디지털 포렌식 관점에서 CD-R 미디어의 해쉬 값 생성에 관한 연구

*박정흠 *김권엽 *이상진 *임종인

고려대학교 정보경영공학전문대학원

*{junghmi, kkyoup, sangjin, jilim}@korea.ac.kr

Study on Hashing of CD-R Media from the Viewpoint of Digital Forensics

*JungHeum Park *KwonYoup Kim *SangJin Lee *JongIn Lim

Graduate School of Information Management & Security, Korea Univ.

요약

해쉬 알고리즘은 디지털 포렌식 수사에서 디지털 증거의 무결성을 증명하기 위해 널리 사용되고 있다. 디지털 증거의 무결성은 동일한 데이터에서만 같은 해쉬 값이 계산된다는 성질에 의하여 증명된다. 일반적으로 동일한 데이터에 대한 해쉬 값은 서로 다른 포렌식 툴을 이용해서 계산을 해도 항상 같은 값이 출력될 것이라고 인식하고 있다. 하지만, CD-R 미디어의 경우에는 해쉬를 계산하는 포렌식 툴에 따라 값이 다르다는 특성이 있다. 이것은 해쉬 값이 CD 제작 도구에서 CD-R 미디어에 데이터를 기록하는 방식과 각 포렌식 툴 별로 CD-R 미디어를 인식하는 방식에 의해 영향을 받기 때문이다. 이러한 특성은 CD-R 미디어의 무결성 증명 시에 문제가 될 여지가 있기 때문에 디지털 포렌식 수사 절차에서 반드시 고려되어야 한다.

본 논문에서는 CD-R 미디어의 해쉬 값에 영향을 주는 요소에 대해 기술하고, 실험용 CD-R 미디어를 제작하여 대표적인 디지털 포렌식 도구들을 이용해서 확인한다. 이를 통해, 디지털 포렌식 수사 절차에서 CD-R 미디어에 대한 해쉬 값을 계산할 때 고려해야 할 사항을 제안한다.

1. 서론

해쉬 알고리즘은 가변 길이의 입력에 대해서 고정된 길이의 출력을 만들고, 일방향성이 있기 때문에 데이터의 무결성을 증명하는데 사용될 수 있다.[1] 디지털 포렌식 수사 절차에서 디지털 증거를 수집할 때 디지털 증거의 무결성을 증명하기 위한 수단이 반드시 필요하며, 이를 위해서 해쉬 알고리즘이 주로 사용된다. 디지털 증거의 무결성은 동일한 데이터에서만 같은 해쉬 값이 계산된다는 성질에 의하여 증명된다.[2]

디지털 포렌식 수사에서 하드디스크, 메모리, CD, DVD 등의 저장매체는 주요 수사 대상이며, 이에 대한 이미지 획득 과정에서 무결성 증명을 위해 해쉬 값 계산이 요구된다.[2] 이러한 요구에 맞추어 대부분의 디지털 포렌식 툴들은 저장 매체에 대한 이미지 획득 시에 해쉬 값을 저장하여 추후에 손쉽게 무결성을 검증할 수 있는 기능을 제공한다.

일반적으로 동일한 데이터에 대한 해쉬 값은 서로 다른 포렌식 툴을 이용해서 계산을 해도 항상 같은 값이 출력될 것이라고 인식하고 있다. 하지만, Chris Marberry et al.[3]에 따르면 CD-R 미디어의 경우에는 해쉬를 계산하는 포렌식 툴에 따라 값이 다르다는 특성이 있다. 이것은 해쉬 값이 CD 제작 도구에서 CD-R 미디어에 데이터를 기록하는 방식과 해쉬를 계산하는 포렌식 툴 별로 CD-R 미디어를 인식하는 방식에 의해 영향을 받기 때문이다. 이러한 특성은 CD-R 미디어의 무결성 증명 시에 문제가 될 여지가 있기 때문에 디지털 포렌식 수사 절차에서 반드시 고려되어야 한다. 최근 경찰청에서 발행한 디지털 증거처리 표준 가이드라인에서는 증거 분석의 기본 원칙으로 '증거 분석 결과의 신뢰성 확보'를 준수하도록 하고 있다. 증거 분석 결과의 신뢰성 확보란 같은 증거물에 대한 분석 결과가 항상 같아야 한다는 것이다. 다시 말해서 증거물이 동일한 경우에는 제 3의 분석관이 다시 분석해도 원래의 분석과 일치하는 결과가 도출되어야 하며, 다른 증거 분석 소프트웨어 및 장비를 사용하여도 원래의 분석과 일치하는 결과가 도출되어야만 증거 분석 결과에 신뢰성이 있다고 명시하고 있다.[4] 이 가이드라인 상으로는 해쉬를 계산하는 포렌식 툴 별로 CD-R 미디어의 해쉬 값이 다르다면 '증거 분석 결과의 신뢰성 확보'를 하지 못하는 것이다. 이렇게 되면 추후에 법정에서 디지털 증거의 신뢰성을 확보하지 못할 것이고, 최종적으로 증거로써 인정받지 못할 것이다.

본 논문에서는 해쉬를 계산하는 포렌식 별로 CD-R 미디어의 해쉬 값이 다를 수 있다는 사실을 설명하고, 디지털 증거의 무결성을 유지하기 위해 디지털 포렌식 수사 절차에서 고려해야 할 사항을 제안한다. 2장에서는 CD-R에 대한 전반적인 내용과 논문에서 사용하는 용어들을 정리한다. 3장에서 CD-R 미디어의 해쉬 값에 영향을 주는 요소들에 대해서 기술하고, 4장에서 실험용 CD-R 미디어를 제작하여 3장의 내용을 확인한다. 5장에서는 3, 4장 내용을 바탕으로 디지털 포렌식 수사 절차에서 CD-R 미디어의 해쉬 값을 계산할 때 고려해야할 사항을 제안하고, 마지막으로 6장에서 결론을 내린다.

* 본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장동력핵심기술개발사업의 일환으로 수행하였음. [2007-S019-01,정보투명성 보장형 디지털 포렌식 시스템 개발]

2. 관련 기술

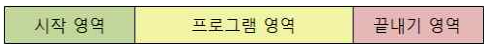
가. CD-Format Book

CD-Format Book은 CD에 대한 포맷을 정리한 문서이며, 1982년 Red Book을 시작으로 기존 포맷이 확장되거나 새로운 포맷이 만들어질 때마다 발표되고 있다. 본 논문에서는 CD에 대한 기본적인 배경 지식으로 Red Book, Yellow Book, Orange Book에 대해서 간략하게 설명한다.

(1) Red Book

오디오 CD에 대한 포맷으로, 아래 [그림 1]과 같이 시작 영역, 프로그램 영역, 끝내기 영역으로 구성된다.[5]

- 시작 영역 (Lead in area) : CD의 시작을 알리는 곳으로서 TOC (Table of contents : 목록에 대한 정보)가 담겨있다. (TOC : CD에 들어갈 데이터의 내용을 미리 알려주는 것으로, 데이터 영역의 전체길이, 데이터의 시작 위치, 이름, 트랙 수 등의 정보가 포함됨)
- 프로그램 영역 (Program area) : 실제로 음악이 저장된 오디오 트랙이 저장된다. (최대 99개의 트랙이 저장될 수 있음)
- 끝내기 영역 (Lead out area) : 프로그램 영역의 끝을 나타내는 부분이다. 실재로 이 영역에는 1분 30초 동안 계속 "0"만을 저장해서, CD 플레이어로 하여금 마지막이라는 것을 알 수 있도록 한다.



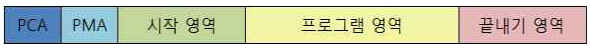
[그림 1] Red Book CD 포맷

(2) Yellow Book

Red Book 포맷에 컴퓨터 데이터를 저장할 수 있도록 확장하여 만든 포맷으로, 이 포맷의 발표와 함께 CD-ROM이 출시되었다. 이 포맷은 나중에 약간의 수정을 거쳐서 매킨토시, MS-DOS, 유닉스 등에서 사용할 수 있는 국제적인 표준으로 인정받아 CD-R과 CD-RW에 적용되는 국제 표준안인 ISO 9660으로 채택되었다.[5]

(3) Orange Book

Red Book과 Yellow Book의 규격을 벗어나지 않으면서, 트랙 단위 쓰기(Track At Once), 디스크 단위 쓰기(Disk At Once), 멀티 세션(Multi-Session) 등의 물리적인 기록 방법에 관한 규정이 포함되었다. Red Book이나 Yellow Book과 동일한 구조에 [그림 2]처럼 PCA(파워 조절영역)와 PMA(임시 저장 메모리)가 시작 영역 앞에 추가된 구조가 되었다.[5]



[그림 2] Orange Book CD 포맷

나. CD 쓰기 옵션

(1) 쓰기 방식

- 트랙 단위 쓰기 (Track at Once, 이하 TAO)

: 디스크 한 장을 트랙 단위로 굽는 방식이다. 가장 기본적인 TAO 방식은 오디오 CD에서 사용하는 Red Book 규격으로 [Lead In] [Audio Track1][Audio Track2]..... [Lead Out] 의 형태로 저장된다. 보통 TAO 방식은 Audio CD 제작 시에만 사용되며, 레이저 빔이 한 트랙별로 On/Off 를 반복하는 방식이다. 여기에서 레이저 빔이 Off된 구간이 오디오 갭(Gap)으로 2초간의 공백이 생긴다.[6]

- 디스크 단위 쓰기 (Disk at Once, 이하 DAO)

: 디스크 한 장을 한 번에 굽는 방식이다. TAO가 먼저 사용되었으며, TAO에서 발생하는 몇 가지 문제점을 해결하기 위하여 개발된 방식이 DAO 이다. [Lead In] [All Audio Track] [Lead Out]의 간단한 구조로 저장되며, 직접 복사를 할 때 일반적으로 사용하는 방법으로 레이저를 중단하지 않고 처음부터 끝까지 연속으로 쓴다. 트랙 간 갭(Gap)을 만들지 않기 때문에 오디오 CD를 만들 때 이 방식을 이용하여 TAO로 제작 시에 생성되는 2초간의 공백이 생기지 않게 할 수 있다.[6]

(2) 멀티 세션

[그림 2]의 구조가 하나의 세션(Session)이 된다. 일반적인 CD는 하나의 세션이 존재하도록 설정되며, 목적에 따라 여러 개의 세션이 존재하도록 설정할 수 있다.

- multi-session

: 멀티 세션은 한 장의 CD에 여러 개의 세션을 만드는 것이다. 대표적으로 [그림 3]의 멀티 볼륨 멀티 세션 (Multi Volume Multi Session) 방식과 [그림 4]의 연결된 멀티 세션 (Linked Multi Session) 방식이 있다. [그림 3]은 최초에 1번 Session이 있었고, 나중에 2번 Session이 멀티 볼륨 멀티 세션 방식으로 써진 것이다. 이 경우에 CD에는 2번 Session만 남은 것처럼 보이고, 일반적인 방식으로는 1번 Session에 접근할 수 없게 된다. [그림 4]는 최초에 1번 Session이 있었고, 나중에 2번 Session이 연결된 멀티 세션 방식으로 써진 것이다. 이 경우에 CD에는 1번 Session과 2번 Session이 같이 써진 것처럼 보인다. 즉, 현재 세션이 이전 세션의 데이터에 접근할 수 있게 되어 하나의 세션이 있는 것처럼 보이는 것이다.[6]

[그림 3] 멀티 볼륨 멀티 세션 (Multi Volume Multi Session)

[그림 4] 연결된 멀티 세션 (Linked Multi Session)

(3) 파일 시스템

- ISO 9660

: CD에서 사용하는 파일 시스템 표준으로 Yellow Book의 포맷이 기반이 되어 제정되었다.[7]

- ISO 9660 + Joliet

: ISO 9660 파일 시스템이 가지고 있던 파일 이름 길이의 제약(8 + 3 bytes)을 극복하기 위한 확장 파일 시스템이다.[8]

3. CD-R 해쉬 값에 영향을 주는 요소

Chris Marberry et al.[3]에 따르면 CD-R 미디어의 경우에 해쉬를 계산하는 포렌식 툴 별로 해쉬 값이 다르다는 특성이 있다. Chris Marberry et al.[3]의 실험 결과로 볼 때, CD-R 미디어에 TAO 방식으로 데이터가 써진 경우에 포렌식 툴 별로 다른 해쉬 값이 계산되는 경우가 있었다. 멀티 세션의 여부, 파일 시스템의 차이는 해쉬 값에 영향을 주지 않는 것으로 나타났다. 또한, 광학드라이브(Optical Disk Drive)의 드라이버에 이상이 있을 경우에도 각 어플리케이션 별로 다른 해쉬 값이 계산되었다.

이러한 결과는 데이터 쓰기 방식에 따라 포렌식 툴에서 CD-R 미디어를 인식하는 방식이 다르기 때문에 서로 다른 해쉬 값이 계산된다고 가정할 수 있게 한다. 이것이 사실이라면, CFTT (Computer Forensic Tool Testing) 를 통과한 상용 포렌식 툴들이 서로 다른 해쉬 값을 계산하는 것은 문제가 될 여지가 충분하다.[9] 그리고 데이터 쓰기 방식과는 별개로 광학드라이브(Optical Disk Drive)의 펌웨어(Firmware) 별로 에러 처리 방식이 다르기 때문에 서로 다른 해쉬 값이 계산된다는 주장도 있다.[10]

위의 내용을 정리하면, 다음 세 가지 요소가 CD-R 미디어의 해쉬 값이 포렌식 툴 별로 다르게 계산되는 요소라고 가정할 수 있다.

- 데이터 쓰기 방식
- 포렌식 툴에서 CD-R 미디어를 인식하는 방식
- 광학드라이브의 펌웨어

다음 장에서 이와 같은 요소들이 해쉬 값에 영향을 주는지 확인하는 실험을 한다.

4. CD-R 해쉬 값 계산 실험

이 장에서는 앞 장에서 제시한 요소들을 확인하기 위한 실험을 한다. 실험을 위한 CD-R 미디어를 [표 1]과 같이 제작하였다.

[표 1] 실험용 CD-R 미디어 제작

CD #1	CD #2
64 KB 실행 파일 DAO (Disk At Once) 하나의 세션 ISO 9660 + Joilet	64 KB 실행 파일 TAO (Track At Once) 하나의 세션 ISO 9660 + Joilet

1, 2번 CD의 메타 데이터를 확인한 결과 섹터 수는 600개이다.

실험은 EnCase, FTK, X-Ways, Final Forensics 등 대표적인 상용 포렌식 툴들을 대상으로 하였다.

먼저 1번 CD에 대한 실험 결과는 [표 2]에서 확인할 수 있다. DAO 방식으로 써진 1번 CD는 FTK, X-Ways, Final Forensics의 경우에 정상적인 섹터 수(600개)로 인식되었고, 같은 해쉬 값이 계산되었다. 그런데, EnCase의 경우 특이한 결과를 나타내었다. EnCase 4.20은 FTK, X-Ways, Final Forensics 보다 섹터 수가 하나 적게 인식되었고, 이것은 다른 해쉬 값이 계산되는 결과를 나타내었다. 한편, EnCase 6.5.1.2는 섹터 수를 11,999개로 인식하는 이상 현상을 보였다. 해쉬 계산 역시, 713개의 Read Error를 출력하며 다른 툴들과는 전혀 다른 값을 출력하였다.

[표 2] CD #1 실험 결과

포렌식 툴	해쉬 값	섹터 수
EnCase 4.20	B3C04AD40F7D59A4DE1DA9C68CE25656	599
EnCase 6.5.1.2	D0C0C0310F979B8A36ACCBEEE4A624B7 (Read errors: 713)	11,999
FTK Imager 2.4	e9f530cd156483106839a5f1ea64aeac	600
X-Ways 14.5	E9F530CD156483106839A5F1EA64AEAC	600
Final Forensics	E9F530CD156483106839A5F1EA64AEAC	600

2번 CD에 대한 실험 결과는 [표 3]에서 확인할 수 있다. TAO 방식으로 써진 2번 CD는 1번 CD의 경우와 마찬가지로 FTK, X-Ways, Final Forensics의 경우에 정상적인 섹터 수(600개)로 인식했다. 하지만, X-Ways의 경우 2개의 Read Error를 출력하며, FTK, Final Forensics와 해쉬 값을 다르게 계산하였다. FTK와 Final Forensic은 1번 실험과 마찬가지로 같은 섹터 수로 인식되었고, 같은 해쉬 값이 계산되었다. 이번에도 역시, EnCase의 경우 특이한 결과를 나타내었다. EnCase 4.20은 1번 실험과 마찬가지로 FTK, X-Ways, Final Forensics 보다 섹터 수가 하나 적게 인식되었고, 더불어서 1개의 Read Error를 출력하며 다른 해쉬 값이 계산되는 결과를 나타내었다. EnCase 6.5.1.2는 1번 실험과 마찬가지로 섹터 수를 11,999개로 인식하는 이상 현상을 보였다. 해쉬 계산 역시, 713개의 Read Error를 출력하며 다른 툴들과는 전혀 다른 값을 출력하였다.

[표 3] CD #2 실험 결과

포렌식 툴	해쉬 값	섹터 수
EnCase 4.20	195B8D4A12FB03CFC6448FFFDCBB6C32 (Read Errors: 1)	599
EnCase 6.5.1.2	E8A2CDB50A3C249A9FD6C914C121EB11 (Read errors: 713)	11,999
FTK Imager 2.4	1579b31d9c0c537f9bc9c72039ca3a36	600
X-Ways 14.5	F10F9ACBE286CE886280B7377FBE1D6D (Read errors: 2)	600
Final Forensics	1579B31D9C0C537F9BC9C72039CA3A36	600

[표 4]에서 위의 실험 결과를 정리하였다.

[표 4] CD #1, #2 실험 결과 정리

CD #1 (DAO)	EnCase를 제외한 다른 툴들은 모두 같은 섹터 수를 인식하고, 같은 해쉬 값을 계산함. EnCase 4.20는 정상적인 섹터 수 -1로 인식. EnCase 6.5.1.2는 섹터 수를 비정상적으로 많게 인식.
CD #2 (TAO)	EnCase를 제외한 다른 툴들은 같은 섹터 수를 인식. FTK, Final Forensics는 같은 해쉬 값 계산. X-Ways는 2개의 Read Error 출력. EnCase 4.20는 정상적인 섹터 수 -1로 인식. EnCase 6.5.1.2는 섹터 수를 비정상적으로 많게 인식.

다음으로 광학드라이브의 펌웨어의 차이가 해쉬 값에 영향을 주는지를 실험하였다. 실험을 위해서 서로 다른 제조사의 광학드라이브가 설치된 PC 3대를 준비하였다. 각 PC에서 1, 2번 CD에 대한 해쉬 값을 계산해 본 결과는 [표 5]와 같다.

[표 5] 광학드라이브 펌웨어 별 실험 결과

	CD #1	CD #2
PC #1	앞의 실험 결과와 같음.	앞의 실험 결과와 같음.
PC #2	EnCase 6.5.1.2는 섹터 수를 590개로 인식. 다른 툴은 PC #1과 같음.	EnCase 6.5.1.2는 섹터 수를 590개로 인식. X-Ways는 섹터 수를 600개로 인식하지만, 20개의 Read Error 출력. 다른 툴은 PC #1과 같음.
PC #3	PC #1의 결과와 같음.	PC #1의 결과와 같음.

실험 결과 2번 PC의 경우에 1, 3번 PC와 비교하였을 때, EnCase 6.5.1.2는 섹터 수를 다르게 인식하였고, X-Ways는 많은 Read Error를 출력하였다.

앞의 실험 결과로 미루어 보아, 데이터 쓰기 방식, 포렌식 툴에서 CD-R 이미지를 인식하는 방식, 광학드라이브의 펌웨어가 CD-R 미디어의 해쉬 값에 영향을 주는 요소라고 할 수 있다. 그러나 실험에 고려한 요소가 적었고 실험 횟수가 충분하지 못했기 때문에 보다 확실히 하려면 추가적인 실험이 있어야 할 것이다.

5. 제안 사항

앞 장에서 확인하였듯이, CD-R 미디어의 경우에 모든 포렌식 툴에서 같은 해쉬 값이 계산되지 않는다는 특성이 있다. 이러한 특성은 디지털 증거의 무결성이 보존되지 못하게 할 수 있으며, 추후에 법정에서 문제가 될 소지가 있다. 그러므로 CD-R 미디어를 증거로 수집할 때 어떤 PC 환경에서 어떤 포렌식 툴을 사용해서 해쉬 값을 계산하였는지 명시하도록 해야 한다. 또한, 증거의 무결성 검증도 수집할 때의 환경에 맞추어 실시해야만 문제가 발생하지 않을 것이다.

6. 결론 및 향후 계획

이상에서 CD-R 미디어의 해쉬 값이 계산하는 포렌식 툴 별로 다를 수 있다는 사실을 설명하였다. 해쉬 값이 다른 요소로는 데이터 쓰기 방식, 포렌식 툴에서 CD-R 미디어를 인식하는 방식, 광학드라이브의 펌웨어가 있으며, 이러한 요소들을 실험을 통해 확인하였다. 하지만, 실험 횟수가 충분하지 못하다는 한계가 존재한다. 또한, 다른 쓰기 옵션, 다른 CD-R 미디어, CD 제작 환경 등의 고려할 수 있는 요소들을 모두 실험하지 못했다는 한계가 있다. 고려할 수 있는 모든 요소들을 실험해야 정확한 결과가 나오겠지만, 같은 데이터에 대한 해쉬 값이 계산하는 포렌식 툴마다 다르다는 사실은 의미 있게 받아들여야 할 것이다. 특히, 디지털 증거가 법정에서 효력을 발휘하려면 무결성이 반드시 검증되어야만 하기 때문에 디지털 증거를 수집하는 수사관들은 증거 수집 단계에서 본 논문에서 기술한 사실들을 인지하고 있어야만 한다.

향후에는 본 논문의 실험에서 고려하지 못한 다양한 요소들 (다른 쓰기 옵션, 다른 CD-R 미디어, CD 제작 환경, 해쉬 값 계산 환경, 더 많은 포렌식 툴 등) 에 대해서 더 폭 넓은 실험을 할 계획이다.

참고 문헌

[1] Mark Stamp, “Information Security : Principles and Practice ”, WILEY, 2005

[2] Kevin Mandia, Chris Prosise & Matt Pepe, “Incident Response & Computer Forensics”, McGraw-Hill Osborne Media, 2003

[3] Chris Marberry, Philip Craiger, “CD-R Acquisition Hashes Affected by Write Options”, Journal of Digital Forensic Practice

[4] 디지털 증거처리 표준 가이드라인, 경찰청

[5] CD-Format Book, http://www.samsungodd.com/kor/Information/ODDTech, 삼성전자 정보자료실 - 기술 정보 - CD-ROM 기술

[6] CD의 레코딩 방식, http://powerodd.com/bbs/board.php?bo_table=odd_lecture&wr_id=11&page=3, ODD mania community powerODD

[7] Standard ECMA-119 - Volume and File Structure of CDROM for Information Interchange, Ecma International, 1987

[8] Joliet Specification, http://bmrc.berkeley.edu/people/chaffee/jolspec.html, Microsoft Corporation, 1995

[9] Computer Forensic Tool Testing (CFTT), http://www.cftt.nist.gov/, National Institute of Standards and Technology (NIST)

[10] Paul Crowley, Dave Kleiman, “CD and DVD Forensics”, SYNGRESS, 2006