Web Server에서 Web URL Page의 Image Log File

유승희, 조동섭
이화여자대학교 컴퓨터정보통신학과

Image Log Files of the URL Page of Web Server

Seung-Hee Yoo, Dong-Sub Cho
Dept. of Computer Science, Ewha Womans University

Abstract - 웹 서버에서 로그파일은 웹 서버에 대한 접속정보를 저장한다. 이 정보를 분석하면 웹 서비스를 하는데 있어서 서비스의 질을 높이는데 좋은 참고자료가 될 뿐 아니라 웹 서버에 이상이 생겼을 경우 발생한 오류를 조기에 발견하는 데에도 사용되는 중요한 자료이다. 현재 이러한 로그파일은 텍스트 파일로 저장이 되어있으며 오랜 시간이 지나 그 웹 페이지가 삭제되었을 경우 로그파일에 기록된 그 시각의 웹 페이지를 찾아보기가 어렵다.
본 연구에서는 로그파일에 기록된 그 시각의 웹 페이지의 이미지를 저장하는 방법으로 이러한 단점을 보안하고 오랜 시간이 지난 후에도 그 웹 페이지를 볼 수 있는 방법을 제안한다. 이 아이디어가 구현되어 실현되면 또한 Digital Forensic으로써 범죄 수사에도 많은 도움이 될 뿐만 아니라 휴대전화로 풀 인터넷 브라우징이 가능한 풀브라우저에도 적용될 수 있다.

1. 서 론

웹서버는 웹서비스에 대한 요청과 제공에 대한 기록을 모두 로그파일에 저장한다. 따라서 로그파일을 이용하여 사용자 정보, 방문 시간, 사용중인 웹 브라우저 종류, 방문 페이지, 다운로드 용량, 서버의 에러원인, 8단계의 에러메시지 등의 다양한 정보를 분석해 낼 수 있다.
하지만 오랜 시간이 지난 후 서버의 그 웹페이지가 삭제되면 더 이상 로그파일에 기록된 그 시간의 웹 페이지를 볼 수 없게 된다.
로그파일을 기록한 시간의 이미지 파일을 저장시켜놓으면 오랜 시간이 지나 웹페이지가 삭제되었을 경우에도 그 시간의 웹페이지를 볼 수 있다.
그리하여 본 논문에서는 많은 부분에 적용될 수 있는 Image Logging 기법을 제안해 보았다.

2. 관련 연구

2-1. Log File

사이트에서 파일을 검색할 때마다 서버 소프트웨어는 그 기록을 남긴다. 로그(Log)란 시스템에 접속한 사용자들의 행위들을 저장해 놓은 기록들이다. 따라서 외부에서 침입을 해온 공격자가 시스템에서 어떠한 일을 행했는지, 또는 사용자가 어떠한 명령어들을 사용해주었는지 등의 보안적인 의미의 정보들과 시스템이 처리한 업무와 에러 등의 시스템 운영 정보들을 가지고 있다. 즉, 시스템에서 작동된 모든 현상들이 저장되어 있는 것이 바로 로그이다. 로그는 문제가 발생하였을 때 그 해결 방안을 제시해주는 가장 기본적인 자료이다.
Apache서버의 경우, Access(Transfer), Error, Referrer, Agent등 4개의 로그파일로 웹서버의 모든 상황을 텍스트 파일 형태로 기록한다.
로그파일의 종류 및 각 파일의 기록정보는 아래 표와 같다.

<표 1> Log File의 종류

로그 파일 종류	설명
Access Log File	Transfer 로그파일이라고도 하며, 일반적인 정보를 기록한다.
Error Log File	웹서버에서 발생하는 모든 에러와 접속실패를 시간과 내용의 두 가지로 기록한다.
Refferer Log File	사용중인 웹서버를 소개해준 사이트와 소개받은 페이지를 화살표로 기록한다.
Agent Log File	사용자가 사용한 웹브라우저와 OS에 대한 정보를 기록하는 파일로, Browser Log File 이라고도 한다.

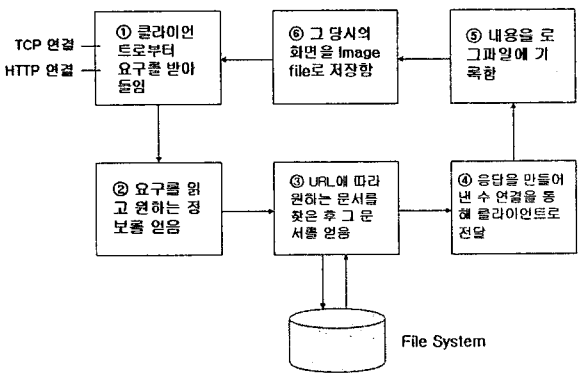
로그파일의 형식에는 CLF(Common Logfile Format), IIS(Internet information Server)와 W3C 등이 있다. CLF는 웹서버의 원조라 할 수 있는 NCSA계열의 웹서버에서 사용하는 파일형식으로 현재 대부분의 웹서버가 지원하고 있다. 물론 웹서버마다 자체적으로 로그파일의 포맷형식을 지원하고 있지만, 대부분의 웹서버 제작사는 이 CLF라는 표준로그파일형식을 지원하고 있다. IIS는 windows NT에서 가장 많이 사용되는 웹서버 소프트웨어로서 자체적으로 분석도구도 제공한다. 로그파일형식은 NCSA계열의 로그파일과는 다르며, 파일의 기록기간단위 즉 일별, 월별 등의 환경설정도 가능하다. W3C는 Extened MS-IIS가 제공하는 다른 로그파일 형식이다. NCSA 또는 IIS 로그파일형식이 고정형식인 것에 비해 이것은 사용자가 로그항목의 위치와 내용을 지정할 수 있다.

2-2 Digital Forensics

Digital Forensics란 디지털 자료가 법적 증거력을 갖기 위해 자료의 수집/보관/분석/보고에 이르는 일련의 모든 절차와 방법들을 총칭하는 말이다. 아직 명확히 규정된 표준은 없고 디스크 포렌식, 네트워크 포렌식, DB 포렌식 등 수만은 종류의 포렌식이 존재한다. 아직 국내에서는 연구 초기단계이지만 미국에서의 경우 특정 주에서는 일정 기업 이상의 규모에서 포렌식 툴을 사용하여 디지털 관련 범죄가 있을 경우, 법적 증거물로 인정이 된다.
본 연구에서는 Image Log File을 Digital Forensics로서 사용방안을 제시해보았다. 웹서버에 외부 침입이 있었을 경우 로그파일만 가지고는 충분한 법적 증거력이 되지 못한다. 예를 들어 오랜 시간이 지났을 경우 그 시간의 웹페이지가 삭제되어 더 이상 어떠한 웹페이지를 봤는지 무슨 작업을 행했는지 알 수 없기 때문이다. 하지만 웹서버에서 인터넷 브라우저의 Full Browser의 Image File은 그 당시의 모든 화면을 보여주기 때문에 충분한 법적 증거자료가 될 것이라 판단된다.

3. Web URL Page의 Image Logging 구성도

서버 구현에는 아파치 서버를 사용할 것이다. Image File을 아래 그림에서 보듯이 로그파일을 기록하고 기록한 시간의 URL Page를 Image Log DB에 저장한다. 본 Image Logging에는 Guangming Software(US)사의 HTML SnapShot이라는 소프트웨어를 사용하여 URL Page를 Image로 만들 것이다. HTML SnapShot은 Visual C++로 구현된 프로그램으로서 HTML파일을 JPG나 BMP등의 그림파일로 만들어 주는 소프트웨어이다.



<그림 1> Image Logging Web Server의 알고리즘

4. 결론 및 향후 과제

본 논문에서는 Web Sever의 Image Logging기법을 제안해 보았다. 이 기법은 다음과 같은 기대를 할 수 있다.
첫째, 시간이 지나 Web Page가 삭제된 후에도 그 시간의 Web Page를 확

인할 수 있다.

둘째, Digital Forensics으로서 법적 증거로뿐만 아니라 많은 부문에서 응용되어 사용 되어질 수 있다.

향후에는 중복되는 Image File을 알고리즘을 통하여 삭제할 수 있는 mining기법 등 구체적인 방법을 고안하고 구현을 통하여 Image Logging기법을 검증 할 필요가 있다.

[참 고 문 헌]

[1] 고광만, 박홍진 "모바일 웹 서버 관리기 구현"
[2] http://www.apache.org
[3] http://www.logger.co.kr
[4] "Web Server Monitoring", White Paper,
http://www.freshtech.com/WhitePaper.htm
[5] The Economics of Digital Forensics
[6] F. Stajano. Will your digital butlers betray you? In Proceedings of A CM Work-
shop on Privacy in the Electronic Society (WPES), 2004.
[7] http://www.guangmingsoft.net/
[8] 안정철 "시스템 로그분석"