ITL 사이버보안 연구보고서

이 보고서는 ITL 테크노트(TechNote) 사이트 있는 연구보고서이며, 문서로 허가되지 않고서는 다른 곳에 재 게시할 수 없습니다.

SQLite 삭제된 레코드 영역 복구 기법

SQLite 는 오픈소스 데이터 베이스 소프트웨어로 작고 빠른 특징 때문에, 소형 임베디드 기기, 브라우저 등에서 많이 사용하고 있다. 최근 포렌식 분석 시 모바일 기기의 분석 수요가 늘어남에 따라 본 문서에서는 이러한 SQLite 데이터 베이스 파일 포맷에서의 데이터 복구 방법에 대해 알아본다. 첫 번째로 SQLite 의 기본 구조를 간략하게 알아본 후, 구조를 통해 복구할 수 있는 비할당영역과 삭제된 셀의 데이터를 확인하는 방법에 대해 알아 본다

SQLite 삭제된 레코드 복구 기법

작성자: 박상호, ddeok9@gmail.com

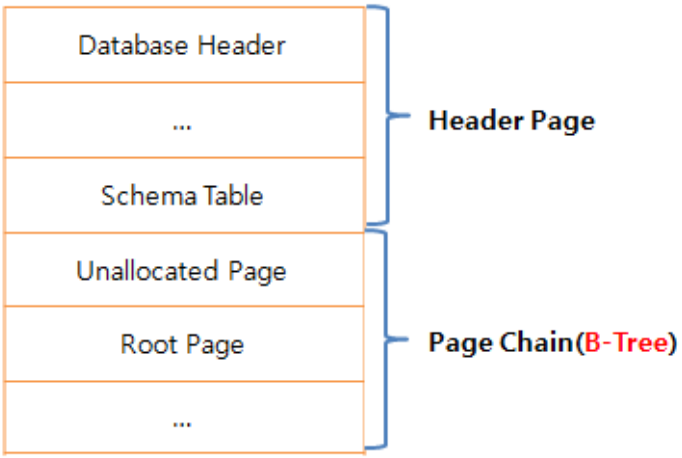
승인 일자: 2014. 7. 7

요 약

SQLite 는 오픈소스 데이터베이스 소프트웨어로 작고 빠른 특징 때문에, 소형 임베디드 기기, 브라우저에서 많이 사용되고 있다. 최근 포렌식 분석 시 PC 뿐만 아니라 모바일 기기도 늘어남에 따라 SQLite 파일이 저장하는 문자메시지, 연락처, 통화내역 등이 중요한 증거로 쓰이는 경우가 있다. 만약 사용자가 임의로 삭제한 데이터가 존재한다면, SQLite 파일의 구조 파악 후 삭제된 레코드 복구 기법을 이용하여 복구를 진행한 후 분석을 해야 한다. 본 문서에서는 간략하게 SQLite 데이터 베이스 파일 포맷에 대해 알아보고, 비할당 영역과 삭제 된 셀에서의 데이터 복구에 대해 알아본다.

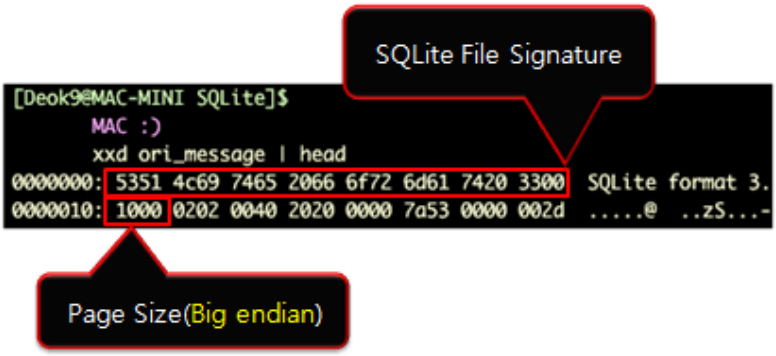
1. SQLite 파일 포맷

SQLite 데이터 베이스 파일의 전체 구조는 아래 [그림 1]과 같다.



[그림 1] SQLite 데이터베이스 파일 전체구조

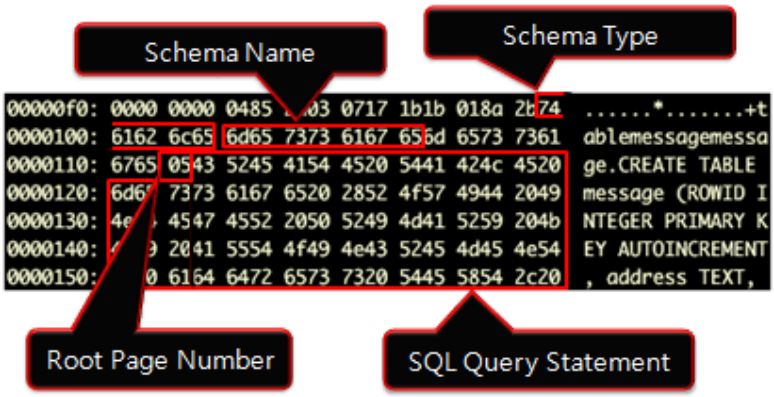
크게 헤더 페이지와 페이지 체인으로 구성되어 있으며, 헤더 페이지에는 SQLite 데이터 베이스 파일 시그니쳐와, 페이지 사이즈(빅 엔디안 표기)를 가지고 있다. 페이지 체인은 루트 페이지를 시작으로 B-Tree 형식으로 이루어져 있다. 아래 [그림 2]는 SQLite 데이터 베이스 헤더를 실제로 확인한 결과이다.



[그림 2] SQLite 데이터 베이스 파일 헤더 확인

최근 사용되고 있는 SQLite 버전은 3이 아니지만, 기존 애플리케이션과의 호환 문제로 인해 시그니쳐는 그대로 "SQLite format 3"으로 사용하고 있다. 위 예제 파일에서의 페이지 크기는 0x1000 사이즈를 가지고 있는 것을 확인 가능하다. 이는 차후 페이지를 탐색할 때 유용하게 쓰일 수 있는 데이터이다.

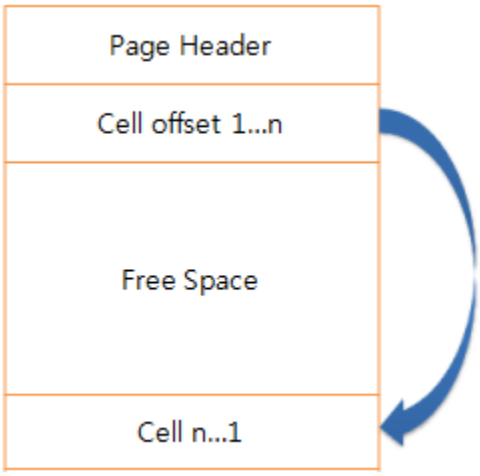
아래 [그림 3]은 스키마 테이블을 확인한 결과이다.



[그림 3] SQLite 데이터 베이스 스키마 테이블 확인

스키마 테이블은 스키마 타입(테이블, 인덱스 등), 스키마 이름, 루트 페이지 번호, SQL 쿼리 상태 등의 정보를 가지고 있다. 본 문서에서 확인 하는 스키마는 테이블 스키마 이며, 삭제된 데이터가 실제로 저장되어 있는 스키마를 의미한다. 루트 페이지 번호란 해당 DB 파일에서 가장 처음 나타나는 페이지의 인덱스를 의미하며, [그림 2]에서 확인한 페이지 크기인 0x1000과 현재 확인한 루트 페이지 번호를 곱하여 0x5000 지점으로 이동하면 실제로 가장 첫 페이지를 확인할 수 있다. SQL 쿼리는 테이블 생성 시 사용한 쿼리로 추후 데이터 복구 시 각 테이블이 가지는 데이터를 유추하기 위해 알아둘 필요가 있다.

실제 데이터들이 저장된 페이지는 아래 [그림 4]와 같은 구조를 가지고 있다.



[그림 4] 페이지 구조

크게 페이지 헤더, 셀 오프셋 정보, 비할당 공간(페이지 내 데이터 양에 따라 존재 유무가 다름), 실제 데이터가 저장되는 최소 단위인 셀으로 구성되어 있다. 페이지 헤더는 아래 <표 1>과 같은 구조를 가지고 있다.

<표 1> 페이지 헤더 구조

오프셋	내용
0	페이지 플래그: 0x05 (Internal 페이지) 페이지 플래그: 0x0D (Leaf 페이지)
1-2	첫 비 할당 블럭 오프셋
3-4	페이지 내의 셀 개수
5-6	첫 번째 셀 오프셋
7	3바이트 이상 비 할당 블럭 개수
8 -	셀 오프셋 체인

페이지는 Internal 페이지와 Leaf 페이지로 이루어져 있으며, Internal 페이지는 B-Tree 이어주는 페이지이고, Leaf 페이지는 실제 데이터가 저장되는 페이지이다. 블록과 셀은 같은 개념이나 삭제되었을 경우 블록, 활성 상태일 경우 셀이라고 표현하고 있다. 각 오프셋이 가지는 정보를 통해 추후 삭제된 데이터 복구에 어떻게 사용하는지 알아보도록 하겠다.

마지막으로 실제 데이터가 저장된 셀은 아래 그림 5와 같은 구조를 가지고 있다.

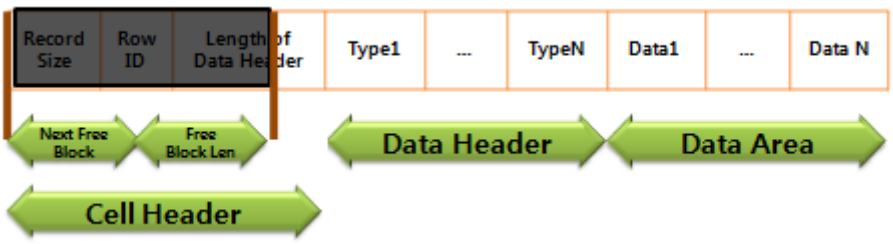
[그림 5] 셀 구조

크게 셀 헤더, 데이터 헤더, 데이터 영역으로 구분되며, 레코드 사이즈와 Row ID는 가변길이 정수로 이루어져 있다. 데이터 헤더는 데이터 영역의 실제 데이터가 어떤 타입인지를 알려주며, 바로 앞에서 데이터 헤더의 총 길이를 나타내고 있다.

지금까지 SQLite 데이터베이스에서 삭제된 데이터 복구 시 필요한 간단한 구조에 대해 살펴보았다. 다음으로 삭제된 페이지 복구 방법에 대해 알아보도록 한다.

2. 삭제된 데이터 복구

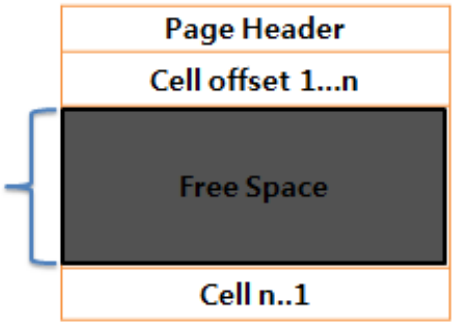
삭제된 데이터 복구하기 전에 실제 데이터가 삭제되면 데이터가 저장된 셀이 어떻게 변화하는지 먼저 알아본다. 아래 [그림 6]은 데이터 삭제 시 셀의 변화이다.



[그림 6] 데이터 삭제 시 셀의 변화

셀의 맨 앞 4바이트가 각각 2바이트씩 다음 삭제된 비할당 블록의 오프셋과, 해당 프리 블록의 길이로 변한다. 이는 기존에 가지고 있던 레코드 크기와 ID 값의 크기에 따라 데이터 헤더 길이 정보가 덮어 쓰여질 수 있기 때문에 복구 시 유의할 사항이다.

삭제된 데이터 복구는 크게 2가지 방법으로 진행할 수 있다. 첫 번째로는 비 할당 영역 복구 방법이다. 이는 데이터가 할당되었다가 사라진 비 할당영역 전체를 복구하는 것이다. 아래 [그림 7]은 페이지 구조에서 비 할당 영역을 나타낸 것이다.



[그림 7] 페이지 내의 비 할당 영역

복구 방법은 아래와 같다.

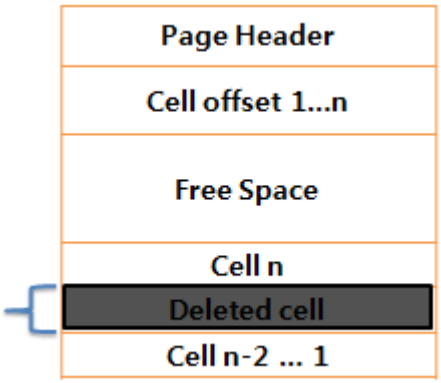
1. SQLite 파일 포맷에서 Leaf 페이지를 찾는다.
   - 데이터 베이스 헤더에서 확인한 페이지 크기 단위로 점프하며, 페이지 플래그가

0x0D인지 여부를 검사하여 스캔한다.

2. 페이지 플래그가 0x0D인 Leaf 페이지의 페이지의 오프셋 8 이후에서 셀 오프셋 체인을 따라가며 오프셋이 0x0000이 될 때까지 확인한다.
   - 비 할당 영역의 첫 부분을 확인할 수 있다.
3. 다음으로 페이지 헤더 오프셋 5-6에 있는 첫 번째 활성 셀 오프셋 정보를 확인한다.
   - 비 할당 영역의 끝 부분을 확인할 수 있다.

위와 같은 방법을 통해 실제 데이터가 있는 페이지에서 데이터가 할당되었다가 비 할당 처리 된 삭제 데이터 블록을 추출할 수 있다. 추출된 영역의 데이터는 텍스트로 확인 가능하기 때문에 추후 정규 표현식 또는 테이블 구조와 매칭을 통하여 가공이 필요하다.

다음으로 할당 셀 사이에 존재하는 삭제된 셀 복구에 대해 알아보도록 한다. 아래 [그림 8]은 간략하게 표현한 그림이다.



[그림 8] 삭제된 셀 복구

삭제된 셀은 페이지 헤더 오프셋 1-2에 존재하는 비할당 블럭 오프셋을 시작으로 삭제된 셀이 가지는 첫 4바이트(비할당 블럭 오프셋과 크기)를 통해 체인을 형성해 복구 가능하다.

위와 같은 과정을 거쳐 삭제된 데이터를 추출하면 아래 [그림 9]와 같이 복구 가능한 것을 확인할 수 있었다.

[그림 9] 삭제된 문자 메시지 복구 확인

3. 결론

지금까지 SQLite 데이터 베이스 파일의 간단한 구조와 이를 이용한 삭제된 비할당 블록, 삭제된 셀을 복구하는 과정에 대해 살펴 보았다. SQLite 데이터 베이스 파일을 복원하여 Viewer로 볼 수 있도록 하는 방법은 아니지만, 분석 시 삭제된 데이터를 명시할 수 있고, 확인할 수 있다는 점에서 필요한 기법이라고 생각한다.

"ITL Cyber Security Summer 2014"

사이버보안, 사이버포렌식의 분야에 6개의 교육과정과 사이버 보안 컨퍼런스가 서울 양재동 엘타워에서 개최됩니다. 국내 최고 전문가가 직접 강의하는 CSS 2014(8월 26일 – 29일) 참여 기회를 놓치지 마십시요!

2014년 8월 26일 ~ 29일, ITL Cyber Security Summer 2014, 서울 양재동 엘타워 5층, 8층

- **2014년 8월 26일(화)**
  - 사이버 시큐리티 써머(CSS) 컨퍼런스 2014
- **2014년 8월 27일(수) – 28일(목)**
  - M230 악성코드 수집 및 분석
  - M300 윈도 침해사고 포렌식 분석
  - M270 HTML5 보안위협 및 공격 기법
- **2014년 8월 29일(금)**
  - M210 메타스플로이트를 이용한 침투시험
  - M330 침입탐지를 위한 로그분석 방법
  - M350 난독화 기법 및 대응

"SANS Korea 2014"

정보보호, 사이버 포렌식의 세계 최고 수준의 교육 'SANS Korea 2014'가 11월 10일 - 15일까지 6일간 서울 코엑스에서 개최합니다. 코스를 놓치지 마십시요!

2014년 11월 10일 ~ 15일, SANS Korea 2014, 서울 코엑스

- **SEC504**: 해킹 기법, 공격 및 사고 대응 방법 (GIAC: GCIH)
- **FOR585**: 고급 스마트폰 및 모바일 기기 포렌식
- **SEC660**: 고급 침투시험, 공격 및 윤리적 해킹 (GIAC: GXPN)

ITL 교육, 국내외 SANS 교육 및 GIAC 자격증과 관련 사항은 아래로 문의 바랍니다.

- 전화: 031)717-1447
- 이메일: itl@itlkorea.kr, sans@sans.or.kr