情報保護學會論文誌
第 21 卷 第 6 號, 2011. 12

이메일 클라이언트 내의 삭제된 이메일 복원에 관한 연구*

정 초 롱† , 이 근 기, 이 상 진‡
고려대학교 정보보호연구원

Recovery Techniques for Deleted Email Items in Email Client*

Chorong Jeong† , Keun-gi Lee, Sangjin Lee‡
Center for Information Security & Technologies, Korea University

요 약

기업은 커뮤니케이션 수단 및 업무처리를 위한 수단으로 이메일을 사용한다. 이메일은 비즈니스의 주요업무를 처리하기 때문에 기업 정보의 많은 부분을 보존하고 있다. 그러므로 기업 기밀 유출과 같은 사건 조사 시에 이메일은 중요한 증거가 될 수 있다. 하지만 용의자가 의도적으로 이메일을 삭제할 가능성도 있기 때문에, 삭제된 이메일을 복원하는 것은 포렌식 관점에서 매우 중요하다. 본 논문은 다양한 이메일 클라이언트 파일 내에 존재하는 삭제된 이메일 아이템 구조를 분석하고, 파일에 존재하는 이메일 복원 가능성과 복원 방안을 설명한다.

ABSTRACT

Corporations use e-mail as their primary method for internal communication and business processes. By their nature, the e-mails are in general used for major business processes that contain large amounts of business information. When there is a critical event, such as Technology leakage, an e-mail message can become important evidence. However, as there is a high likelihood that a suspect will intentionally erase an e-mail message, the ability to recover deleted e-mail is very important. This pater analyzes the deleted e-mail item structure in files of various e-mail clients, and explains the possibility and methods of recovery.

Keywords: Email Forensics, Email Recovery, Recovery, Email

I. 서 론

이메일은 개인의 의사소통 수단이기도 하지만 일반적으로 기업에서 주요한 업무 처리를 위한 핵심 수단이라 할 수 있다. 기업은 직원들과의 의사소통을 원활하게 하고 빠른 업무 처리를 위해 이메일을 사용한다.

이메일은 기업 기밀 유출과 같은 범죄의 수단으로 사용할 수 있다. 이러한 사건의 조사 과정에서 조사관은 이메일에 대한 포렌식 조사를 하게 된다. 용의자는 자신의 잘못을 은폐하기 위해 사건과 관련된 이메일을 삭제했을 가능성이 크다. 그렇기 때문에 삭제된 이메일을 복원하는 것은 사건의 흐름을 파악하는데 도움을 주며, 사건에 중요한 단서를 제공할 수 있기 때문에 매우 중요하다.

이메일을 삭제하는 방법은 지운 편지함에서 삭제하는 방법과 지운 편지함을 거치지 않고 삭제하는 방법

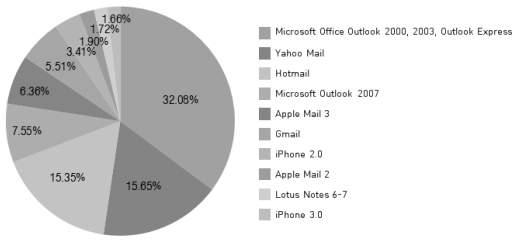
접수일(2011년 3월 23일), 수정일(2011년 7월 13일), 게재확정일(2011년 9월 6일)
* 본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발 사업의 일환으로 수행하였음 〔10035157, 실시간 분석을 위한 디지털 포렌식 기술 개발
† 주저자, cutycom_@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr

〔그림 1〕 Kernel for Outlook Express(왼쪽)와 PST Walker(오른쪽)

이 있다. 지운 편지함에서 삭제하는 방법은 이메일을 삭제하여 지운 편지함으로 이동시킨 후, 지운 편지함을 비우는 방식이다. 지운 편지함을 거치지 않는 이메일 삭제 방법은 마이크로소프트에서 정의한 윈도우 운영체제 단축키에 제시된 Shift+Delete 키를 사용하여 이메일을 영구적으로 삭제하는 방법을 말한다〔1〕. 삭제된 이메일 정보는 비할당 영역에 그대로 존재는 경우가 있다. 하지만 비할당 영역 분석은 데이터가 완전한 형태로 존재하지 않는 경우가 많기 때문에 분석에 어려움이 따른다〔2〕.

이메일 복원을 위한 도구가 많이 존재한다. 그 중 분석에 주로 활용되는 도구인 Kernel for Outlook Express〔3〕와 PST Walker〔4〕는 각각 DBX 파일과 PST 파일 내의 삭제된 이메일을 복원하는 도구이다. 각각의 도구는 해당하는 파일 내의 삭제된 이메일만 복원 가능하다. 또한 PST, DBX, EML 등의 다양한 파일을 지원하는 통합 복원 도구가 존재하지 않기 때문에 다양한 파일 내의 삭제된 이메일을 한 번에 분석하기 어렵다. 복원 기능을 테스트하기 위해 3개의 삭제된 이메일을 저장한 DBX 파일과 6개의 삭제된 이메일을 저장한 PST 파일을 생성하였다. 각각의 파일을 앞에서 언급한 두 복원 도구를 이용하여 삭제된 이메일을 복원한 결과, 〔그림 1〕과 같이 Kernel for Outlook Express는 3개의 삭제된 이메일 중 1개의 이메일을 복원하였으며, 한글을 지원하지 않아 복원한 이메일의 내용을 알아볼 수 없었다. PST Walker의 경우는 6개의 삭제된 이메일 중 3개를 복원하였다. 앞에서 언급한 도구뿐만 아니라 많은 다른 도구도 완벽한 복원 기능을 제공하지 못하며, 통합 분석을 할 수 없기 때문에 새로운 복원 도구가 필요하다.

2009년 6월, CampaignMointor는 6개월 동안 고객 3만 명을 대상으로 현재 사용하고 있는 이메일



〔그림 2〕 이메일 클라이언트 사용 통계

클라이언트에 대한 설문조사를 실시했다〔5〕. 〔그림 2〕와 같이 고객이 사용하는 이메일 클라이언트는 크게 웹 메일과 이메일 클라이언트로 분류할 수 있다. 웹 메일은 사용자가 웹브라우저를 이용하여 장소에 상관없이 사용할 수 있으며, 웹 서버에 메일 목록과 메일 본문, 전송 확인, 첨부파일 등의 데이터가 저장되어 있다〔6〕. 이런 환경에서 포렌식 조사를 하기 위해서는 웹브라우저 사용 흔적을 분석해야 한다. 본 논문에서는 메일을 저장하고 있는 파일을 대상으로 분석하기 때문에 웹 메일은 분석 대상에서 제외한다.

이메일 클라이언트는 Microsoft의 Office와 같이 응용프로그램을 사용하여 이메일을 송·수신한다. 설문조사의 결과를 보면, 대부분의 사람들이 사용하는 클라이언트는 Office Outlook 2000, 2003과 Outlook Express이며, 웹 메일인 Yahoo Mail, Hotmail, Gmail를 제외하고 그 다음으로 Office Outlook 2007을 많이 사용하는 것을 알 수 있다. 많은 고객이 사용하는 Apple Mail 2, 3은 Mac 기반의 운영체제에 기본적으로 설치되어 있다. iPhone 2.0과 3.0은 이메일 클라이언트를 사용하지 않고 아이폰이나 아이팟을 이용하여 이메일을 송·수신하는 것을 의미한다.

본 논문에서는 Windows OS 기반의 이메일 클라이언트인 Mozilla Thunderbird(MBOX)와 Outlook Express(DBX), Microsoft Office Outlook(PST, OST)에서 사용하는 이메일 파일을 대상으로 삭제된 이메일 구조와 복원 방안에 대해 설명한다.

II. 관련 연구

이메일 포렌식과 관련한 논문들은 Rachid Hadjidj et al.(2009)과 Farkhund Iqbal et al.(2010)의 연구와 같이 익명의 이메일에서 문체 특

성을 분석하여 저자를 판단하는 연구가 주로 이루어졌다[7,8]. 하지만 본 논문은 익명의 저자 판별이 아닌 이메일 저장 파일 내의 삭제된 이메일 복원에 초점을 맞춘다.

삭제된 이메일 구조 분석은 처음으로 Arne Schloh(2000)이 연구하였다[9]. 이 연구는 Microsoft Outlook Express의 이메일 파일인 DBX 파일을 대상으로 하였으며, 삭제되지 않은 이메일의 구조와 삭제된 이메일의 구조를 분석하여 파일 포맷 문서로 정리하였다. 또한 삭제되지 않은 이메일을 쉽게 분석할 수 있는 오픈 소스를 Arne Schloh 홈페이지에 공개하였다. 그 이후, J.B. Metz가 Microsoft Outlook 파일인 PST와 OST 파일을 대상으로 연구하였다[10]. 이 연구는 2008년부터 현재까지 꾸준히 PST 구조를 분석해왔다. Arne Schloh의 DBX 파일 분석과 마찬가지로 J.B. Metz의 PST 분석도 삭제되지 않은 이메일의 구조를 분석하여 파일 포맷 문서로 정리하였으며, PST 분석 도구를 개발하여, Sourceforge와 같은 오픈 소스 사이트에 공개하였다. J.B. Metz가 공개한 PST 분석 도구는 삭제되지 않은 이메일을 분석하여 그 결과를 파일로 저장한다. 또한 완벽하지 않지만 삭제된 이메일 복원 기능도 제공한다. 그리고 2010년, Microsoft에서 PST 포맷 관련 문서를 공개하였다.

본 논문에서는 DBX 파일 구조를 실제 파일에 적용하여 분석하고, 기존의 PST 복원 방안보다 더 많은 이메일 정보를 추출하는 향상된 이메일 복원 기법을 제시한다.

III. 삭제된 이메일 구조 및 복원 방안

[표 1]은 주로 사용되는 이메일 클라이언트와 각 이메일 클라이언트에서 사용하는 파일의 확장자명, 지원하는 OS를 정리한 것이다. 본 논문에서는 지원하는 OS가 Windows인 이메일 클라이언트를 대상으로 분석하였다.

Windows Live Mail에서 사용하는 파일 형식은 EML이다. EML은 MIME형식으로 구성되며, 하나의 파일에 하나의 이메일이 저장된다. 이메일을 삭제하는 것은 EML 파일을 삭제하는 것을 의미하므로 파일 내에 삭제된 이메일은 존재할 수 없다. 하지만 파일을 삭제할 때, 파일시스템의 메타 영역에 존재하는 삭제 플래그 값만 변경하므로 파일의 메타 정보와 데이터 영역에 있는 파일의 데이터는 그대로 남아 있다.

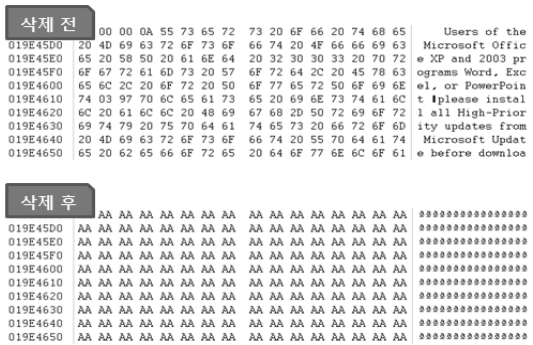
[표 1] 이메일 클라이언트 이름 및 확장자명, OS

이메일 클라이언트 이름	파일 확장자명	OS
Office Outlook (2000, 2003, 2007, 2010)	PST, OST	Windows
Outlook Express	DBX	Windows, MAC OS
Thunderbird	MBOX	Windows, Mac OS X, Linux
Windows Live Mail	EML	Windows Vista, 7
Lotus Notes 6-7	NSF	Windows, Unix, Linux, IBM
Apple Mail 2,3	EMLX	MAC OS X, iOS

그러므로 디스크 내에서 삭제된 EML 파일을 복원할 경우, 파일이 덮어써지지 않은 비할당 영역에 존재한다면 복원이 가능하다.

IBM Lotus Notes는 기업에서 사용하는 그룹웨어로 이메일을 저장하는 파일 형식은 NSF이다. 이 파일은 INBOX, SENT, TRASH 등의 편지함에 속한 모든 이메일을 저장한다[11]. 하지만 이메일을 삭제하면, [그림 3]과 같이 메일의 영역이 0xAA 등의 문자로 덮어써진다. 그러므로 NSF에서 삭제된 이메일 복원은 불가능하다.

하지만 나머지 3개의 이메일 클라이언트인 Outlook과 Outlook Express, Thunderbird는 삭제된 이메일 복원이 가능하다. 3개의 이메일 클라이언트는 파일에 삭제된 이메일 정보가 남아있다. 다음은 Mozilla의 Thunderbird(MBOX)와 Microsoft의 Outlook(PST, OST), Outlook Express(DBX)에 저장되어 있는 삭제된 이메일 정보의 구조와 복원 방안에 대해 설명한다.



[그림 3] NSF 파일 내의 일반 이메일 정보와 삭제된 이메일 정보

3.1 Mozilla Thunderbird (MBOX)

Thunderbird는 Mozilla에서 제공하는 이메일 클라이언트이다. Thunderbird는 〔표 2〕와 같이 MIME 기반의 MBOX 형식으로 이메일이 저장된 확장자가 없는 파일과 확장자가 .msf인 파일이 한 쌍으로 구성된다. 확장자가 없는 MBOX 형식의 파일은 다수의 이메일이 연속적으로 저장되어 있다〔12〕. 이 파일은 INBOX(받은 편지함), Sent(보낸 편지함), Trash(지운 편지함), Drafts(임시 보관함), Unsent message(보낼 편지함) 등의 편지함 당 하나의 파일이 생성된다. .msf 파일은 Mail Summary File로 MBOX 형식의 파일 내의 이메일을 관리하는 인덱스 파일이다. MBOX 형식의 파일과 같이 편지함 당 하나의 파일로 존재한다.

Thunderbird에서 삭제된 이메일을 복원하기 위해서는 MBOX 형식의 파일을 추출해야 한다. MBOX 형식의 파일은 삭제된 이메일을 포함하여 저장한다. 또한 텍스트로 구성되어 있기 때문에 MIME 형태의 단위로 계속 추출하면 삭제된 이메일까지도 쉽게 복원할 수 있다. MBOX 형식의 파일은 이메일의 상태 정보를 저장하는 헤더와 MIME 형식의 이메일 데이터 영역으로 구성된다. 이메일의 상태 정보를 저장하는 헤더에는 X-Mozilla-Status가 존재하며, 이 정보는 사용자가 이메일을 읽었는지, 이메일을 삭제했는지 플래그로 나타낸다. 삭제된 이메일을 복원하기 위해서는 X-Mozilla-Status가 0x0008을 포함하는지 확인하면 된다. 이는 MSG_FLAG_EXPUNGED를 나타내는 플래그 값이다.

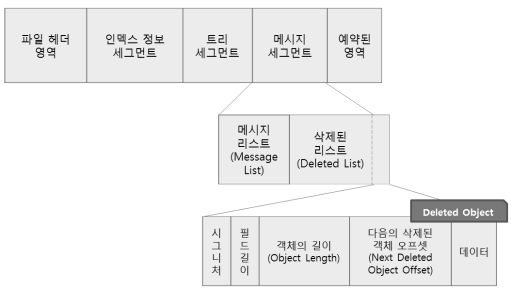
〔표 2〕 Inbox(MBOX 형식의 파일)와 Inbox.msf 파일 예

MBOX 형식의 파일 예	.msf 파일 예
From - Mon Feb 21 20:09:06 2011 X-Account-Key: account3 X-UIDL: 1297918244-525.42499.koreaLINK,S=1297918244525.4249900000 X-Mozilla-Status: 0009	// 〈!-- 〈mdb:mork:z v="1.4"/〉 --〉 〈 〈(a=c)〉 // (f=iso-8859-1) (B8=numHdrsToKeep)(B9=daysToKeepBodies)(BA=keepUnreadOnly) (BB=useServerDefaults)(BC=cleanupBodies) …

3.2 Microsoft Outlook Express (DBX)

Microsoft Outlook Express는 Windows에 기본적으로 설치되어 있는 이메일 클라이언트이다. 따라서 Windows 95, 98, XP에서 별도로 Microsoft Office를 설치하지 않아도 사용할 수 있다. Outlook Express에서 이메일을 저장하는 파일 형식은 DBX이다. DBX는 받은 편지함, 보낸 편지함, 지운 편지함, 보낼 편지함, 임시 보관함이라는 파일명으로 편지함 당 하나의 파일로 존재한다. 이 편지함은 이메일이 저장된 경우에 생성된다. 또한 이메일을 관리하기 위한 파일로 Folders.dbx, Pop3uidl.dbx, Offline.dbx가 존재한다. Pop3uidl.dbx는 이메일 서비스를 POP3 서버로 이용할 때 필요하다. 이 파일은 POP 서버에 메시지 복사본을 보관하도록 설정하면 생성되는 이메일의 고유 ID를 관리한다. Offline.dbx는 IMAP이나 Hotmail 계정이 존재하는 경우, 오프라인에서 동작한 IMAP이나 Hotmail의 행동을 저장하는 파일이다. Folders.dbx는 Outlook Express에서 받은 편지함, 보낸 편지함, 보낼 편지함, 지운 편지함, 임시 보관함과 같은 폴더 목록을 만드는 것을 지원하는 파일이다. DBX 파일의 구조는 〔그림 4〕와 같다.

DBX 파일은 크게 헤더 영역, 3개의 세그먼트, 예약된 영역으로 나눌 수 있다. 첫 번째로 파일 헤더 영역은 DBX 파일 자체의 정보와 각 세그먼트의 시작 오프셋, 세그먼트 크기 등의 정보를 담고 있다. 두 번째 영역은 인덱스 정보 세그먼트이다. 이 영역은 folders.dbx의 정보와 메일 제목, 시간, 발신자 정보, 수신자 정보 메시지 ID 등 메시지의 정보를 저장한다. 세 번째 영역인 트리 세그먼트는 다수의 이메일 관리를 하기 위해 사용하는 정보를 저장한다. 네 번째 영역인 메시지 세그먼트는 메시지를 저장한다. 메시지 세그먼트는 삭제되지 않은 이메일의 본문(text)과 삭



〔그림 4〕 DBX 파일 구조

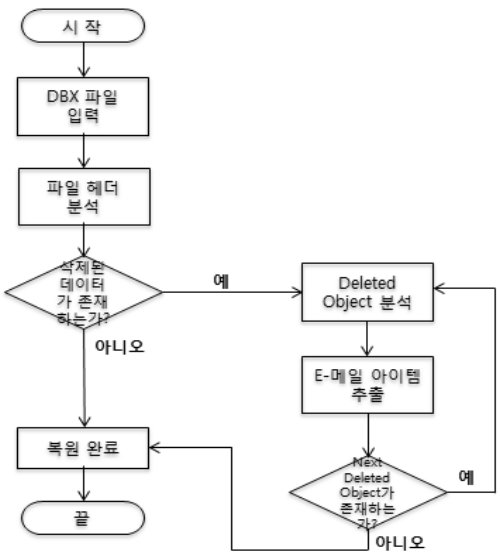
제된 이메일의 본문을 각각의 리스트로 관리한다. 삭제된 이메일 리스트는 〔그림 4〕와 같이 Deleted Object로 구성된다. 마지막으로 예약된 영역은 3개의 세그먼트 중 새로운 영역이 필요한 경우 파일의 크기를 다시 조절할 필요 없이 사용할 수 있다.

DBX 파일 내의 삭제된 이메일을 복원하기 위해서는 Deleted Object를 중심으로 분석해야 한다. Deleted Object 내에는 해당하는 객체의 길이와 다음 삭제된 객체(Next Deleted Object)의 오프셋이 저장되어 있으며, 그 외의 영역에는 발신자, 수신자, 본문 등 삭제된 이메일의 데이터가 저장되어 있다. DBX 파일 내의 삭제된 이메일 복원 알고리즘은 〔그림 5〕와 같다.

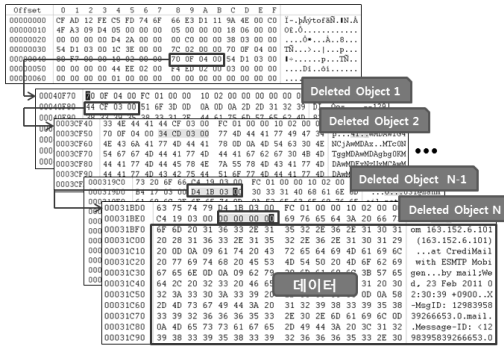
DBX 파일 내의 삭제된 이메일 리스트는 가장 최근의 삭제된 이메일이 첫 Deleted Object에 저장되므로, 최근에 삭제된 이메일 순으로 복원이 가능하다. 하지만 삭제된 메일은 본문 내용의 끝부분이 덮어써진 경우가 많아 완벽한 복원은 불가능하다. 〔그림 6〕은 실제 DBX 파일 내에서 삭제된 이메일 정보를 찾는 과정이다.

3.3 Microsoft Office Outlook (PST, OST)

Microsoft Outlook은 Microsoft Office에 속한 이메일 클라이언트로, 이메일을 저장하기 위한 파일 형식은 PST와 OST이다. PST는 Personal



〔그림 5〕 DBX 파일 내의 삭제된 이메일 복원 알고리즘



〔그림 6〕 DBX의 삭제된 이메일 본문 검색 과정

Storage Table로 이메일 메시지와 기타 항목을 저장하는 파일이다. Microsoft Outlook은 Microsoft Exchange Server와 연동하여 사용할 수 있다. OST는 Offline Storage Table로 Microsoft Exchange Server에 연결할 수 없을 때 오프라인으로 작업할 수 있도록 로컬에 복사본을 저장하여 사용하기 위한 파일이다. PST와 OST 파일 구조는 차이가 없으며, 구분하는 값이 파일 헤더에 저장되어 있다〔10〕.

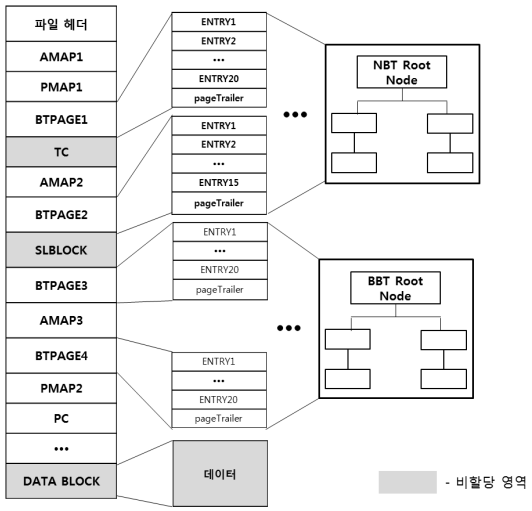
PST는 Page(512바이트) 단위로 표현하는 PMAP(Page MAP)과 Allocation(64바이트) 단위로 표현하는 AMAP(Allocation MAP)으로 파일 내의 이메일 할당 정보를 관리한다. PMAP과 AMAP의 크기는 512바이트이며, 1이면 할당, 0이면 비할당을 나타낸다. PMAP의 1비트는 페이지 크기인 512바이트와 맵핑되며, 하나의 PMAP은 2,031,616바이트의 데이터 섹션을 나타내기 때문에 2,031,616바이트 간격마다 저장되어 있다. AMAP의 1비트는 64바이트의 Allocation 블록과 맵핑되며, 하나의 AMAP은 253,952바이트의 데이터 섹션을 나타내기 때문에 253,952바이트 간격마다 존재한다〔10,13〕.

그리고 PST는 이메일을 B-Tree로 관리한다. B-Tree의 시작 오프셋은 파일 헤더에 존재하며, B-Tree는 BTPAGE로 구성된다. BTPAGE는 B-Tree를 구성하기 위한 노드 정보를 저장하며, 하나의 BTPAGE는 최대 20개의 노드 정보를 저장할 수 있다. 각 노드는 하위 BTPAGE를 가리키며, Leaf 노드는 이메일의 정보를 담고 있는 TC와 PC, SLBLOCK 그리고 Data Block을 가리킨다. TC(Table Context)는 이메일을 관리하기 위한 메타정보를 저장하며, 64바이트 배수 단위로 구성된다. PC는 이메일 본문, 제목, 발신자, 수신자 등의 이메

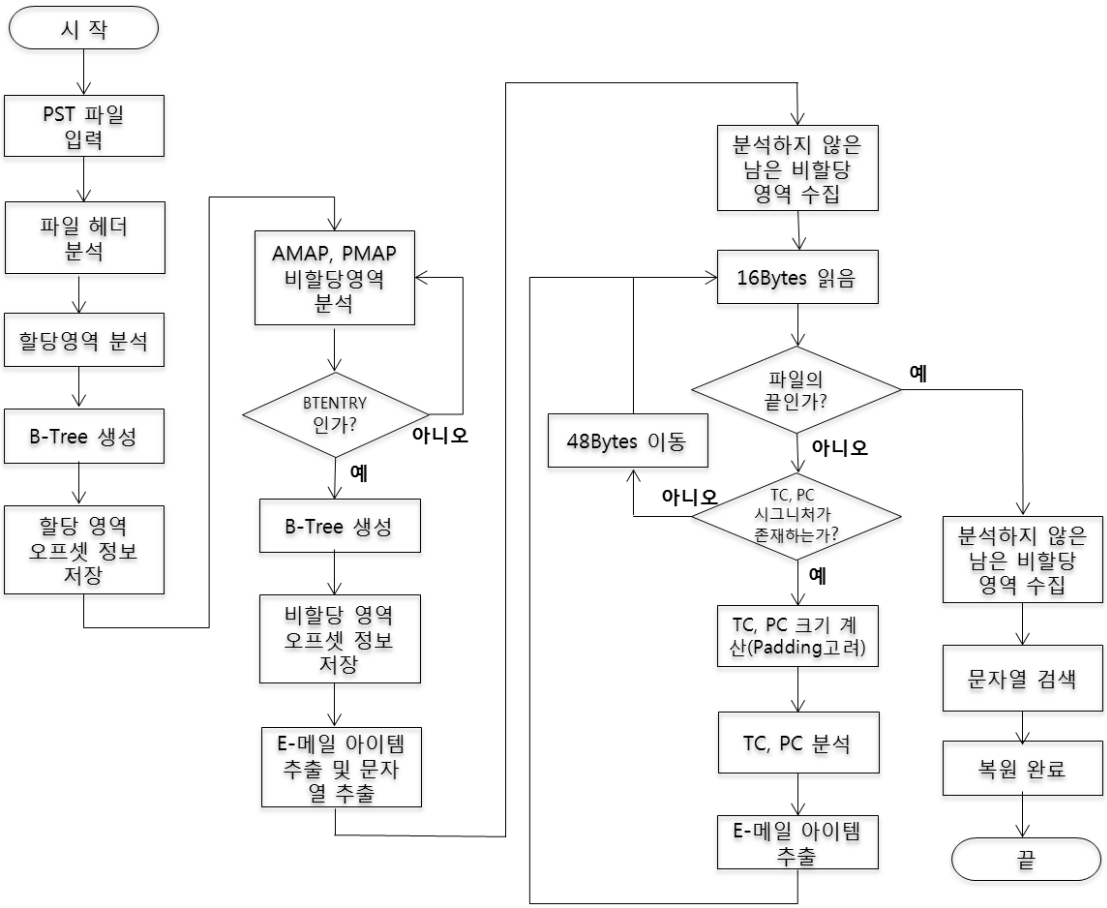
일 속성이나 정보를 저장하며, TC와 같이 64바이트 배수의 단위로 구성된다. 만약 BTPAGE에 하나의 이메일 정보를 담지 못한 경우 추가적인 노드 정보를 SLBLOCK에 저장한다. BTPAGE와 마찬가지로 TC, PC, Data Block을 가리킨다. 마지막으로 Data Block은 이메일의 첨부파일이나 이메일의 HTML과 같이 데이터가 저장된 블록이다. Data Block의 크기는 보통 한 페이지(Page) 배수로 구성된다. 전체적인 PST 구조는 〔그림 7〕과 같다.

〔그림 7〕에서 회색으로 표시된 영역은 파일의 비할당 영역을 나타낸다. 비할당 영역은 0x00값으로 비어있거나 현재는 삭제했지만 예전에 송·수신 했던 이메일 정보가 저장되어 있을 가능성이 있다.

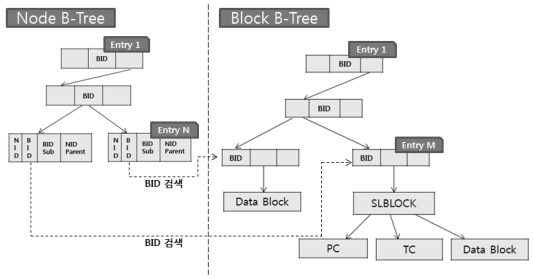
PST 파일은 이메일을 두 개의 B-Tree인 NBT(Node B-Tree)와 BBT(Block B-Tree)로 관리한다. BBT는 이메일의 실제 데이터를 저장하는 Node



〔그림 7〕 PST 파일 구조



〔그림 8〕 PST 파일 내의 삭제된 이메일 복원 알고리즘



〔그림 9〕 NBT와 BBT의 구조

를 관리하며, NBT는 효율적인 BBT의 Node 검색을 위해 인덱스 정보를 저장한다[10]. 〔그림 9〕와 같이 NBT와 BBT는 관리하는 데이터가 다르기 때문에 서로 다른 구조의 Entry를 가지고 있다.

PST에서 정상적인 메일을 추출하기 위해서는 이메일을 관리하는 B-Tree를 분석해야 한다. BTPAGE에 존재하는 Leaf 노드의 오프셋을 분석하여 각각의 TC와 PC, Data Block을 추출한 후 통합하여 하나의 이메일로 표현한다.

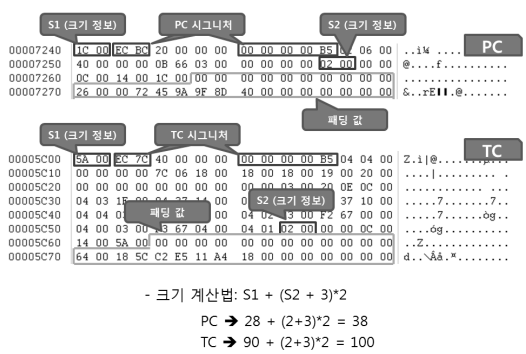
PST 내의 삭제된 이메일을 복원하기 위해서는 AMAP과 PMAP을 분석하여 삭제된 이메일이 위치한 비할당 영역을 분석한 후, 비할당 영역에 존재하는 BTPAGE를 추출하여 B-Tree를 구성한다. 만약 구성한 B-Tree의 Leaf 노드가 가리키는 데이터(TC, PC, Data Block)가 모두 비할당 영역에 존재한다면, 완전한 이메일 복원이 가능하다. 하지만 정상적인 이메일이 비할당 영역을 부분적으로 덮어쓴 경우, Leaf 노드가 가리키는 데이터는 할당 영역과 비할당 영역에 걸쳐 있을 수 있다. 이 경우는 완전한 복원이 어렵기 때문에 문자열 검색을 이용하여 비할당 영역에 존재하는 문자열을 추출한다. 그리고 비할당 영역에서 분석하지 않은 BTPAGE 외의 영역은 TC와 PC 내에 존재하는 고정적인 데이터를 시그니처로 지정하고 그것을 기준으로 다시 검색하여 TC와 PC를 추출하고 분석한다. 그리고 시그니처와 시작 오프셋이 존재하지 않는 Data Block을 추출하기 위해 지금까지 분석한 영역을 제외한 나머지 영역을 모아 카빙 기법을 적용하여 첨부파일이나 다른 정보를 추출한다. 〔그림 8〕은 본 논문에서 제시하는 PST 파일 내의 삭제된 이메일 복원 알고리즘이다.

PST는 표현할 수 있는 최대 블록 크기가 7KB이다. 그러므로 만약 삭제된 이메일 본문의 크기가 7KB 미만일 경우, 이메일의 본문이 TC나 PC에 포함되기 때문에 정상적으로 복원이 가능하다. TC와 PC는 〔그림 10〕과 같이 7바이트의 값을 시그니처로 지정하고, 크기를 계산하여, 비할당 영역에서 추출할 수 있다. 이 때, 크기는 가장 작은 블록 단위인 64바이트의 배수로 맞춰야 한다. 하지만 이메일의 본문 또는 첨부파일의 크기가 7KB 이상이거나 지운 편지함을 거치지 않은 삭제 방법으로 이메일의 메타정보가 덮어써진 경우에는 TC와 PC, Data Block에 이메일 데이터가 저장되기 때문에 분석이 어렵다. 그러므로 추출한 TC와 PC를 분석하고, 나머지 비할당 영역에서 Data Block을 분석해야 한다. 하지만 Data Block은 표현할 수 있는 시그니처가 없고 시작 오프셋도 알 수 없기 때문에 나머지 비할당 영역에서 문자열을 검색하여 읽을 수 있는 데이터를 저장한다. 이렇게 추출한 데이터는 어느 메일에 속해 있는지는 모르지만, 최소한 삭제하기 전에 송·수신한 데이터임을 확신할 수 있다. 또한 카빙 기법으로 이메일 속성 및 데이터를 효과적으로 추출할 수 있다.

IV. EMREAT(EMail REcovery & Analysis Tool) 성능 실험 및 결과

본 절에서는 앞서 설명한 정보를 바탕으로 구현한 EMREAT의 성능을 실험하기 위해 기존에 존재하는 이메일 분석도구와 EMREAT의 분석 결과를 비교한다. 성능 실험의 대상은 Microsoft Oulook의 이메일 저장 파일인 PST이며, PST의 삭제된 이메일 복원 성능에 대해 실험한다.

성능 실험 할 도구는 PST 복원 도구로 잘 알려진 PSTWalker[18]와 Kernel for Outlook PST Repair[19], Stellar Phoenix Outlook PST Repair[20]이다. 분석 대상 파일은 삭제된 이메일과



〔그림 10〕 TC, PC의 시그니처 및 크기 계산법

〔표 3〕 샘플 파일의 정보

	이메일 제목	삭제여부
1	20110124 email test No.1	O
2	20110124 email test No.2	X
3	20110124 email test No.3	X
4	20110124 email test No.4	X
5	20110124 email test No.5	X
6	20110124 email test No.6	O
7	20110124 email test No.7	O
8	20110124 email test No.8	O
9	20110124 email test No.9	O
10	COMPLETED:Sales Order…	O

정상적인 이메일의 정보를 알고 있은 샘플 파일과 실제 사용하고 있는 PST 파일을 수집하였다. 먼저 샘플 파일의 정보는 다음과 같다.

샘플 파일은 2011년 1월 19일부터 24일까지 사용한 이메일이 저장되어 있으며 크기는 4,729KB이다. 총 10개의 이메일을 저장하고 있으며, 그 중 6개의 이메일이 삭제되어 있었다. 이메일 구조를 분석한 결과, 6개의 삭제된 이메일 중 하나의 이메일이 다른 데이터로 덮어써진 것을 확인할 수 있었다. 4가지의 복원 도구를 사용하여 샘플 파일을 복원한 결과는 다음과 같다.

비할당 영역 내의 문자열 추출은 이메일 파일 내의 삭제된 이메일 내용 중에 일정 영역이 다른 정보로 덮어써져서 복원할 수 없는 경우에 수행한다. 이 경우 비록 삭제된 메일을 올바르게 복원할 수 없지만 삭제된 이메일 영역에 존재하는 문자열을 추출하여 송수신한 메일의 일부분을 복원할 수 있다.

복원 결과 EMREAT이 가장 좋은 성능을 보였다. EMREAT의 속도에서 소괄호 안의 속도(6sec)는 문자열 추출까지 실행하였을 때의 시간이며, 문자열 추출을 제외하고 비할당 영역 내의 삭제된 이메일만 복원한 속도는 소괄호 밖의 시간(2sec)이다. EMREAT에서 문자열 추출까지 실행한 결과, 238KB 크기의 비할당 영역의 문자열이 추출된 것을 확인할 수 있다.

〔표 4〕 샘플 파일의 복원 결과

도구	복원된 이메일 개수	속도	스트링 추출	복원율 (%)
EMREAT	6	2sec (6sec)	238KB	100
PSTWalker	2	2sec	-	33
Kernel for Outlook PST Repair	1	1sec	-	17
Stellar Phoenix Outlook PST Repair	5	2sec	-	83

〔그림 11〕 추출된 8번 이메일의 일부 데이터

〔표 5〕 실제 파일의 복원 결과

도구	복원된 이메일 개수	속도	스트링 추출
EMREAT	57	18sec (95sec)	23,296 KB
PSTWalker	5	4sec	-
Kernel for Outlook PST Repair	156	52sec	-
Stellar Phoenix Outlook PST Repair	52	128sec	-

EMREAT은 비할당 영역에 존재하는 이메일이 다른 정보로 일부 덮어써질 경우에도 문자열 추출을 통해 덮어지지 않은 부분에 존재하는 문자열을 추출하기 때문에 가장 좋은 성능을 보인다. 그 결과, 덮어써져 복원할 수 없는 8번 이메일의 일부 데이터를 추출할 수 있다.

다음은 실제 사용하는 PST 파일을 수집하여 성능 실험을 진행하였다. 실험 파일은 2009년 6월 12일부터 2011년 6월 13일까지 사용하였으며, 크기는 285,465KB이다. 파일의 복원 결과는 다음과 같다.

파일의 복원 결과는 Kernel for Outlook PST Repair가 156개를 복원하여 가장 많은 이메일을 복원한 것처럼 보인다. 하지만 복원된 156개의 이메일을 분석한 결과, 중복된 이메일이 많았다. 그리고 복

〔표 6〕 실제 파일의 오탐(false positive) 결과

도구	복원된 이메일 개수	오탐 개수 (중복)	실제 복원 개수	정확도 (%)
EMREAT	57	0(1)	56	98.2
PSTWalker	5	0(0)	5	100
Kernel for Outlook PST Repair	156	156 (156)	0	0
Stellar Phoenix Outlook PST Repair	52	9(5)	41	78.8

원된 이메일은 대부분 정상적인 이메일을 삭제된 것이라고 판단하였다. 실제 삭제된 이메일을 복원하지 못했다. Stellar Phoenix Outlook PST Repair는 9개의 오탐이 발견되었으며, EMREAT은 1개의 중복된 이메일이 있었다. EMREAT은 56개의 이메일을 복원하고, Stellar Phoenix Outlook PST Repair에서 41개의 이메일을 복원한 것과 달리 PSTWalker는 5개의 이메일을 복원하였다. 5개의 이메일 중 중복된 이메일이나 오탐인 경우는 존재하지 않았으나 다른 도구에 비해 복원율이 현저히 떨어지기 때문에 실제 파일의 이메일 복원은 EMREAT이 가장 좋다는 것을 알 수 있다.

복원 실험의 결과, 4가지 도구의 특징은 다음과 같다. PSTWalker는 전체적으로 복원율이 현저히 떨어진다. 또한 Kernel for Oulook PST Repair는 중복된 이메일이 너무 많이 추출되어 이메일을 분석하는데 어려움이 있다. 그리고 복원된 이메일을 확인한 결과, 삭제되지 않은 일반 이메일임에도 불구하고 삭제되었다고 분석되는 경우도 존재했다. Stellar Phoenix Outlook PST Repair는 복원율은 PSTWalker나 Kernel for Outlook PST Repair보다 좋았으나 삭제된 이메일이 표시가 되지 않아 정상적인 이메일과 구분을 할 수 없어 분석에 어려움이 있다. 이에 반해 EMREAT은 Stellar Phoenix Outlook PST Repair보다 좋은 복원율을 보였으며, 속도도 다른 도구에 비해 좋은 것을 확인할 수 있다. 또한 중복된 메일의 개수도 적었으며, 삭제된 이메일의 일부가 다른 데이터로 덮어써져 복원을 할 수 없는 경우, 문자열 추출 기능을 사용하여 남아있는 이메일 정보 중 일부를 추출할 수 있다.

V. 결론 및 향후 연구

개인 정보 및 기밀 정보와 같은 데이터의 송·수신 흔적을 지우기 위해 삭제한 이메일을 복원하는 것은 포렌식 관점에서 매우 중요하다. 본 논문에서는 다양한 이메일 파일에서 삭제된 이메일 복원이 가능함을 보였다. Thunderbird(MBOX)는 텍스트 파일이므로 파싱하여 추출하는 방법이며, Outlook Express(DBX)는 파일 포맷에 삭제된 이메일을 위한 deleted list 구조체가 존재하여 복원이 가능했으나 부분적인 복원이 가능하다. 마지막으로 Microsoft Office Outlook(PST, OST)은 복잡하지만 할당정보를 갖고 있는 AMAP, PMAP을 기반으로 정보를 수집한 후 복원할 이메일로 B-Tree를 새로 생성하며, 생성한 B-Tree에서 일반 메일 추출방법과 동일하게 삭제된 이메일 아이템을 추출할 수 있다. 또한 기존 분석보다 향상된 방법으로 비할당 영역에서 시그니처 기반의 카빙 분석과 문자열 검색으로 더 많은 이메일 정보를 획득할 수 있다.

본 논문에서는 다양한 이메일 데이터 파일에 저장되어있는 정상 메일을 추출하였으며 삭제된 메일을 복원할 뿐만 아니라, 삭제된 후 덮어써진 메일의 일부를 추출하는 기법에 대해 설명하였다. 향후에는 단순한 메일 추출이나 복원뿐만 아니라 이메일 분석에 필요한 송수신 이메일에 대한 연관 분석, 이메일 통계 분석 및 이메일 리뷰 시스템 등 다양한 분석 기법을 연구하여 효율적인 포렌식 분석을 가능하게 할 예정이다.

참고문헌

〔1〕 Microsoft, "Windows XP에서 사용할 수 있는 바로 가기 키 목록" URL: http://support.microsoft.com/kb/301583/ko, 2005

〔2〕 유병영, 박정흠, 방제완, 이상진, "비할당 영역 데이터 파편의 문서 텍스트 추출 방안에 관한 연구" *정보보호학회논문지*, Vol.20 No.6, pp. 43-51, 2010년 12월

〔3〕 PST Walker Software, PST Walker, URL:http://www.pstwalker.com/download.html

〔4〕 KERNEL DATA RECOVERY, Kernel for Outlook Express, URL : http://www.kerneldatarecovery.com/outlook-express-recovery.html

〔5〕 Campaign monitor at : http://www.campaignmonitor.com/stats/email-clients/

〔6〕 박상헌, 김역, 이상진, "HTML 파일 복원 및 웹메일 분류에 관한 연구", *KoreaCrypt 2007*, pp. 113-121, 2007년 11월

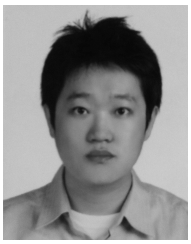
〔7〕 Rachid Hadjidj, Mourad Debbabi, Hakim Lounis, Farkhund Iqbal, and Adam Szporer, Djamel Benredjem, "Towards an integrated e-mail forensic analysis framework", *Digital Investigation 2009*, pp. 124-137, Jan. 2009

〔8〕 Farkhund Iqbal, Hamad Binsalleeh, Benjamin C.M. Fung, Mourad Debbabi,

"Mining writeprints from anonymous e-mails for forensic investigation", *Digital Investigation 2010*, pp. 1-9, Mar. 2010
〔9〕 Arne Schloh, OE dbx file format : 'file header', 'OE_Dbx_Deleted.html' URL : http://oedbx.aroh.de/zip/oedbx.zip/
〔10〕 Microsoft, 〔MS-PST〕 Outlook Personal Folders File Format (.pst) Structure Specification, Jun. 2010 URL : http://msdn.microsoft.com/en-us/library/ff385210(v=office.12).aspx
〔11〕 Lotus Notes URL : http://www-01.ibm.com/software/lotus/products/notes/
〔12〕 Importing_and_exporting_your_mail URL : http://kb.mozillazine.org/Importing_and_exporting_your_mail
〔13〕 Joachim Metz, Personal Folder File (PFF) file format specification, Jan. 2010 URL : http://sourceforge.net/projects/libpff/files/documentation/

〈著者紹介〉

정 초 롱 (Cho-rong Jeong) 학생회원
2009년 8월: 성신여자대학교 컴퓨터정보학부 졸업
2011년 8월: 고려대학교 정보경영공학전문대학원 석사
〈관심분야〉 디지털 포렌식, 이메일 포렌식



이 근 기 (Keun-gi Lee) 학생회원
2007년 2월: 부경대학교 전자컴퓨터정보통신공학부 졸업
2010년 8월: 고려대학교 정보경영공학전문대학원 석사
2010년 9월 ~ 현재: 고려대학교 정보보호학과 박사과정
〈관심분야〉 디지털 포렌식, 모바일 포렌식, 기업 포렌식

이 상 진 (SangJin Lee) 정회원
1987년: 고려대학교 학사
1989년: 고려대학교 석사
1994년: 고려대학교 박사
1989년 ~ 1999년: ETRI 연구원 역임
1992년: 국가안전기획부장 표창
1999년 ~ 현재: 고려대학교 정교수
〈관심분야〉 디지털 포렌식, 모바일 포렌식, 심층 암호, 해쉬 함수