이종 장치간 전송 파일의 추적 정보 연구

조을한*, 김지선**, 조태남***
*기전대학교 디지털포렌식학과
**우석대학교 정보보안학과
***우석대학교 IT전자융합공학과
joeulhan@gmail.com, rlawltjs122@gmail.com, tncho@ws.ac.kr

Research on tracking information of file transferred between heterogeneous devices

Eulhan Jo*, Jisun Kim**, Taenam Cho***
*Dept. of Digital Forensics and Information Security, Kijeon University
**Dept. of Information Security, Woosuk University
***Dept. of IT and Electronics Engineering, Woosuk University

요 약

파일 추적은 디지털 포렌식에서 매우 중요한 요소이며, 파일 추적에는 파일의 원본 확인과 이동 경로 분석이 수반된다. 본 논문은 다양한 매체를 통해 이미지 파일이 전송될 때 변화하는 시각정보와 원본 확인에 사용되는 해시값의 변화를 분석함으로써 파일 추적 시 고려해야 할 사항을 연구하였다.

1. 서론

디지털 포렌식에서 중요한 요소 중의 하나인 파일 추적은 기밀 파일의 배포 여부와 배포경로를 알아내는 것이다. 이를 위해서는 어떤 파일이 원본 파일로부터 배포된 것이며 배포경로를 입증할 수 있어야 한다. 파일 추적은 한 가지 기술로만 해결하기 어려우며 시스템 로그, 워터마크, 이동 매체 확인 등 많은 기술이 복합적으로 요구된다.

두 파일이 동일하다는 것은 두 파일의 해시값을 비교함으로써 확인할 수 있으며, 파일에 연관된 시간 정보는 파일의 원본과 사본을 구분 및 배포경로를 파악하는 중요한 정보가 된다[1]. 그러나 파일의 시간 정보와 해시값은 매체를 통해 다른 단말장치에 저장될 경우 매체나 단말장치에 따라 시간 정보가 달라지기도 하고[2][3] 파일이 변형되기도 한다[4]. 최근에는 PC와 이메일 뿐만 아니라 스마트폰, USB, 클라우드 등 다양한 저장 매체가 존재하며, 이들 저장 장치 간의 이동 수단도 USB, 이메일, 클라우드 저장소 등 매우 다양해졌다.

본 논문에서는 특히 아이폰의 이미지 파일이 대표적인 여러 매체를 통해 전달될 때 나타나는 시간 정보의 변화와 해시값의 변화를 분석하였다.

2. 관련 연구

(1) 파일 시간 정보

파일에는 파일시스템이 제공하는 정보와 파일 자체가 보유하고 있는 메타정보가 존재한다.

널리 사용되고 있는 FAT과 NTFS 파일시스템의 메타정보에서 파일의 생성, 수정, 접근시간 정보를 확인할 수 있으며 이 정보는 파일 자체에 포함되지 않고 파일시스템이 제공하는 정보이다. FAT과 NTFS 파일시스템에서 시간 정보를 가지고 있는 영역은 각각 표 1과 표 2 와 같다.

<표 1> FAT의 DATA 영역

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td></tr>
<tr><td></td><td colspan="7">Name</td><td colspan="3">Extension</td><td>Attr</td><td>Reserved</td><td colspan="2">Created Time</td></tr>
<tr><td>Created Date</td><td colspan="2">Last Accessed Date</td><td>Starting Cluster HI</td><td colspan="2">Last Written Date</td><td colspan="2">Last Written Date</td><td>Starting Cluster Low</td><td colspan="6">File Size</td></tr>
</table>

<표 2> NTFS 의 $STANDARD_INFORMATION

Attribute Header			
Created Time		**Modified Time**	
MFT Modified Time		**Accessed Time**	
Flag	Max Number Of Version	Version Number	Class ID
Owner ID	Security ID	Quota Charged	
Update Sequence Number (UCN)			

이미지 파일의 경우에는 EXIF(Exchangeable Image File Format)라고 하는 이미지 파일 메타데이터 포맷을 통해 사진에 대한 정보가 메타데이터로 제공되며 이 정보는 파일 일부로 포함된다.

EXIF에는 카메라 제조사(Marker), 카메라 모델(Model) 등 매우 많은 정보가 수록되어 있으며, 우리가 관심 있는 시간정보로는 생성, 수정, 접근시간 외에도 다양한 세부적인 시간 정보를 담고 있다.

(2) 해시함수

해시함수(hash function)는 데이터의 무결성을 증명하는 가장 널리 쓰이는 효율적인 방법이다. 해시함수 h(x)는 함수값 y=h(x)로부터 x를 알아낼 수 없다는 일방향성과 동일한 해시값을 가지는 서로 다른 입력값을 알아낼 수 없다는 충돌 회피성을 가진다. 따라서 두 파일이 해시값이 같으면 동일한 파일이라고 인정할 수 있으며, 주어진 해시값을 이용하여 원본 파일을 생성해 낼 수 없다. 이러한 성질에 근거하여 해시함수는 파일의 무결성과 원본 확인에 사용된다. 가장 널리 사용되는 해시함수로는 MD5, SHA1, SHA2(SHA128, SHA256, SHA512) 등이 있다.

3. 실험 환경 및 방법

이미지의 전달 매체에 따라 달라지는 파일 정보를 실험하기 위해서 그림 1과 같이 아이폰으로 촬영한 jpg 타입의 사진 파일을 여섯 가지 전송 매체를 통해 전송한 후 수신 단말기에서 해시값과 시간 정보를 확인하였다. 원본 파일을 원격 드라이브에 업로드할 때는 선택자에 따라 여러 가지 파일 타입으로 업로드하였다(표 3 참조). 안드로이드, 아이폰에 있는 jpg의 해시값은 PC에 연결하여 HashCalc 프로그램[5]을 이용하여 SHA256 값을 계산하였다.

(그림 1) 실험 방법

<표 3> 실험 환경

Source		전달 매체	확장자	Destination	
Device	확장자			Device	해시함수
iPhone Xs, iPhone 11 Pro	jpg	USB	jpg	PC, Android (Galaxy A5), iPhone	SHA256
		Gmail			
		카카오톡			
		네이버 클라우드	jpg		
			heic		
		구글 드라이브	jpg		
			heic		

4. 실험 결과

(1) 시간 정보

아이폰에 저장된 원본 사진 파일을 PC나 안드로이드로 전송했을 때, 파일시스템이 제공하는 시간정보는 모두 수신/다운로드한 시간으로 변경된다. 아이폰으로 이메일을 통해 전송했을 때도 모두 수신한 시간으로 변경된다. 그 외의 방법으로 아이폰으로 전달했을 때는 생성시간이 원본과 그대로 유지된다.

EXIF에 저장된 다양한 시간 정보도 전송 방식에 따라 삭제되기도 하였다. 그러나 삭제되지 않은 경우에는 원본과 동일한 시간정보로 유지된다. 표 4는 각각의 수신 단말에 사본이 만들어 졌을 때 시간정보가 유지되는지를 나타낸 것이다.

<표 4> 파일 전송으로 인한 시간 정보의 유지 여부

Source	전송 매체	Destination		
		PC	안드로이드	아이폰
아이폰	USB	○	-	-
	Gmail	×	×	×
	카카오톡	○	×	○
	네이버 클라우드	○	○	○
		○	○	○
	구글 드라이브	○	○	○

(2) 해시값

아이폰 단말장치 사이에서 전송 방식을 달리하여

사본을 생성하고, 원본 이미지와 사본 이미지의 해시값을 비교하였다. SHA256 결과값이 커서 동일한 값만 구분하기 위하여 알파벳으로 표기하였다.

파일을 전송 방법은 아이폰 갤러리에서 전송 프로그램으로 공유하는 방법과 전송 프로그램에 접속하여 업로드하는 방법이 존재하는데, 전송 방법에 따라 각각 다른 해시 패턴을 보이기 때문에 표 5와 같이 6개로 나누어 실험하였다. 실험 결과의 일관된 분석을 위해, 여러 개의 파일에 대하여 동일한 실험을 실행하였으며, 하나의 파일에 대해서도 3번 이상 동일한 실험을 실행하여 항상 같은 결과가 나타나는지 관찰하였다.

<표 5> 사진 파일 공유 방법

실험 구분	전송 파일	전달 방법
실험 1	원본 파일 f0	프로그램에서 업로드
실험 2		갤러리에서 업로드
실험 3	사본 파일 f1	프로그램에서 업로드
실험 4		갤러리에서 업로드
실험 5	사본 파일 f2	프로그램에서 업로드
실험 6		갤러리에서 업로드

표 6은 실험 1에 대한 결과로서, 원본 파일 f0을 전송 프로그램에 접속하여 파일을 업로드/다운로드했을 때 해시값을 비교한 결과이다. 네이버 클라우드에 heic로 업로드하여 아이폰에서 다운로드했을 때만 원본 해시값과 같고, 다른 경우에서는 해시값을 가진다.

<표 6> 전송 프로그램에서 원본 파일 업로드(실험 1)

Source		전송 방식	확장자	Destination		
단말	해시			PC	안드로이드	아이폰
아이폰	A	USB	jpg	A	-	-
		Gmail	jpg	B1	B1	B2
		카카오톡	jpg	B3	B3	B4
		네이버 클라우드	jpg	B5	B5	B5
			heic	B6	B6	A
		구글 드라이브	heic	B6	B6	B7

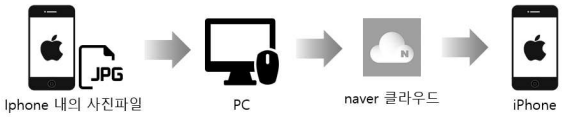
표 7은 원본 파일 f0을 아이폰 갤러리에서 공유 버튼을 이용하여 jpg 파일을 업로드했을 때의 결과이다. 여섯 가지 전송 방식 모두 jpg 확장자로 업로드되며 모두 원본과 다른 해시값을 가진다.

<표 7> 갤러리에서 원본 파일 업로드(실험 2)

Source		전송 방식	확장자	Destination		
단말	해시			PC	안드로이드	아이폰
아이폰	A	USB	jpg	A	-	-
		Gmail	jpg	C1	C1	C2
		카카오톡	jpg	C1	C1	C3
		네이버 클라우드	jpg	C1	C1	C3
		구글 드라이브	jpg	C1	C1	C1

표 6에서 본 바와 같이 2가지 경우에서 원본과 해시값이 동일하게 나타난다. 우리는 이 사본들을 이용하여 다시 실험을 수행하였다. 2번째 전송 방식에서 생성된 사본을 f1이라고 하고(그림 2), 첫 번째 전송 방식에서 생성된 사본을 가지고 2번째 전송 방식을 이용하여 생성된 파일을 f2라고 하자(그림 3).

(그림 2) 사본 파일 f1 생성 방법



(그림 3) 사본 파일 f2 생성 방법

f1을 가지고 표 6과 표 7의 실험을 수행한 결과 결과는 각각 표 8 및 표 9와 같다. 표 6은 표 8과 동일하고 표 7은 표 9와 동일한 것으로 나타난다. 즉, 예상대로 원본을 가지고 실험했을 때와 동일한 결과를 보인다.

<표 8> 전송 프로그램에서 f1 업로드(실험 3)

Source		전송 방식	확장자	Destination		
단말	해시			PC	안드로이드	아이폰
아이폰	A	USB	jpg	A	-	-
		Gmail	jpg	B1	B1	B2
		카카오톡	jpg	B3	B3	B4
		네이버 클라우드	jpg	B5	B5	B5
			heic	B6	B6	A
		구글 드라이브	heic	B6	B6	B7

<표 9> 갤러리에서 f1 업로드(실험 4)

Source		전송 방식	확장자	Destination		
단말	해시			PC	안드로이드	아이폰
아이폰	A	USB	jpg	A	-	-
		Gmail	jpg	C1	C1	C2
		카카오톡	jpg	C1	C1	C3
		네이버 클라우드	jpg	C1	C1	C3
		구글 드라이브	jpg	C1	C1	C1

f2를 가지고 표 6과 표 7의 실험을 수행한 결과 결과는 각각 표 10 및 표 11과 같다. 이 경우에는 예상과 달리 f2가 원본과 해시값이 동일한 사본임에도 불구하고 실험 결과가 원본에 대한 실험 결과와 전혀 다른 결과를 보이고 있다.

<표 10> 전송 프로그램에서 f2 업로드(실험 5)

Source		전송 방식	확장자	Destination		
단말	해시			PC	안드로이드	아이폰
아이폰	A	USB	jpg	A	-	-
		Gmail	jpg	D1	D1	D2
		카카오톡	jpg	A	A	D3
		네이버 클라우드	jpg	D4	D4	D4
			heic	A	A	A
		구글 드라이브	heic	A	A	A

<표 11> 갤러리에서 f2 업로드(실험 6)

Source		전송 방식	확장자	Destination		
단말	해시			PC	안드로이드	아이폰
아이폰	A	USB	jpg	A	-	-
		Gmail	jpg	A	A	B3
		카카오톡	jpg	A	A	B4
		네이버 클라우드	jpg	A	A	A
		구글 드라이브	jpg	A	A	A

5. 결론

본 논문에서는 아이폰에 저장된 이미지 파일이 다양한 매체를 통하여 다른 장치로 전송되었을 때 발생 되는 시간 정보의 변화와 해시값의 변화를 조사하였다. 저장 단말장치나 전송 매체에 따라 시간 정보가 달라지고, 파일의 해시값도 달라지는 것을 확인하였다. 해시값은 원본과 동일함을 확인하는 중요한 수단으로서, 해시값이 다를 경우 동일한 파일이라고 단정하기 어렵다. 원본에 아무 수정이 가해지지 않은 사본에 대해서 해시값이 달라지기 때문에 단순히 해시값의 비교로서 원본을 확인하는 것은 위험한 일이다.

향후에는 다양한 이미지 타입의 파일과 동영상 및 다양한 전송 방식에 대해 분석할 것이다. 또한 해시값의 변화를 야기시키는 원인에 대해 조사하고, 원본 확인을 위한 방법을 연구하고자 한다.

Acknowledgement

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2017R1D1A3B03032637).

참고문헌

[1] J. Kim, Mo. Kwak, S. Lee, and T. Cho, "File Tracking Technique with Active Directory Event Log," World IT Congress, paper no. 23, 2020.

[2] K. Jin, J. Yang, S. Lee, S. Han, T. Cho, "A Study on the File Trace using File Metadata," Journal of KICS, Republic of Korea, 2018. pp.873-874.

[3] J. Bang, B. Yoo, S. Lee, "Timestamp Analysis of Windows File Systems by File Manipulation Operations," Journal of KIISC, 20(3), pp.79-91, 2010.

[4] S. Han, Practice on Digital Forensics, BOOKK, 2019.

[5] HashCalc, https://www.slavasoft.com/hashcalc/.

[6] naver Cloud, www.cloud.naver.com.

[7] Google Drive, www.google.com/drive/.