이전 파일시스템에 기록되었던 파일을 복구하기 어렵다. Windows 2000 시스템에서 파일시스템을 포맷할 경우 저장매체의 용량이 1GB를 초과하면 클러스터가 자동적으로 4K(4,096 Byte)로 설정된다[16]. 하지만 이 단위는 임의로 변경할 수 있기 때문에 클러스터 크기를 변경한 경우, 현재 파일시스템이 사용하는 클러스터 크기만 고려한다면 이전에 사용하던 데이터는 복구하기 어렵다. 따라서 이를 능동적으로 변경할 수 있도록 지원해야 한다.

4.4 복구 파일 형식 지정

앞서 살펴본 바와 같이 메타 정보를 이용한 파일 복구는 비교적 짧은 시간에 완료할 수 있지만 데이터 카빙의 경우 저장매체의 용량에 따라 많은 시간을 필요로 한다. 따라서 사건의 성격에 따라 복구 대상을 지정하여 복구 시간을 줄일 필요가 있다. 예를 들어, 포르노그래피와 관련된 사건을 조사하는 경우에는 전체 파일형식을 대상으로 데이터 카빙을 수행하는 것이 아니라 멀티미디어 파일을 우선적으로 복구하는 것이 바람직 할 것이다. 또한 악성코드에 의해 피해를 받은 시스템이라면 실행 파일을 우선적으로 복구해야 할 것이다. 이처럼 저장매체와 관련된 사건의 성격에 따라 선택적으로 복구할 수 있도록 복구 파일 형식을 지정할 수 있어야 한다.

4.5 복구 파일 검증

일반적으로 100GB의 저장매체의 경우 10만개가 넘는 파일이 저장된다. 이와 같은 저장매체에서 데이터 카빙을 수행할 경우 카빙 알고리즘에 따라 2~4배의 추가적인 파일들이 생성된다. 그래픽 파일의 경우에는 미리보기가 가능하기 때문에 잘못 복구된 파일을 쉽게 제거할 수 있으나 문서 파일의 경우에는 직접 응용프로그램을 통해 확인하는 방법 외에는 특별한 방법이 없다. 이러한 오탐은 추가적으로 많은 분석시간이 필요하므로 수사에 큰 어려움으로 작용한다. 오탐을 해결하기 위한 방안은 이미 연구가 진행되어 있다[17, 18]. 따라서 이러한 방법을 이용하여 응용프로그램 수준에서 표현이 가능한 파일만 선별할 수 있는 방안이 필요하다.

4.6 복구 파일 이름 생성

메타 정보를 이용한 파일 복구의 경우 파일시스템에 의존하기 때문에 삭제된 파일 이름을 복구할 수 있다. 하지만 파일 카빙의 경우 파일의 메타 정보가 없기 때문에 파일 이름을 복구하기 어렵다. 따라서 대부분의 데이터 복구 도구들은 중복을 방지하기 위해 저장매체에서 해당 파일의 물리적인 위치 값을 이용하여 파일 이름을 생성한다. 복구 파일의 수가 많을 경우 파일 이름을 단순히 물리적인 위치 값으로 생성하는 것은 수사관에게 아무런 도움도 주지 못한다. 따라서 파일 이름을 생성할 때 각 파일 형식에 맞는 의미있는 정보로 파일 이름을 생성한다면 특정 파일을 찾는데 매우 효과적일 것이다. 예를 들어, 일반적으로 문서 파일의 경우 파일의 처음에는 문서의 제목을 쓴다. 따라서 문서의 처음 일정 바이트로 파일이름을 생성한다면 파일을 응용프로그램을 통해 열어보기 전에 중요하다고 판단되는 파일을 선별하여 우선적으로 분석할 수 있을 것이다[19].

5. 디지털 포렌식 관점의 효과적인 데이터 분석을 위한 기능 요구사항

5.1 파일 분류 분석

최근 대용량의 저장매체로 인해 저장매체마다 작게는 수만에서 많게는 수십만 개의 파일을 사용한다. 따라서 데이터 복구를 수행할 경우에도 많은 파일이 복구 된다. 이 경우 효과적인 분석을 위해서는 복구된 파일을 적절히 분류할 수 있는 기능이 필요하다.

첫 번째 분류 방법은 파일의 시그니처별 분류이다. 윈도우의 경우7에는 어플리케이션 바인딩을 위해 확장자를 사용하지만, 리눅스, MacOS의 경우에는 확장자가 기반이 되지 않는다. 따라서 정확한 파일의 분류를 위해서는 시그니처별로 파일을 분류하는 것이 바람직하다. 시그니처별로 분류된 파일은 수사관의 용도에 따라 빠르게 분석이 가능하다. 두 번째 분류 방법은 시간별 가상 폴더와 타임라인 분류이다. 디지털 포렌식 관점에서 파일의 시간 정보는 사건이 발생한 시점과 연관될 수 있기 때문에 매우 중요한 정보이다. 메타 정보를 이용한 파일 복구의 경우 파일시스템에 기반하기 때문에 파일의 시간 정보를 함께 복구할 수 있다. 따라서 각 시간별로 가상폴더를 만들거나 타임라인을 생성하여 파일을 분류한다면 사건이 일어난 시점을 중심으로 효과적인 분석이 가능하다. 세 번째는 파일의 특성에 의한 분류이다. 파일

의 특성은 파일시스템에 의해 부여되는 파일시스템 수준의 암호화, 압축, 숨긴 특징과 어플리케이션에 의해 파일 자체에 부여되는 암호화, 압축 등의 특징이 있다. 이러한 특징은 용의자가 능동적으로 파일을 조작했다는 것을 알 수 있다. 따라서 파일의 특성에 의한 분류는 다른 파일에 비해 정밀하게 분석해야 할 파일들을 선별해 준다.

5.2 정규표현식 검색

대용량의 저장매체로부터 복구된 다수의 파일을 대상으로 분석하기 위해서는 검색과 필터 기능이 필수적이다. 복구된 파일은 다양한 포맷과 형식을 가지고 있으므로 강력한 분석을 위해서는 단순한 키워드가 아닌 정규표현식의 사용이 가능해야 한다.

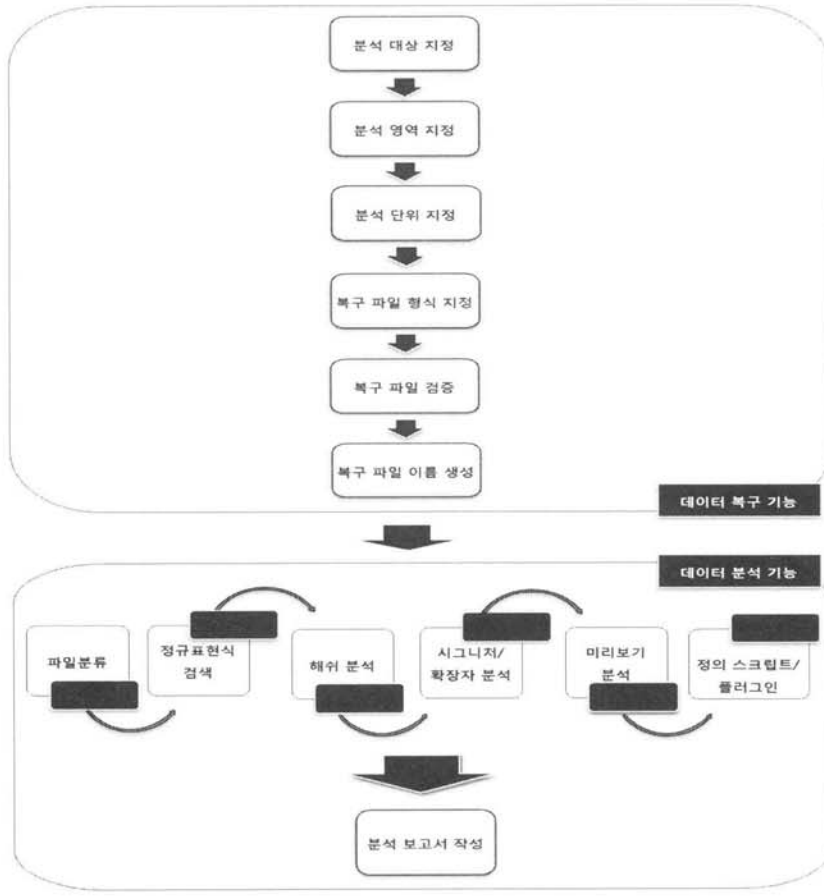
우선 파일 이름을 대상으로 정규표현식에 의한 필터 기능이 필요하다. 메타 정보를 이용한 파일 복구로 복구한 파일과 데이터 카빙으로 복구한 파일 이름에 대해 필터링을 수행하게 되면 원하는 파일은 효과적으로 분류할 수 있다. 두 번째로 논리적인 파일 데이터를 대상으로 한 검색 기능이 필요하다. 논리적인 파일 내부에는 파일에서 사용하는 의미 있는 데이터를 포함한다. 따라서 각 파일을 어플리케이션으로 직접 확인하기 전에 적절히 분류할 수 있다. 마지막으로 슬랙(Slack) 영역을 대상으로 검색이 이루어져야 한다. 저장매체 상에 존재하는 슬랙은 파일 슬랙, 파일시스템 슬랙, 볼륨 슬랙이 있다[20, 21]. 이러한 슬랙들은 일반적으로 운영체제에서 다루지 않는 영역이기 때문에 악성코드나 중요한 데이터를 은닉하는데 사용하곤 한다. 따라서 슬랙 영역을 대상으로 검색을 수행할 수 있는 기능이 필요하다.

5.3 해쉬(hash) 분석

해쉬 분석은 알려진 파일들에 대한 해쉬셋을 이용하여 분석하고자 하는 저장매체에서 알려진 파일을 제거하기 위한 목적으로 수행된다. 특히 악성코드에 피해를 입은 시스템이나 침입에 의해 손상된 시스템의 경우 운영체제의 실행파일을 변경하는 경우, 해쉬 분석을 통해 손쉽게 변경된 파일을 찾아낼 수가 있다. NIST의 NSRL(National Software Reference Library)을 이용하면 운영체제 버전별로 알려진 파일에 대한 해쉬셋을 얻을 수 있다[22]. 또한 생성된 각 파일의 해쉬값은 보고서 작성 시 증거의 무결성을 입증하기 위해 사용될 수 있다.

5.4 시그니처/확장자 분석

파일의 경우 형식에 따라 파일의 헤더 부분에 고유한 시그니처를 가지고 있다. 파일은 시그니처를 통해 구분되기도 하지만 윈도우 운영체제의 경우 확장자를 기준으로 파일을



(그림 2) 디지털 포렌식 관점의 데이터 복구 및 분석 프레임워크

구분한다. 이러한 특성을 이용하여 파일을 은닉하기 위해 시그니처 혹은 확장자를 임의로 변경시키는 경우가 많다. 따라서 시그니처와 파일 확장자간의 매칭을 수행하여 용의자의 의도적인 조작여부를 파악하는 것이 필요하다.

5.5 미리보기 분석

디지털 포렌식 도구로 데이터를 효과적으로 분석하기 위해서는 해당 파일을 새로운 저장매체로 저장하기 전에 파일 내용을 확인할 수 있어야 한다. 그래픽 파일이나 문서 파일의 경우 미리보기를 통해 해당 파일의 내용을 살펴볼 수 있다면 도구를 통한 효과적인 분석이 될 수 있다. 또한 파일 형식이나 볼륨의 특정 영역을 정밀하게 살펴볼 수 있도록 HexView를 지원하는 것도 필요하다.

5.6 사용자 정의 스크립트 및 플러그인

개발자에 의해 개발되는 도구들은 사용자의 요구를 모두 반영하기 힘들다. 대부분의 널리 사용되는 도구들은 추가적인 사용자의 요구에 적절히 대응할 수 있도록 별도의 스크립트나 플러그인을 통해 도구의 기능을 보안한다. 디지털 포렌식 도구인 EnCase Forensic, 디스어셈블러인 IDA Pro의 경우에도 사용자가 직접 작성한 스크립트를 실행할 수 있는 기능을 지원하며, 동적 디버깅 프로그램인 OllyDbg는 플러그인을 추가할 수 있다[9, 23, 24].

5.7 디지털 포렌식 분석 보고서 작성

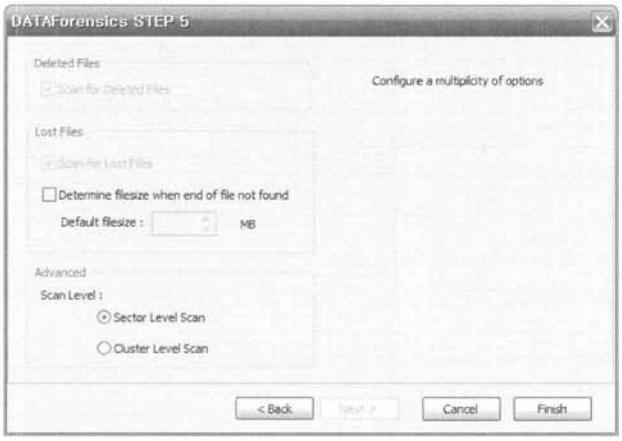
디지털 포렌식 분석 결과는 법정에서 효력을 가져야 하므로 증거에 대한 무결성과 정확성을 입증할 수 있도록 신뢰성 있게 작성되어야 한다. 대부분 디지털 포렌식 분석관을 제외하고는 디지털 포렌식 기술에 대한 이해가 낮다. 따라서 디지털 포렌식 분석 보고서는 분석 전체과정에 대해 논리적으로 기술하면서 쉽게 알 수 있도록 작성되어야 한다. 분석보고서에는 기본적으로 조사자 이름, 사건 이름, 사건 번호을 비롯하여 증거로 사용될 각 파일의 이름, 경로, 크기, 시간정보, 해쉬값 등이 포함되어야 한다. 그래픽 파일의 경우에는 그래픽이미지도 분석 보고서에 추가되어야 한다.

6. 구현 결과

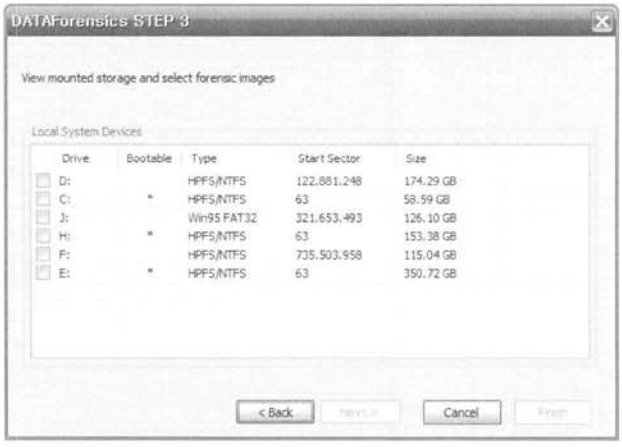
앞서 제시한 프레임워크를 기반으로 디지털 포렌식 관점의 데이터 복구 및 분석 도구인 DATAForensics 구현하였다. 현재 제시한 기능 중 복구 대상 지정, 복구 영역 지정, 복구 단위 지정, 복구 파일 형식 지정, 복구 파일 검증, 파일 분류 분석, 정규표현식 검색, 분석 보고서 작성 기능의 일부를 구현하였다. 앞으로 해쉬 분석, 시그니처/확장자 분석, 미리보기 분석, 사용자 정의 스크립트 및 플러그인을 추가해 나갈 예정이다.

삭제된 파일 복구 기능은 실제 저장매체를 통한 테스트가 중요하다. 따라서 현재 의뢰받은 복구 저장매체를 대상으로 전문 복구 도구와의 성능을 비교하고 있다. 그리고 전문 디지털 포렌식 도구와 디지털 포렌식 분석 성능에 대해 오탐이나 잘못된 결과가 없는지 비교하고 있다. 디지털 포렌식 분석 결과는 법정에서 증거로 인정받아야 하므로 결과에 대해 신뢰할 수 있어야 한다. 따라서 지속적인 전문 디지털 포렌식 도구와의 비교 테스트가 필요할 것으로 판단된다.

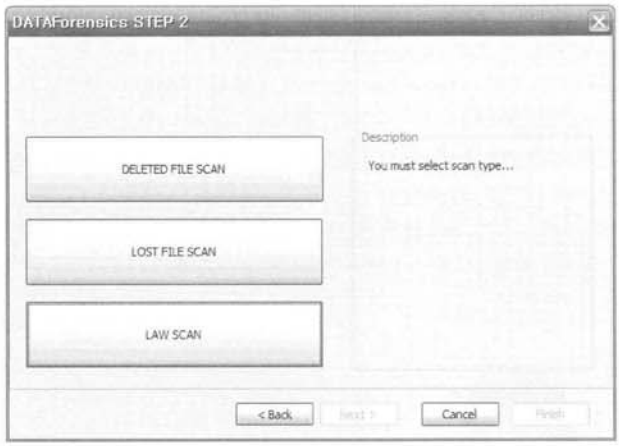
(그림 3) DATAForensics 주 화면



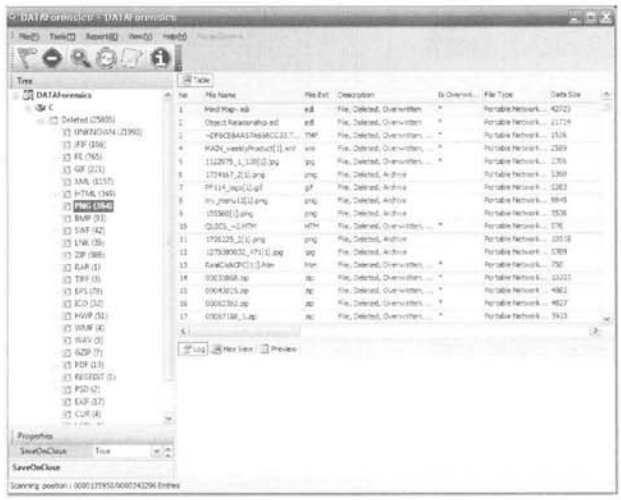
(그림 4) 복구 단위 지정



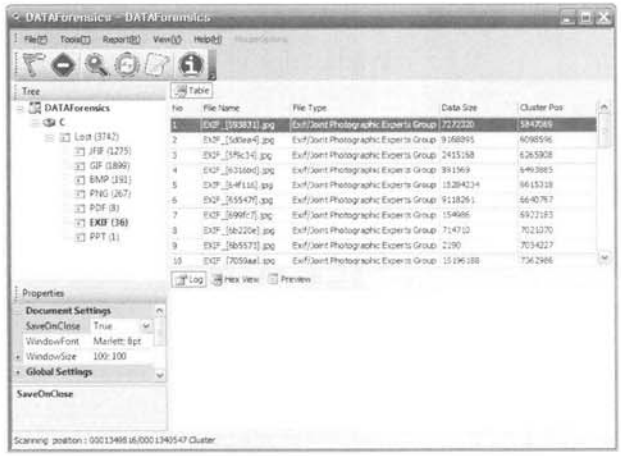
(그림 5) 복구 대상 지정



(그림 6) 복구 영역 지정



(그림 7) 메타 영역을 이용한 파일 복구



(그림 8) 파일 카빙 복구

7. 결 론

디지털 기기가 현대 생활에서 뗄 수 없을 만큼 개인의 디지털 기기 의존도가 점점 더 증가하고 있다. 따라서 범죄 수사에서 디지털증거가 활용될 가능성이 더욱 높아졌다. 현재 널리 사용되고 있는 디지털 포렌식 도구들은 저장매체의 할당 영역에 있는 파일을 대상으로 디지털 포렌식 분석을 수행한다. 일부 비할당 영역의 데이터를 복구하기 위한 방법을 제공하지만 기능이 제약적이거나 부족하다. 또한 비할당 영역의 데이터를 전문으로 복구해주는 도구들은 디지털 포렌식 분석에 활용하기에 많은 제약이 있다. 따라서 디지털 포렌식 분석에서 활용 가능한 데이터 복구 및 분석 도구가 필요하다.

디지털 포렌식 분석을 위한 데이터 복구를 수행하기 위해서는 다양한 대상을 마운트 할 수 있는 복구 대상 지정을 비롯하여 복구 영역 지정, 복구 단위 지정, 복구 파일 형식 지정, 복구 파일 검증, 복구 파일 이름 생성의 기능이 필요하다. 또한 복구된 데이터를 디지털 포렌식 관점에서 분석하기 위해 파일 분류 분석, 정규표현식 검색, 해쉬 분석, 시그니처/확장자 분석, 미리보기 분석, 사용자 정의 스크립트 및 플러그인 추가 기능이 필요하며 법정에서 증거의 효력을 가지기 위해 신뢰할 수 있는 분석 보고서가 작성되어야 한다.

용의자가 의도적으로 파일을 삭제하였거나 이전에 사용되었던 파일들이 사건에 결정적인 증거로 활용될 가능성은 항상 존재한다. 디지털 포렌식 도구들은 수사에 활용 가능해야하며 다른 도구와 다르게 분석된 결과가 법정에서 증거로 효력을 인정받을 수 있어야 한다. 따라서 본 논문에서 제시한 프레임워크는 데이터복구라는 중요한 포렌식 기법을 활용하여 제한된 시간 내에 효과적인 분석 기능들에 기반한 프레임워크를 제시하였다. 이러한 프레임워크는 데이터 복구뿐만 아니라 다른 디지털 포렌식 데이터를 분석하기 위한 필수적인 기능으로 활용될 수 있을 것이다.

현재 제시한 프레임워크를 기반으로 한 1차 버전의 데이터 복구 및 분석 도구를 개발하였다. 추후 프레임워크에서 제시한 기능을 추가하여 완성된 도구를 개발할 예정이다. 실제 수사 환경에서 데이터 복구에 초점을 맞춘 디지털 포렌식 도구가 활용되면 좀 더 정밀한 증거 분석이 가능할 수 있을 것이다.

참 고 문 헌

[1] Nicole Lang Beebe, Jan Guynes Clark, "A hierarchical, objectives-based framework for the digital investigations process," Digital Investigation, Vol.2, June, 2005.

[2] Brian D. Carrier, Eugene H. Spafford, "An Event-Based Forensic Investigation Framework," Digital Forensic Research Workshop (DFRWS), Aug., 2004.

[3] Ricci S.C. leong, "FORZA - Digital forensics investigation framework that incorporate legal issues," Digital Investigation, Vol.3, September, 2006.

[4] M.I. Cohen, "PyFlag - An advanced network forensic framework," Digital Investigation, Vol.5, September, 2008.

[5] Dr. Ren Wei, "A Framework of Distributed Agent-based Network Forensics System," Digital Forensic Research

Workshop (DFRWS), Aug., 2004.
[6] Nick L. Petroni, Jr., AAron Walters, Timothy Fraser and William A.Arbaugh, "FATKit : A framework for the extraction and analysis of digital forensic data from volatile system memory," Digital Investigation, Vol.3, December, 2006.
[7] Mark Reith, Clint Carr and Gregg Gunsch, "An Examination of Digital Forensic Models," International Jounal of Digital Evidence, Vol.1, Issue 3, Fall 2002,
[8] Jeroen van den Bos and Ronald van der Knijff, "TULP2G-An Open Source Forensic Software Framework for Acquiring and Decoding Data Stored in Electronic Devices," International Journal of Digital Evidence, Vol.4, Issue 2, Fall 2005.
[9] EnCase Forensic, http://www.guidancesoftware.com/computer-forensics-ediscovery-software-digital-evidence.htm
[10] FTK(Forensic Toolkit), http://www.accessdata.com/forensictoolkit.html
[11] FINALForensics, http://www.finaldata.co.kr/Forum/?s=FRM.FF2
[12] TopTenREVIEWS, http://data-recovery-software-review.toptenreviews.com/
[13] 김진국 외 4명, "파일카빙 기법에 관한 연구", 디지털 포렌식연구 2호, 2008년 6월.
[14] Joachim Metz, Bas Kloet and Robert-Jan Mora, "Analysis of 2007 DFRWS Forensic Carving Chllenge - The 07 smart carving approach," Digital Forensic Research Workshop (DFRWS), July, 2007.
[15] Peter Gutmann, "Secure Deletion of Data from Magnetic and Solid-State Memory," Sixth USENIX Security Symposium, July, 1996.
[16] Microsoft support site, http://support.microsoft.com/kb/140365.
[17] Simson L. Garfinkel, "Carving contiguous and fragmented files with fast object validation," Volume 4, Supplement 1, September 2007.
[18] 김진국 외 2명, "Flase-Positive를 줄이기 위한 향상된 파일 카빙 기법 연구", 안티 포렌식 대응기술 워크샵, 2008년 8월.
[19] 김진국 외 4명, "파일카빙에서의 파일이름 결정 방법", 한국정보보호학회 동계학술대회, 2008년 12월.
[20] Brian Carrier, "File System Forensic Analysis," Addison Wesley Professional, P139-140, 2005.
[21] Steve Bunting, "The Official EnCase Certified Examiner STUDY GUIDE Second Edition," Sybex, 2007.
[22] National Software Reference Library, http://www.nsrl.nist.gov/
[23] IDA Pro, http://www.hex-rays.com/idapro/
[24] OllyDbg, http://www.ollydbg.de/

김 진 국

e-mail : proneer@gmail.com
2008년 강원대학교 컴퓨터정보통신공학과 (학사)
2008년~현 재 고려대학교 정보경영공학 전문대학원 석사과정
관심분야 : 디지털 포렌식, 데이터 복구 및 분석

박 정 흠

e-mail : junghmi@korea.ac.kr
2007년 한양대학교 정보통신대학 컴퓨터전공 (학사)
2009년 고려대학교 정보경영공학전문대학원 (석사)
2009년~현 재 고려대학교 정보경영공학 전문대학원 박사과정
관심분야 : 디지털포렌식, 안티-안티 포렌식

이 상 진

e-mail : sangjin@korea.ac.kr
1987년 고려대학교 수학과(학사)
1989년 고려대학교 수학과(이학석사)
1994년 고려대학교 수학과(이학박사)
1989년~1999년 ETRI 연구원
1999년~현 재 고려대학교 정보경영공학 전문대학원 교수
관심분야 : 디지털 포렌식, 모바일 포렌식, 심층 암호, 해쉬 함수