

Web URL Page 의 Image Logging 기법

유승희, 조동섭
이화여자대학교 컴퓨터정보통신학과
e-mail : yooyyy@gmail.com

Image Logging Technique of a Web URL Page on the Tiny Web Server

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

요 약

기존의 웹서버는 텍스트형식의 로그파일을 기록한다. 웹 서버에서 로그파일은 클라이언트의 웹 서버에 대한 접속정보와 웹서버의 처리상황 등 모든 상황을 기록하고 저장한다. 이 정보를 분석하면 웹 서비스를 하는데 있어서 웹 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 저장되어있기 때문에 조작의 가능성도 있고 오랜 시간이 지나 해당 웹 페이지가 삭제되었을 경우 로그파일에 기록된 그 시각의 웹 페이지를 볼 수 없다.

본 연구에서는 로그파일에 기록된 그 시각의 웹 URL 페이지 이미지를 저장하여 이미지 로그파일을 만드는 시스템을 구현해 봄으로써 텍스트형식 로그파일의 단점을 보완하고 오랜 시간이 지난 후에도 그 웹 페이지를 볼 수 있는 기법을 연구하였다. 이 기법은 로그파일로써의 역할 뿐만 아니라 Digital Forensics 로 범죄 수사에도 많은 도움이 될 수 있고 휴대전화에서 폴 인터넷 브라우징 연구에도 적용될 수 있다.

1. 서론

웹서버는 클라이언트의 웹서비스에 대한 요청과 제 공에 대한 기록을 모두 텍스트형식의 로그파일에 저장한다. 따라서 로그파일을 이용하여 사용자 정보(IP Address), 방문 시간, 사용중인 웹 브라우저 종류, 방문 페이지, 다운로드 용량, 서버의 에러원인, 8 단계의 에러메시지 등의 다양한 정보를 분석해 낼 수 있다.

텍스트파일 형식의 로그파일은 보안상 문제점을 가지고 있고, 오랜 시간이 지난 후 서버의 그 웹페이지가 삭제되어 더 이상 데이터 베이스가 남아있지 않을 경우 더 이상 로그파일에 기록된 그 시각의 웹 페이지를 볼 수 없다.

로그파일을 기록한 시각의 웹 페이지를 이미지 파일로 저장시켜놓으면 오랜 시간이 지나 웹페이지가 삭제되었을 경우에도 그 시각의 웹페이지를 볼 수 있고, 쉽게 조작할 수 있는 텍스트파일 형식의 로그파일보다 보안상 훨씬 안전하다.

그리하여 본 논문에서는 많은 부분에 활용될 수 있는 이미지 로깅 서버 시스템을 구현하여 보았다.

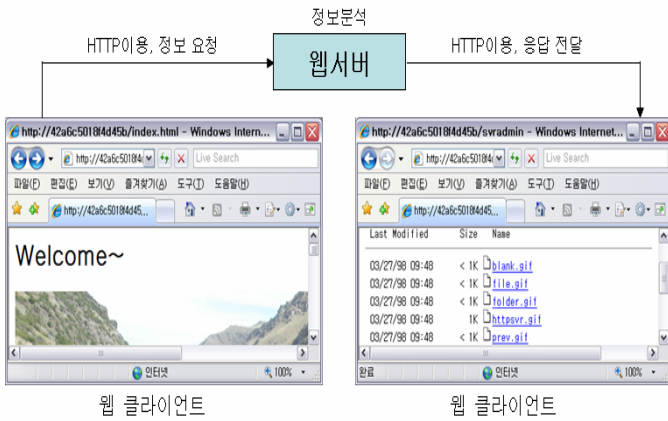
2. 관련연구

2-1. 웹서버

웹서버는 HTTP 를 통해 클라이언트가 요청하는 HTML 문서나 이미지파일 등을 전송해주는 서비스 프로그램이다. 클라이언트인 웹 브라우저의 URL 에서 나타내는 HTML 문서 및 각종 정보를 HTTP 로 송신하며 대부분 클라이언트의 웹 브라우저와의 사이에서 다수의 커넥션을 연결하여 HTML 문서 및 기타 파일들을 병렬로 전송하여 처리시간을 단축하는 서비스를 제공한다. 또한, HTML 문서 처리를 위한 프로그램으로 CGI 스크립터나 자바 서블릿등을 사용하여 웹 URL 화면에 연동되는 동적 처리를 수행한다. CGI 처리를 위해 펄(Perl), 루비(Ruby), PHP 등의 스크립트 언어가 있다.

웹서버 제품에는 아파치 웹서버, CERN 웹서버, NCSA 웹서버, IIS, 웹투비(WebtoB), HTTPSvr 등이 있다.

(그림 1)은 웹서버와 웹클라이언트가 동작하는 모습이다. PC 에서 사용하는 대표적인 웹 클라이언트인 인터넷 익스플로러는 사용자가 입력한 주소를 참조하여 접속 대기 중인 웹서버에 접속한 후, HTTP 를 이용하여 요청 메시지를 보낸다. 웹서버는 클라이언트가 보낸 데이터를 분석하고 응답메시지를 다시 보낸다. 인터넷 익스플로러는 웹서버가 보낸 데이터를 받아 화면에 표시한다.



(그림 1) 웹서버/클라이언트

2-3. 로그파일

클라이언트가 웹서버에게 정보를 요청하고 그에 대한 응답을 할 때마다 서버 소프트웨어는 로그파일로 그 기록을 남긴다.

(그림 2)는 웹서버 로그파일의 예를 보여준다. 그림에서와 같이 로그파일에는 웹서버의 처리상황, 클라이언트의 아이피주소, 접속시간, 클라이언트의 요구사항, 서버의 에러원인 등의 정보가 저장된다. 이를 통해 다양한 정보를 분석해 낼 수 있다.

Hit Log for Sunday, March 30, 2008				
304	203.255.177.166	03/30/08	20:59:44	/index.html
304	203.255.177.166	03/30/08	20:59:44	/SvrAdmin/Blank.gif
304	203.255.177.166	03/30/08	20:59:44	/SvrAdmin/Folder.gif
304	203.255.177.166	03/30/08	20:59:44	/nz.jpg
200	203.255.177.166	03/30/08	20:59:47	/svradmin
304	203.255.177.166	03/30/08	20:59:47	/SvrAdmin/Blank.gif
304	203.255.177.166	03/30/08	20:59:47	/SvrAdmin/File.gif

(그림 2) 웹서버의 로그파일

로그파일은 확장자가 .txt 인 텍스트파일 형식으로 되어있으며 파일 특성상 조작이 쉬워 보안상 문제점을 가지고 있다.

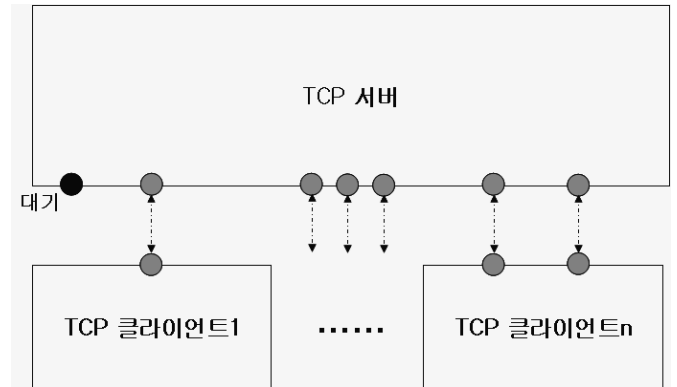
2-2. TCP 서버

TCP 서버는 TCP/IP 에 기반한 서버로 웹서버 또한 TCP 서버의 대표적인 예라 할수 있다.

서버는 소켓을 생성하고 클라이언트가 접속하기를 기다린다. 이때 서버가 사용하는 소켓은 특정 포트번호와 결합되어 있어서 이 포트번호로 접속하는 클라이언트만 수용할 수 있다. 클라이언트가 접속하면 TCP 프로토콜 수준에서 연결 설정을 위한 패킷교환이

이루어진다. 이 연결 절차가 끝나면 서버는 접속한 클라이언트와 통신할 수 있는 새로운 소켓을 생성한다. 서버가 클라이언트와 데이터를 주고받을 때는 이 소켓을 사용하고 기존의 소켓은 새로운 클라이언트 접속을 수용하는 용도로 사용된다.

그림 3 은 TCP 서버/클라이언트가 통신하는 상황을 보여준다. 그림에서와같이 서버측 소켓과 클라이언트측 소켓이 1 대 1로 대응한다.

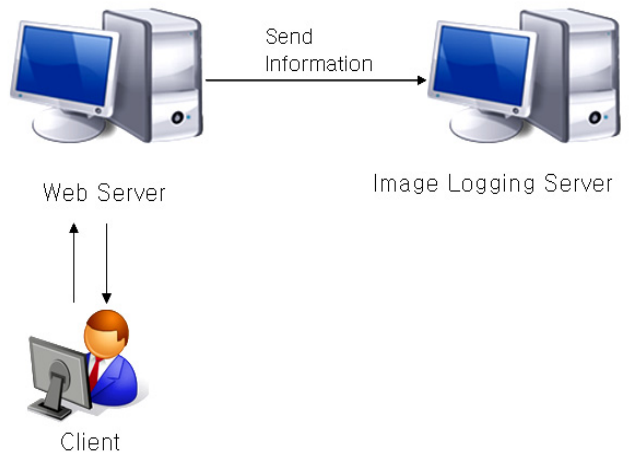


(그림 3) TCP 서버/클라이언트 통신

3. 구현

본 시스템은 웹 URL 페이지를 클라이언트에게 제공하는 웹서버와 웹서버로부터 정보를 전달받아 이미지로깅을 해주는 이미지로깅 서버로 구현하였다.

(그림 4)는 본 시스템의 동작방식을 보여준다.



(그림 4) 구현한 시스템의 동작방식

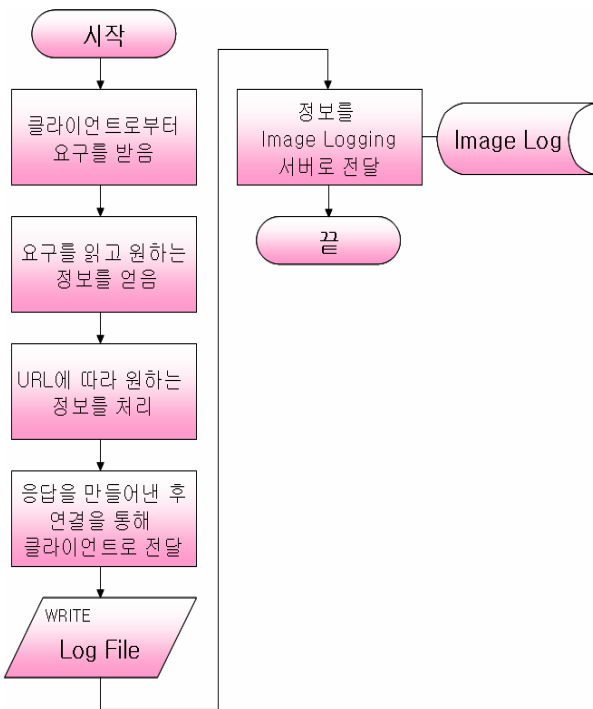
3-1. 미니 웹서버

웹서버는 Microsoft(US)사의 미니웹서버인 HTTPSVr Version 2.0 을 사용하였다. HTTPSVr 은 오픈소스 웹서버로써 사용자가 원하는대로 소스를 수정하여 사용할 수 있다.

(그림 3)은 구현된 시스템에서 미니 웹서버의 동

작 순서를 보여준다. 웹서버는 클라이언트로부터 요청이 들어오면 들어온 요청을 처리한다. 매번 요청이 들어오고 처리할 때마다 로그파일에 그 기록을 남긴다. 본 웹서버의 로그파일에는 앞절에서 본 바와 같이 클라이언트의 아이피주소, 접속시간, 요구사항, 서버의 에러원인 등의 정보가 저장된다.

구현된 웹서버는 로그파일에 이러한 정보들이 기록될 때마다 소켓을 생성하여 이미지로깅 서버로 클라이언트의 아이피주소와 요구사항(클라이언트가 방문중인 웹 URL 페이지)을 보낸다. 이때 웹서버는 이미지로깅 서버의 클라이언트가 되는 것이다.



(그림 3) 미니 웹서버의 처리 순서도

3-2. 이미지로깅 서버

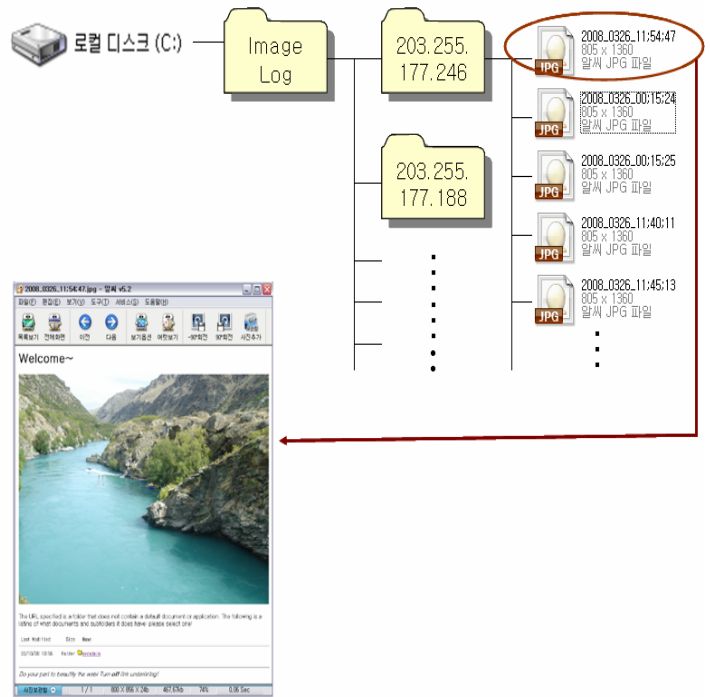
이미지로깅을 하기 위한 서버는 TCP 서버로 구현하였다. 본 서버는 웹서버에서 받은 정보를 바탕으로 이미지 로그파일을 만들어서 이미지로그 DB에 저장한다.

이미지로깅 기법에는 Guangming Software(US)사의 HTML SnapShot 이라는 소프트웨어를 사용하였다. 본 소프트웨어는 웹 URL 페이지를 이미지(JPG 파일)로 만들어주는 소프트웨어로써 Visual C++로 구현된 프로그램이다.

본 서버에서는 웹서버에서 받은 정보인 클라이언트의 아이피주소와 요구사항으로 웹 URL 페이지를 이미지파일로 만들어서 이미지로그 DB에 저장시킨다.

이미지파일의 이름은 웹서버로부터 정보를 받은 시간으로 저장되며 이 시간은 웹서버에서 로그파일이 기록된 시간과 동일하다. 예를 들어 2008년 1월 1

일 11시 26분 54초에 클라이언트가 웹서버에게 요청하여 그 기록이 로그파일에 저장이 되었다면, 그 시간에 이미지로깅 서버에 이러한 정보를 보내고 이미지로깅 서버는 웹서버에서 전송받은 이 정보를 바탕으로 클라이언트의 아이피 주소가 이름인 폴더를 생성하고 그 폴더에 년월_시간:분:초 형식의 이미지로그파일을 생성, 저장한다. 이미지로그파일은 파일 이름에서 시간정보를 알 수 있다. (그림 4)는 디렉토리 구조와 생성된 이미지로그파일이 저장된 모습을 보여준다.



(그림 4) 이미지로그파일

3-3. 구현된 시스템의 장점

이미지로깅기법은 다음과 같은 기대를 할 수 있다. 첫째, 시간이 지나 웹 URL 페이지가 삭제된 후에도 그 시간의 웹 페이지를 확인할 수 있다.

둘째, 조작이 어려운 이미지파일로 로그파일을 저장함으로써 보안상 취약한 텍스트파일 형식의 로그파일의 문제점을 개선할 수 있다.

셋째, 파일 시스템에 디렉토리 관리 기능을 추가하여 디렉토리를 만들고, 파일을 저장함으로써 데이터의 관리 및 사용을 용이하게 하였다.

4. 결론 및 향후과제

본 논문에서는 웹서버의 이미지로깅 시스템을 구현해 보았다. 이 기법은 여러가지 장점이 있지만 이미지파일의 용량문제, 속도 등 문제점을 가지고 있다.

또한 아직 연구 초기단계이기 때문에 미흡한 점도 많이 발견된다.

향후에는 중복되는 이미지 로그파일을 알고리즘 등

을 통하여 클러스터링 할 수 있는 마이닝 기법, 이미지로그파일 분류 기법 등의 구체적인 방법을 고안하고 속도나 용량 문제 등도 개선점을 찾아야 할 것이다.

또한 실험을 통하여 본 시스템의 기법의 성능을 검증할 필요성이 있다.

참고문헌

- [1] 고광만, 박홍진 “모바일 웹 서버 관리기 구현”
- [2] 김선우, “윈도우 네트워크 프로그래밍”
- [3] <http://www.apache.org>
- [4] <http://www.logger.co.kr>
- [5] “Web Server Monitoring”, White Paper, <http://www.freshtech.com/WhitePaper.htm>
- [6] F. Stajano. Will your digital butlers betray you? In Proceedings of ACM Workshop on Privacy in the Electronic Society (WPES), 2004.
- [7] <http://www.guangmingsoft.net/>
- [8] Ross S. Finlayson, David R. Cheriton Log Files: An Extended File Service Exploiting Write-Once Storage, ACM 1987