ISSN(Print) : 1976-5304, ISSN(Online) : 2713-931X
https://kdfs.jams.or.kr, DOI : 10.22798/kdfs.2022.16.1.131

문서보안이 설정된 이메일 첨부파일의 효율적 분석을 위한 도구 설계
- 범죄수사와 e-Discovery 활용을 목적으로 -

장 동 근*, 김 기 범**
성균관대학교 과학수사학과(대학원생)*, (교수)**

Implementation of the Tool for Efficient Review of Email Attachments which Document Security is applied
- For the purpose of Criminal investigation and e-Discovery -

Donggeun Jang*, Gibum Kim**
Dept. of Forensics, Sungkyunkwan University (Graduate Student)*, (Professor)**

요 약

디지털 포렌식 분석관은 이메일을 선별·수집하기 위해 사건과 관련된 키워드나 기간을 설정한다. 이메일 첨부파일에 문서보안이 적용되어 있다면 추출한 후 복호화하여 수집한다. 분석관은 이메일과 추출된 첨부파일의 끊어진 연결성을 확인하며 분석해야 한다. e-Discovery 전문 솔루션을 통해 연결성을 확인할 수 있지만, 비용 문제로 파일명을 직접 비교하여 확인하면 착오가 발생할 수 있어 분석이 지연된다.

본 논문에서는 범죄수사와 e-Discovery에서 복호화된 첨부파일을 이메일에 다시 첨부하여 연결성을 확보할 수 있는 도구를 구현한다. EML·MSG 형태의 이메일 파일에서 문서보안이 적용된 첨부파일들을 추출하여 연결성을 기록하고, 관리자에 의해 복호화된 첨부파일을 재첨부하여 저장하도록 설계하였다. 사용자는 해시값이 변경된 증거의 동일성을 증명하기 위해, 실행 과정에서 생성되는 해시값의 변경 기록을 확인할 수 있다. 실험 결과, 해당 재첨부 도구를 통해 상용 솔루션 없이 이메일과 첨부파일의 연결성을 유지하였다. 분석에 필요한 과정을 단축할 수 있고, 선별·수집 시 누락을 방지할 수 있음을 확인하였다.

주제어 : 문서보안, DRM, 디지털 포렌식, 이메일 첨부파일, 범죄수사, e-Discovery

ABSTRACT

When collecting e-mails from custudians in digital forensics, e-mails are selected through keywords or periods related to the event. If the attached file in the e-mail is applied with document security, the attachments would be extracted. Then document security manager remove the protection of attachments, and investigator collect them separately. For this reason, investigators should check the parental relationship between the extracted attachment and the original e-mail, which has lost connectivity. To address this inconvenience, investigators can use the E-Discovery solution that can be reviewed while maintaining connectivity between the contents of the email and the attachment, but if it is difficult to use due to cost issues, it may take a long time to review and cause errors due to the same file name.

In this paper, the implemented tool presents the method to reattach the decrypted attachment to the original email in order to solve the problem of the existing method in which the connection between the contents of the email and the decrypted attachment is disconnected. The tool extracts attachments that need to be decrypted targeting EML or MSG-type email files and records connectivity with the original email. it attaches the attachments decrypted by the security manager to the email again and stores it. In order to prove the identity of the evidence in which the hash value has been changed, the user may check a record of the change of the hash value generated during the execution process. As a result of the experiment, the connection between the email and the attachment was maintained without a commercial solution through the relevant reattachment tool. It was confirmed that the process necessary for analysis can be shortened, and omission can be prevented during search and seizure.

Key Words : Document Security, DRM, Digital Forensics, E-mail Attachments, Criminal Investigation, e-Discovery

Ⅰ. 서 론

최근 기업들은 개인정보 보호법 개정과 재택근무의 활성화로 문서보안 솔루션 도입을 확대하고 있다. 암호화 방식의 솔루션은 기업 내부망에 있는 모든 문서를 암호화하고 열람 권한을 제한한다. 권한 없는 분석관은 외부망에서 문서를 열람할 수 없어 디지털 포렌식에 한계가 있다. 일반적으로 디지털 포렌식으로 수집한 문서들을 선별할 경우에는 파일의 수정된 날짜와 사건 관련 키워드의 포함 여부를 기준으로 한다. 문서 내 텍스트를 처리하는 인덱싱(Indexing) 작업을 통해 키워드 포함 여부를 확인할 수 있지만, 암호화된 문서는 작업이 불가능하다. 분석관들은 문서보안이 적용된 자료를 선별하기 전, 기업의 보안 담당자에게 복호화를 요청한다. 요청에 앞서 압축파일이나 이메일 첨부파일은 추출 작업이 필요하다. 압축파일은 폴더구조로 추출되는 반면, 첨부파일은 폴더구조가 없어 추출되면 연결성을 확인하기 어렵다. 한번 추출된 첨부파일은 이메일에 다시 첨부할 수 없어 이메일과 첨부파일 간의 연결성을 유지하며 분석하기 위해서는 파일 간의 부모 관계를 파악할 수 있는 Family 기능이 탑재된 e-Discovery 전문 솔루션이 필요하다[1]. 고가의 솔루션을 사용하기 어려운 분석관은 이메일 파일명과 추출된 첨부파일명을 비교하는 방식으로 분석한다. 해당 방식은 분석에 많은 시간이 소요되고, 동일한 파일명이 존재할 경우 혼란이 발생하는 문제가 있다.

따라서 본 논문에서는 이메일에서 추출되어 복호화가 완료된 첨부파일을 다시 이메일에 첨부하여 연결성을 확보할 수 있는 도구를 구현하고 검증하였다. 범용적으로 사용되는 EML 혹은 MSG 형태의 이메일에 대해서 복호화가 필요한 첨부파일을 추출하고, 추출 과정에서 기록된 원본에 복호화된 첨부파일을 재첨부하여 저장하도록 설계하였다. 재첨부되어 저장된 이메일(이하, '재첨부된 이메일'이라고 칭한다)의 보낸 시간, 보낸 사람, 받은 사람 등의 주요한 메타데이터가 원본과 동일하게 유지되는 것을 확인하였다. 제2장에서는 실험 대상 문서보안 솔루션, 이메일과 상용 솔루션에 대해 분석하고, 기존 관련 연구와 한계를 살펴본다. 제3장에서는 이메일 재첨부 실험 설계 및 구현 방안에 대해 제시한다. 제4장에서는 구현된 도구의 검증과 활용 방안에 관해 설명한다. 도구는 문서보안이 설정된 이메일 분석에 필요한 과정을 단축시키고, 오류를 줄일 수 있을 것으로 기대된다.

Ⅱ. 선행 연구

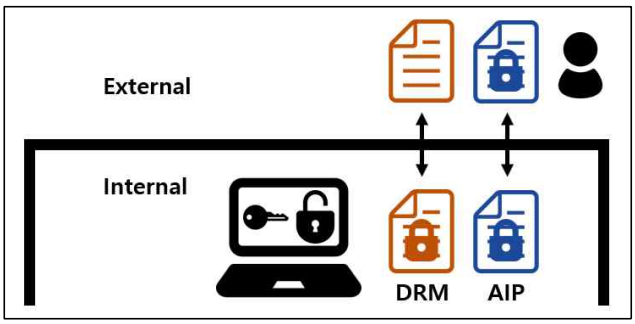
2.1. 문서보안의 종류

문서보안은 기업에서 데이터 유출을 방지하기 위해 사용되고, 보호 방식에 따라 다양한 솔루션이 존재한다. 사용자 계정별 권한 부여와 제한을 통한 접근 통제, 외부 저장장치 복사 및 인터넷 업로드의 통제, PC 내 자료 저장 통제, 그리고 문서 자체를 암호화하여 인가받지 못한 사용자들이 열람할 수 없게 하는 방법 등이 존재한다. 디지털 포렌식에서는 암호화 방식의 문서보안 솔루션이 문제가 된다. 국내에서는 DRM(Digital Rights Management)으로 알려진 솔루션 에이전트 기반의 암호화 문서보안을 주로 사용하는데, PC에 저장되는 문서를 암호화하여 열람 시에만 에이전트 프로그램이 복호화한다. 수집된 문서들은 보안 해제 권한이 있는 계정으로 에이전트 프로그램에 로그인하여 복호화를 진행해야만 외부망에서도 열람할 수 있다.

암호화 방식의 문서보안 솔루션에는 DRM으로 불리는 솔루션 에이전트 기반의 문서보안 외에도, AIP(Azure Information Protection)로 불리는 클라우드 플랫폼 기반의 솔루션이 존재한다. AIP는 Microsoft社가 개발한 문서보안으로, DRM과 달리 기업 내부망 밖으로 공유할 때 복호화를 진행하지 않아도 Azure RMS 포탈에서 권한을 인증받아 열람할 수 있다[그림 1]. DRM은 자료가 복호화되어 유출되면 제어할 수 없는 반면, AIP는 읽기 권한을 회수하여 공유를 제어할 수 있다. AIP를 통해 암호화된 자료들은 외부망에서 복호화 없이도 공유될 수 있지만, 인덱싱이 불가능해서 디지털 포렌식을 위해 복호화 과정이 필요하다.

이러한 이유로 수사기관은 기업 담당자의 참관 하에 수집된 자료를 복호화하고 서명을 받아 문서보안이 적용되지 않은 자료를 인계한다.

본 논문은 성균관대학교 및 교육부, 한국연구재단의 4단계 두뇌한국21 사업 대학원혁신으로 지원된 연구임
▪ Received 24 November 2021, Revised 24 November 2021, Accepted 4 March 2022
▪ 제1저자(First Author) : Donggeun Jang (Email : s4mjang@gmail.com)
▪ 교신저자(Corresponding Author) : Gibum Kim (Email : freekgb02@gmail.com)



〈Figure 1〉 Difference between AIP and DRM

국내 기업은 AIP에서 한컴오피스 문서를 지원하지 않는다는 점 때문에 DRM만 사용하거나 DRM과 AIP 모두를 사용하는 경우가 많아, 두 종류의 문서보안이 적용된 자료들을 선별하여 복호화를 진행할 필요가 있다. 〔표 1〕은 Microsoft社의 AIP 외에도 DRM 솔루션 중 국내에서 가장 많이 사용되는 Softcamp, Fasoo, Markany에서 지원하는 확장자이다〔2〕. 이 외에도 사진 파일에 대한 암호화도 지원하나, 기업에서는 해당 기능을 비활성화한 경우가 많다.

〈Table 1〉 Extensions supported by encryption-based document security

Category	DRM (Based on Solution Agent)	AIP (Based on Cloud Platform)
MS Office	DOC, DOCX, PPT, PPTX XLS,XLSX,XLSM,XLSB etc.	DOC, DOCX, PPT, PPTX XLS,XLSX,XLSM,XLSB etc.
Hancom Office	HWP, CELL, NXL, SHOW etc.	Not Supported
Text	TXT, RTF, CSV	TXT, RTF, CSV
Adobe Reader	PDF	PDF

복호화 시간을 최소화하기 위해서는 DRM 혹은 AIP가 설정된 문서를 선별해야 한다. 〔표 1〕의 확장자를 기반으로 1차 선별한 후, DRM이 설정된 파일의 헤더 부분에 존재하는 〔표 2〕의 시그니처를 통해 탐지할 수 있으며, AIP가 설정된 문서는 Microsoft社에서 제공하는 SDK〔3〕로 탐지할 수 있다.

〈Table 2〉 DRM Signatures

Solution	Signature	Text
Softcamp	\x53\x43\x44\x53\x41\x30\x30\x32	SCDSA002
	\x53\x43\x44\x53\x41\x30\x30\x34	SCDSA004
	\x53\x43\x44\x53\x53\x30\x30\x32	SCDSS002
Fasoo	\x3C\x21\x2D\x2D\x20\x46\x61\x73\x6F\x6F\x53\x65\x63\x75\x72\x65	〈!-- FasooSecure
	\x2E\x20\x44\x52\x4D\x4F\x4E\x45\x20\x20\x54\x68\x69\x73\x20\x44	. DRMONE This D
	\x3C\x21\x2D\x2D\x20\x49\x4E\x43\x4F\x50\x53\x20\x53\x45\x43\x55	〈!-- INCOPS SECU
Markany	\x3C\x44\x4F\x43\x55\x4D\x45\x4E\x54\x20\x53\x41\x46\x45\x52	〈DOCUMENT SAFER
	\x3C\x4D\x41\x52\x4B\x41\x4E\x59\x5F\x44\x4F\x43\x55\x4D\x45\x54	〈MARKANY_DOCUMET
	\x8C\x17\x2E\x9A\x10\xB6\xF6\xF0\xD4\x9C\xA9\xCA\xEE\x80\xBC\x9E	Œ.š¶öðÔœ©Êî€‡ž
	\xC5\x63\x4A\x59\x8C\x1B\xCE\xBD\xBB\xDE\xC3\x59\x3E\xD1\x78\x1C	ÅcJYŒÎ½»ÞÃY〉Ñx

2.2. 이메일 파일의 종류

기업 내 업무환경에서 이메일은 개별 저장되어 있기도 하지만, 이메일 클라이언트마다 다른 고유 형태의 아카이브 파일로도 저장된다. 기업은 자체 개발한 이메일을 사용하거나, Microsoft社의 Outlook이나 Exchange를 주로 사용한다. 자체 개발한 이메일은 내보내기 기능을 통해 EML 형태로 저장할 수 있고, 한 번에 여러 이메일을 압축 파일 형태로 저장할 수도 있다. Outlook은 윈도우즈의 경우 PST, OST 형태〔4〕로, 맥OS의 경우는 OLM 파일로 저장하고, Exchange는 EDB 혹은 DBX 형태로 저장한다. 이 외의 이메일 클

라이언트로 Mozila社의 Thunderbird는 우편함 별로 MBOX 형태로 저장하며, MBOX는 EML 형태의 집합 구조로 구성되어 있어 개별의 EML로 저장할 수 있다.

특정 이메일 클라이언트의 아카이브 파일이 아닌, 개별의 이메일로 저장되는 유형은 EML, EMLX, MSG가 존재한다. EML 파일은 MIME RFC 822 Standard 형식의 이메일 파일로, Gmail, 네이버 메일 등과 같은 웹메일은 물론, 대부분의 이메일 클라이언트 솔루션들이 지원하고 있다. EMLX 파일은 맥OS 환경에서 사용되는 EML과 유사한 형태의 이메일 파일이다. MSG 파일은 Microsoft社가 만든 MAPI(Microsoft Outlook Messaging API) 형식의 이메일이며, Microsoft社의 이메일 클라이언트에서 개별로 저장할 때에 사용되는 형식이다.[5] 대부분의 이메일 아카이브 파일들은 개별 이메일 형태로 변환할 수 있고, EML 혹은 MSG 형태로 사건과 관련이 있는 이메일만을 선별 수집할 수 있다[표 3].

〈Table 3〉 Convertable file format of email client archive

Solution	Email client archive format	Convertable file format
Microsoft Outlook	PST, OST, OLM	EML, MSG
Microsoft Exchange	EDB, DBX	EML, MSG
Mozila Thunderbird	MBOX	EML

2.3. 상용 솔루션의 한계

e-Discovery 전문 데이터 처리 솔루션인 Nuix는 기존 파일들을 새로운 파일들로 대체하는 Replace 기능이 존재한다. Family 기능과 함께 활용하면 파일의 부모 관계가 유지되기 때문에 문서보안이 적용된 첨부파일 대신 복호화된 첨부파일로 대체할 수 있다. 하지만 솔루션 내에서만 확인할 수 있고, 이메일을 추출하면 첨부파일이 대체되지 않은 이메일 원본이 저장된다. 이러한 이유로 상용 솔루션에 대한 의존도를 낮출 수 있는 기술 개발이 필요하다.

2.4. 관련 연구

김기범[6]은 디지털 증거의 동일성은 원칙적으로 해시값을 통해 증명할 수 있으나, 예외 상황으로 해시값이 다르더라도 동일성을 증명할 방법이 존재한다면 인정할 수 있다고 보았다. 문서보안을 사용 중인 기업의 데이터를 수집할 때 복호화 과정에서 발생하며, 같은 작업을 여러 번 반복하여도 해시값이 동일하게 변경된다는 점을 증명하는 방안을 제시하였다. 메타데이터를 활용한 동일성 검증 방식은 구현 도구의 동일성 검증에 대한 근거를 제시해주고 있다.

김상윤[7]은 이메일 내 대용량 첨부파일 URL에서 일괄적으로 다운로드할 수 있는 도구를 구현하였다. 대용량 첨부파일은 일반적인 첨부파일 구조로 되어 있지 않고, 다운로드할 수 있는 URL이 참조되는 형식이기 때문에 이를 탐지하여 다운로드하는 방식으로 설계하였다. 원본 이메일과 다운로드 된 자료를 이중 해시 계산을 통해 증거가 처리되는 과정에서의 위·변조되지 않음을 증명할 수 있다고 설명하였다.

Lucia[8]는 e-Discovery 도입 초기에 발생한 사례를 통해 메타데이터의 중요성에 관해 설명한다. 엑셀 스프레드시트가 TIFF 형태로 변환되어 제출되었을 때, 일부 열과 수식을 숨긴 뒤에 변환되었다는 사실이 발견된 사례를 소개했다. 재판부는 원본 내 파일 이름, 수정 날짜, 작성자 등의 메타데이터를 지운 흔적을 근거로 온전한 메타데이터와 스프레드시트를 제출하도록 명령하였다. 제출 형식의 변환이 있다면 증거의 동일성을 위해 메타데이터 비교가 필수적임을 시사한다.

이신형[9]은 국내 환경에 맞게 법률 및 기업문화를 고려한 e-Discovery 대응 절차를 제시한다. 국내 기업은 문서보안 솔루션을 구축하는 경우가 많아 복호화 작업이 데이터 수집 전에 진행되어야 한다고 말한다. 분석을 목적으로 생산(Production) 형태를 변형하는 e-Discovery 관행 때문에 복호화 과정에서 변형되는 문서와 본 논문에서 변형되는 이메일 또한 증거 능력을 인정받을 수 있을 것이다.

이해진[10]은 인터넷 포털사의 이메일 압수수색을 위해 메일 헤더 정보를 이용한 선별수집에 대해 연구하였다. 사건과 관련된 기간에 생성된 이메일을 선별하는 기존 방법에 비해 헤더 정보를 활용하면 개인의 프라이버시 침해를 방지하고, 사건과 연관성이 있는 자료를 선별할 수 있다고 설명한다. 이메일의 헤더 정보의 중요성을 확인할 수 있었지만, 키워드 검색을 통한 선별 방법은 다루지 않았다.

신연식[11]은 DRM과 같은 문서보안을 사용 중인 기업을 대상으로 압수수색을 진행하였을 때, 이메일 내

첨부된 검색 불가능한 파일들에 대해서도 현장에서 선별이 가능한 도구의 필요성을 설명한다. 비밀번호가 설정되어 있거나 텍스트가 없는 스캔 문서, 문서보안이 설정된 첨부파일들에 대해서 선택적으로 추출하여 복호화 시간을 단축하고, 선별 과정에서 중요 증거가 누락될 수 있는 오류를 방지하도록 하였다. 하지만, 현장에서의 수집 단계에 관한 연구를 중심으로 진행되었고, 선별된 이메일의 분석 과정에 대해서는 논하고 있지 않다.

Tingen[12]은 e-Discovery에서 분석 시간을 줄일 수 있는 키워드 검색과 개념 검색의 중요성을 말한다. 수동 분석은 e-Discovery 과정에서 높은 비용이 발생한다. 이를 해결하기 위한 개념 검색은 제공된 정보와 개념적으로 유사한 정보를 자동으로 검색할 수 있는 기능으로, 분석 시간을 줄일 수 있다고 설명한다. 다양한 기술적 방법을 통한 분석 방법을 제시하지만, 상용 솔루션에 의존적인 한계가 있다.

Grossman[13]은 TAR(Technology-Assisted Review) 사용의 장단점에 관해 설명한다. TAR는 분석한 자료들을 반복 기계학습하여 관련성 순위가 높은 자료를 우선하여 분석하게끔 한다. 검색자의 역량에 따라 효과가 결정되는 키워드 검색 방식에 비해 비용 절감의 효과를 볼 수 있다고 말한다. 기술적으로 분석할 자료의 양을 줄이는 분석 개선 방안을 설명하고 있으나, 상용 리뷰 솔루션에 의존적이며 한국어는 영어에 비해 기능의 신뢰도가 낮다는 문제점이 있다.

김헌[14]은 e-Discovery를 위한 도구로 자료를 분류 및 검색할 수 있도록 오픈소스를 통해 구현하였다. K-means 클러스터링을 활용하여 연관성이 있는 문서 집합을 분류함으로써 분석에 필요한 문서를 선별할 수 있고, 검색된 결과를 확인할 수 있도록 하였다. 해외의 e-Discovery 솔루션에 의존적인 문제를 프로토타입 도구를 통해 개선하였으나, 기존 상용 솔루션의 한계점은 논하지 않았다.

이처럼, e-Discovery와 디지털 포렌식 관점에서 증거의 동일성, 수집과 분석방안에 관한 연구는 다양하게 진행되고 있으나, 문서보안이 적용된 자료의 분석 개선 방안은 아직 연구되지 않았다.

Ⅲ. 실험 및 도구 구현

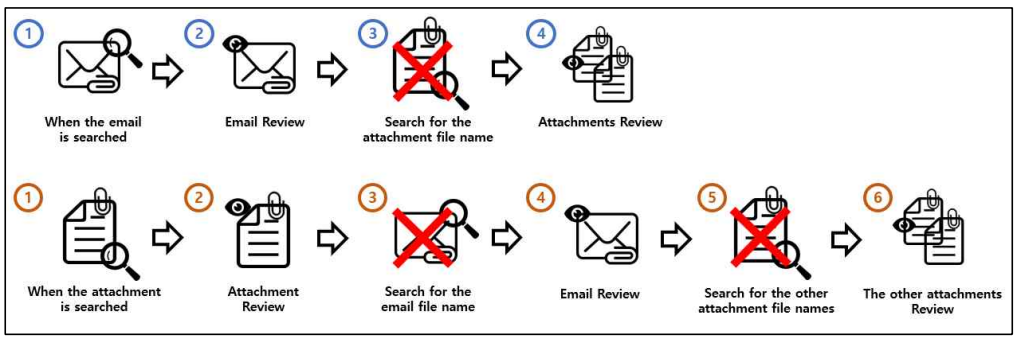
3.1. 실험 설계

실험은 우리나라 기업이 주로 사용하는 4개의 문서보안 솔루션과 3개의 이메일 클라이언트를 대상으로 하였다[표 4]. 입력 데이터는 기존의 디지털 포렌식 수집 방법에 따라 저장한 이메일 파일들을 모은 폴더 혹은 파일로 한다. PST, OST 등과 같은 이메일 아카이브 파일은 모두 EML 혹은 MSG 형태로 변환하여 저장해야 하고, 실험 대상의 EML, MSG 파일에 첨부된 첨부파일은 일반적으로 기업의 업무를 위해 사용되는 오피스 문서, 압축파일, 사진, 실행파일 등으로 구성한다.

〈Table 4〉 The solution conditions in the experiment.

Category	Solutions
Document Security	Softcamp, Fasoo, Markany, AIP(Azure Information Protection)
Email Client	Outlook, Exchange, Thunderbird

실험은 이메일 분석을 진행하는 기존 방식에서 파일명 탐색 단계를 생략하여 분석 시간을 단축하고, 이메일과 추출된 첨부파일의 연결성을 유지할 수 있는 도구 구현을 목표로 한다[그림 2].



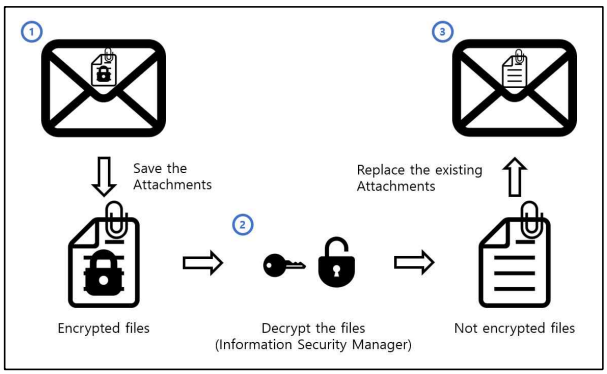
〈Figure 2〉 Process of reviewing Emails with reattached files

3.2. 실험 내용

첫 번째로, 지정한 폴더에서 최하위 폴더까지 확장자가 EML 혹은 MSG인 파일들에 대해 탐색한다. 탐색된 파일들은 순번과 함께 목록화 되어 지정한 추출 폴더 내에 생성된 번호에 맞는 폴더에 첨부파일들을 추출한다. 첨부파일 중 문서보안 솔루션이 지원하는 문서 파일들과 압축 파일(zip, alz, 7z, egg, rar)만을 추출하여 저장한다. 이후, DRM과 AIP의 적용 여부를 확인하여 적용되지 않은 문서들은 삭제한다. 구현된 도구는 추출된 첨부파일, 첨부되어 있던 이메일의 파일명과 MD5 해시값을 리스트에 저장하여 원본이 훼손되지 않았음을 증명한다.

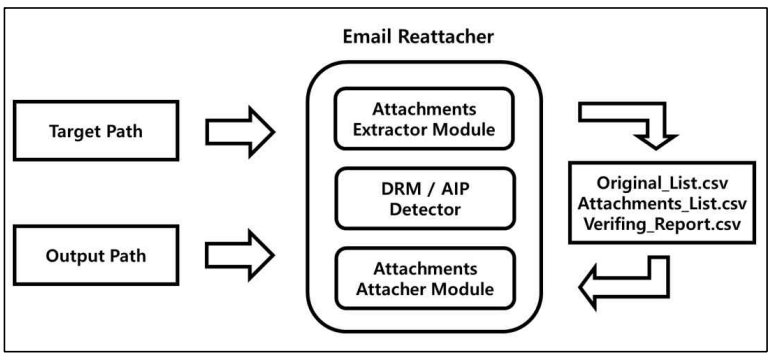
두 번째로, 첨부파일이 저장된 추출 폴더 내에 파일들을 기업의 보안 관리자에게 전달하여 복호화를 요청한다. 만약, 첨부파일 중에 압축파일들이 존재한다면, 일괄적으로 원본 압축 파일명과 동일한 폴더에 압축을 해제하여 저장한 뒤에 전달한다. 전달한 파일들이 복호화가 끝나면, 본래 압축이 되어있었던 폴더들을 동일한 확장자로 다시 압축을 진행해야 한다.

마지막으로, 복호화된 첨부파일 추출 폴더와 원본 이메일 파일들이 저장된 폴더를 지정하여 기존의 원본에 첨부된 파일 중 추출되어 복호화를 진행했었던 파일들을 첨부 해제한다. 리스트에 있는 이메일 순번대로 복호화된 첨부파일들을 이메일에 다시 첨부할 때, 파일 리스트에는 복호화된 첨부파일과 재첨부된 이메일의 경로명, 그리고 모든 파일의 MD5 해시를 기록함으로써 원본 이메일이 변경되는 실험 과정을 기록하여 증거의 동일성을 나타낸다[그림 3].



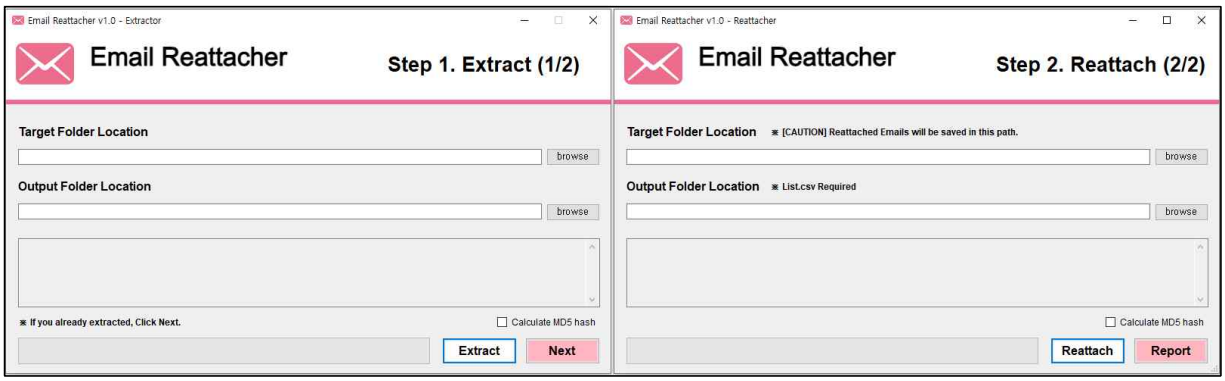
〈Figure 3〉 How to reattach attachments to the Email

3.3 도구 구현



〈Figure 4〉 Structure of Email Reattacher

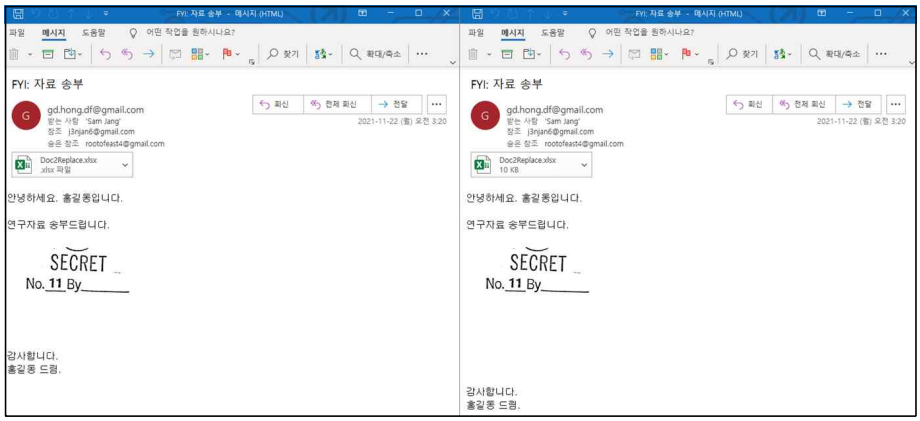
[그림 5]에서 Step 1의 Extract 버튼은 Target 이메일 파일들에 대해 복호화가 필요한 첨부파일들을 추출한다. 이후 사용자는 한 폴더에 추출된 첨부파일을 복호화 작업을 한다. Step 2의 Reattach 버튼을 통해 복호화된 첨부파일들을 재첨부할 수 있다. 전체 과정에서 손상된 파일 등의 이유로 에러가 발생하는 경우, 로그 창에 출력되어 확인할 수 있다. 결과로 생성되는 Original_List.csv, Attachments_List.csv, Verifying_Report.csv는 작업 내용을 기록하는 기능뿐만 아니라, 원본 이메일과 첨부파일들의 MD5 해시와 재첨부된 이메일들의 MD5 해시를 기록하여 실행 전후의 동일성을 입증하는 용도로 사용된다.



〈Figure 5〉 UI of Email Reattacher

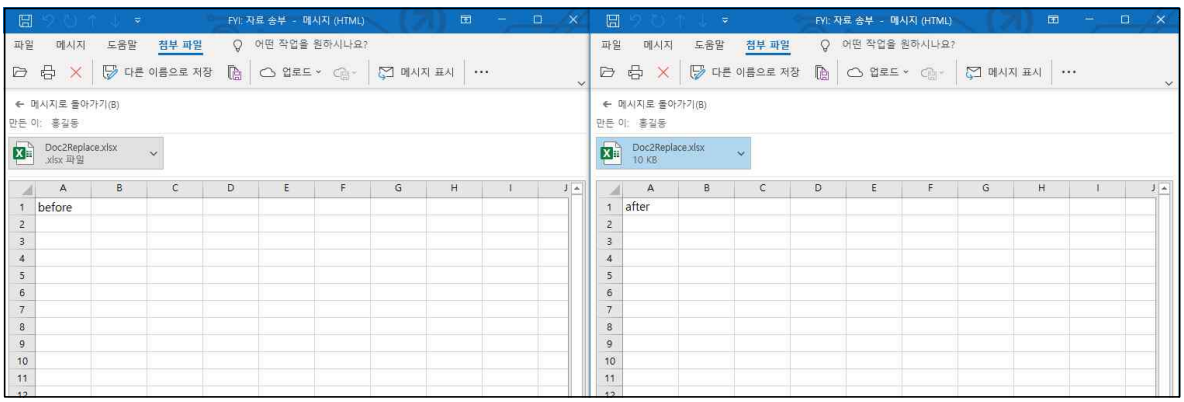
3.4. 실험 결과

구현된 도구를 이용하여 재첨부한 이메일 내 본문 내용, 보낸 사람, 보낸 시간 등의 속성 정보들은 모두 동일하며, 〔그림 6〕과 같이 본문 내용의 줄 바꿈 수를 제외하고 주요한 메타데이터가 동일한 것을 확인할 수 있었다. 프로필 사진은 현재 Outlook 솔루션에 로그인된 계정에서 연동되어 불러오는 사진이며, 원본 이메일 파일에 저장되는 사진이 아니므로 검증에서는 다루지 않는다.



〈Figure 6〉 Comparison between original Email and Email with reattached file

재첨부 실험 이후에, 원본과 다른 첨부파일이 첨부된 것을 확인할 수 있었다. 첨부파일이 변경되는 실험 결과를 확인하기 위하여, 실험 조건의 문서보안이 적용된 문서가 아닌 일반 엑셀 문서의 내용을 변경하여 재첨부하는 방식으로 실험하였다. 〔그림 7〕과 같이 복호화한 첨부파일을 재첨부한 후에는 이메일 본문 내용을 분석한 뒤, 바로 첨부파일들을 분석할 수 있게 되었다. 재첨부 과정에서 첨부파일의 메타데이터는 훼손되지 않으며, 최초 작성자 및 마지막 작성자 등과 같은 오피스 문서 내 정보는 동일하였다. 이는 문서보안이 적용된 첨부파일을 복호화하였을 때도 유지되는 정보이다.



〈Figure 7〉 Comparison between original attachment and reattached attachment

Ⅳ. 실험 검증 및 활용 방안

4.1. 결과물 검증

구현된 도구로 재첨부된 이메일은 첨부파일 영역뿐만 아니라 전체 이메일의 구조가 재구성되어 다른 형태로 저장된다. 실험 결과물 내에서 분석에 중요한 증거로 사용되는 메타데이터 값들을 비교하여 재첨부된 이메일 파일이 증거로써 활용 가능한지 검증하였다. 실험에 사용한 1만여 개의 EML, MSG 파일들에 대해 검증한 결과, 이메일 구조와 첨부파일의 용량이 변경됨에 따라 재첨부된 이메일의 용량은 달라질 수 있었으며, EML 형태의 파일은 MSG 형태에 비해 단순하다는 파일 구조의 특성으로 재첨부 전후의 용량 차이가 적다는 특징을 보였다. 일부 MSG 파일에서 이메일 본문 내용의 줄 바꿈 수가 변경되었으나, 사진이나 중요한 내용이 소실되거나 추가되는 경우는 존재하지 않았다. 따라서 분석 과정에 있어 문제가 되지 않는다고 판단되며, 해시값이 달라지더라도 이메일 고유의 ID인 Message-ID가 동일하다는 점, 항상 같은 결과물이 저장된다는 점을 이유로 증거의 동일성을 증명할 수 있을 것이다[표 5].

〈Table 5〉 Verification of the MSG file with Reattached file

Category	Original Email	Email with Reattached file	Verify
Title	FYI: 자료 송부	FYI: 자료 송부	Same
SentDate	2021-11-22 (월) 오전 3:20	2021-11-22 (월) 오전 3:20	Same
From	'gd.hong.df@gmail.com' 〈gd.hong.df@gmail.com〉	'gd.hong.df@gmail.com' 〈gd.hong.df@gmail.com〉	Same
To	'Sam Jang' 〈s4mjang@gmail.com〉	'Sam Jang' 〈s4mjang@gmail.com〉	Same
CC	'j3njan6@gmail.com' 〈j3njan6@gmail.com〉	'j3njan6@gmail.com' 〈j3njan6@gmail.com〉	Same
BCC	'rootofeast4@gmail.com' 〈rootofeast4@gmail.com〉	'rootofeast4@gmail.com' 〈rootofeast4@gmail.com〉	Same
Contents	안녕하세요. 홍길동입니다. 연구자료 송부드립니다.  감사합니다. 홍길동 드림.	안녕하세요. 홍길동입니다. 연구자료 송부드립니다.  감사합니다. 홍길동 드림.	Almost Same
Importance	중간	중간	Same
Message-ID	SL2P216MB1032445529EA7A73C6106B53BC9E9@SL2P216MB1032.KORP216.PROD.OUTLOOK.COM	SL2P216MB1032445529EA7A73C6106B53BC9E9@SL2P216MB1032.KORP216.PROD.OUTLOOK.COM	Same
last modified date	2021-11-22 (월) 오전 3:20	2021-11-22 (월) 오전 3:20	Same
File Size	41.5 KB	40.0 KB	Not Same
MD5 Hash	605F3CD1205DACB9CB661B09066DA04B	CB284C1B023A2EFE18E8F21DF9A044FE	Not Same

4.2. 활용 방안

대상 기업이 문서보안을 사용하는 경우, 고가의 상용 솔루션을 사용하기 어려운 분석관들은 이메일과 첨부파일의 연결성을 파악하기 위해 많은 시간과 노력을 소모한다. 구현 도구인 Email Reattacher는 상용 솔루션 없이 연결성을 유지해 분석 시간을 단축할 수 있다. 기존 압수·수색 과정에서 수사관은 이메일과 복호화된 첨부파일을 별도로 선별하기 때문에, 첨부파일에는 없는 키워드가 이메일 본문에 존재한다면 첨부파일을 수집하지 못한다. 구현한 도구로 재첨부할 경우, 한 번에 선별하여 누락하지 않고 모두 수집할 수 있다. 또한, 기존 방식에서는 이메일 내 첨부파일과 추출된 첨부파일이 중복으로 저장되어 수집되는 데이터의 용량이 증가하는 반면,

재첨부된 이메일은 원본 용량에 유사하게 저장된다. 구현한 도구를 통해 이메일 전처리를 했을 경우, 작업 내용이 기록되는 Verifying_Report.csv에 작업 전후의 해시값이 기록된다. 암호화 방식의 문서보안을 해제했을 경우에도 해시값은 변경되며, 복호화된 문서보안 파일의 수집 과정과 동일하게 피조사자의 참관하에 재첨부 과정을 진행하여 피조사자의 서명을 받았을 경우 증거로 활용할 수 있을 것이다.

4.3. 한계

Email Reattacher로 재첨부한 결과물은 파일 구조의 전체적인 변경을 통해 저장되기 때문에 전체 혹은 부분 해시를 계산하는 방법으로는 동일성을 증명할 수 없다. 분석에 필요한 정보를 비교하여 검증하는 과정을 통해 차이점이 분석에 미치는 영향은 미미하다는 결론을 냈지만, 일반적인 증거의 동일성을 판단하는 기준인 해시값과 용량이 변경되기 때문에 추가 설명이 필요하다. 재첨부된 이메일은 분석의 효율성을 목적으로 수집되고, 원본 이메일과 문서보안이 해제된 첨부파일 또한 수집되기 때문에 일부 차이가 있더라도 비교할 수 있다. 복호화 및 재첨부의 반복 실험을 통해 정보의 변경이 불가피한 점을 증명하고, 주요 메타데이터가 유지되는 것을 보여 증거의 동일성을 입증해야 할 것이다.

첨부파일이 압축파일인 경우, 사용자가 별도로 압축을 해제하고, 문서보안 복호화 이후에 다시 동일한 파일명으로 압축해야 한다. 이메일 내에 이메일 파일이 첨부되는 경우에는 재귀적으로 도구를 실행시켜야만 모든 이메일의 첨부파일들을 재첨부할 수 있다. 이처럼 사용자는 도구의 한계를 인지하여 활용해야만 오류를 줄일 수 있을 것이다.

Ⅴ. 결 론

본 연구에서 실험을 통해 이메일에 첨부된 문서보안이 적용된 첨부파일에 대해 효율적으로 분석하는 방법을 도구로 구현하여 제시하였다. 제3장에서 암호화 방식의 문서보안과 이메일 데이터의 유형에 대해 살펴보았다. 이를 통해 실험에서 사용되는 이메일과 첨부파일의 종류를 한정할 수 있었으며, 복호화된 이메일 첨부파일들을 원본 이메일에 재첨부하는 도구를 구현할 수 있었다. 제4장에서 도구 및 결과물을 검증함으로써, 해당 재첨부 방식이 주요 정보를 훼손하지 않으며 기존 방식에 비해 저장용량 감소, 분석 과정 개선, 수집 시 오류 발생 가능성을 개선할 수 있다는 결과를 도출할 수 있었다. e-Discovery 전문 솔루션에 본 연구에서 구현한 재첨부 기능이 추가된다면, Replace 기능을 통해 대체한 파일을 저장할 수 있게 되어 솔루션 사용에 따른 제약 없이 분석하는데 용이하고, 부모 관계를 유지할 수 있을 것이다. 향후에도 기업의 내부 자료 유출 방지를 위한 문서보안의 도입은 확대될 것으로 예상되며, 문서뿐만 아니라 이메일 자체, 영상, 사진, 음성 파일 등에 대한 암호화까지 확대될 수 있다. e-Discovery 및 압수·수색의 관점에서 이러한 암호화된 자료의 수집 방법과 효율적인 분석 방법의 연구가 계속되기를 바란다.

참 고 문 헌 (References)

〔1〕 Nuix, 'Nuix Workstation all about processing', https://www.nuix.com/nuix-workstation-all-about-processing, 2022. 3. 7. confimed.

〔2〕 Microsoft, 'File types supported by the Microsoft Information Protection SDK', https://docs.microsoft.com/en-us/information-protection/develop/concept-supported-filetypes; SOFTCAMP, '암호화 지원 대상', https://docs.softcamp.co.kr/article.php?id=311; FASOO, 'Fasoo Enterprise DRM', https://www.fasoo.com/products/fasoo-enterprise-drm; MarkAny, 'Document SAFER', https://www.markany.com/products/data-security/markany-drm/, 2022. 3. 7. confimed.

〔3〕 Microsoft, 'Azure Information Protection Samples', https://docs.microsoft.com/ko-kr/samples/azure-samples/azure-information-protection-samples/azure-information-protection-samples/, 2022. 3. 7. confimed.

〔4〕 Microsoft, 'Introduction to Outlook Data Files (.pst and .ost)', https://support.microsoft.com/en-us/office/introduction-to-outlook-data-files-pst-and-ost-222eaf92-a995-45d9-bde2-f331f60e2790, 2022. 3. 7. confimed.

〔5〕 Microsoft, '〔MS-OXMSG〕: Outlook Item (.msg) File Format', https://docs.microsoft.com/en-us/openspecs/exchange_server_protocols/ms-oxmsg/b046868c-9fbf-41ae-9ffb-8de2bd4eec82, 2022. 3. 7. confimed.

〔6〕 Gi-bum Kim, "A Study on Hash Function in Criminal Procedure", Korean Criminology Review, Vol.29, No.2, pp.199-225, 2018.06

〔7〕 Sang Yoon Kim, "Extract Method of Large Attachments in E-mail Forensics", https://dcollection.korea.ac.kr/public_resource/pdf/000000056756_20211121201752.pdf

〔8〕 Cucu Luicia, "The Requirement for Metadata Production under Williams v. Sprint/United Management Co.: An Unnecessary Burden for Litigants Engaged in Electronic Discovery", Cornell Law Review, Vol.93, No.1, pp.221-242, 2007.11

〔9〕 Shinhyung Lee, "Effective Domestic e-Discovery Procedures", Journal of The Korea Institute of Information Security and Cryptology, Vol.26, No.5, pp.1171-1183, 2016.10

〔10〕 Hae-Jin Lee, "E-mail Header-Based Search and Seizure for Internet Portal DigitalForensics", Journal of The Korea Institute of Information Security and Cryptology, Vol.28, No.5, pp.1129-1140, 2018.10

〔11〕 신연식, "검색불가능 이메일 첨부파일에 대한 현장 탐지기법 연구", https://s-space.snu.ac.kr/bitstream/10371/151400/1/000000154166.pdf

〔12〕 Jacob Tingen, "Technologies-That-Must-Not-Be-Named: Understanding and Implementing Advanced Search Technologies in E-Discovery", Richmond Journal of Law & Technology, Vol.19, No.1, pp.1-49, 2012.01

〔13〕 Maria R. Grossman, "Comments on The Implications of Rule 26(g) on the Use of Technology-Assisted Review", The Federal Courts Law Review, Vol.7, No.1, pp.285-313, 2014.06

〔14〕 Hun Kim, "Design and Implementation of a tool for ESI Categorization and Search in e-Discovery", Proceedings of the Korea information Processing Society Conference, Vol.18, No.2, pp.819-822, 2011.11

저 자 소 개

장 동 근 (Donggeun Jang)
준회원

2019년 8월 : 가천대학교 컴퓨터공학과 졸업
2021년 2월 : 성균관대학교 과학수사학과 석사과정(디지털 포렌식 전공)
2019년 7월~현재 : HM Company Digital Forensic Service 본부 매니저
관심분야 : 디지털 포렌식, 기업범죄, 금융범죄, e-Discovery

김 기 범 (Gibum Kim)
정회원

1997년 3월 ~ 2014년 2월 : 경찰청, 서울경찰청 사이버수사 부서 등 근무
2014년 3월 ~ 2020년 2월 : 경찰대학 경찰학과 학과장/교수요원
2017년 2월 : 고려대학교 정보보호대학원 공학박사
2020년 3월 - 현재 : 성균관대학교 일반대학원 과학수사학과 학과장/부교수
관심분야 : 사이버범죄, 디지털 포렌식, 과학기술치안, 암호화폐