웹 포렌식을 위한 이미지 로깅 서버 구현

유승희, 조동섭
이화여자대학교 컴퓨터공학과
e-mail : shyoo@ewhain.net

Implement Image Logging Server for Web Forensics

Seung-hee Yoo, Dong-sub Cho
Dept. of Computer Engineering, Ewha Womans University

요 약

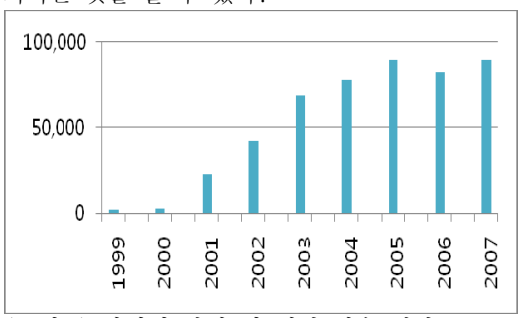
현재 컴퓨터와 인터넷, 정보기술의 발전은 사용자에게 편리함을 가져다 주었으나, 사이버 범죄라는 새로운 역기능을 가지게 되었다. 결국 이는 특정 목적을 가진 범죄자를 낳게되고 정보화 사회의 발전을 저해하는 커다란 걸림돌로 작용하게 되었으며, 이에 대응하는 정보보호기술은 개인의 사생활 보호와 국가 경쟁력을 판단하는 척도로 자리잡게 되었고, 현대에는 정보보호 기술 자체가 국가간 정보전 형태를 띠면서 그 중요성은 매우 커지고 있다. 이러한 정보보호 기술은 방화벽과 침입탐지 시스템의 꾸준한 개발로 이어졌으나, 아직 컴퓨터 범죄를 다루는 피해 시스템의 증거수집, 복구 및 분석을 하는 컴퓨터 포렌식 기술은 아직 활발히 연구되지는 않고 있다.

본 연구에서는 로그파일이 기록되는 시간의 웹 URL 페이지 이미지를 저장하여 이미지 로그파일을 만드는 멀티 쓰레드 TCP 서버를 구현하여 컴퓨터 사이버범죄에 대한 증거자료로서 디지털 포렌식인 이미지 로그파일을 제안하여 보았다

1. 서론

정보보호 기술은 기술 자체가 국가의 경쟁력이 되어가고 있다. 그 중 사이버테러에 대비하는 정보전 대응체계는 중요한 이슈로 떠오르고 있다.

현재 사이버 범죄에서 디지털 증거의 압수· 수색 및 분석을 위한 다양한 방법이 연구되어 활용되고 있다. 그림 1 과 같이 경찰청 사이버테러대응센터의 사이버방지 및 해킹기술 동향자료[1]에 의하면 2003 년 이후 사이버 방지 및 해킹기술이 급격히 증가하는 것을 볼 수 있다.



(그림 1) 사이버 방지 및 해킹 기술 경향

따라서 사이버 범죄의 급속한 증가에 따라, 증거수집의 기술 발전의 필요성도 증가한다. 컴퓨터를 이용한 사이버범죄를 수사하고 재판할 때 디지털 자료에 대한 증거로써 포렉식의 필요성과 중요성이 증대되고 있다.

이러한 컴퓨터 관련 사이버범죄가 발생한 후, 침입자의 흔적을 찾고자 할 때, 우리가 가정 먼저 취하는 행동은 침입자의 흔적(Digital Evidence)을 찾는 행위이다[2]. 이러한 행위에 가장 잘 사용되는 정보가 웹서버 내에 남아있는 로그파일에 관한 정보라 하겠다. 이러한 이유로 로그 정보는 불법적인 범죄자를 수사하기 위한 최소한의 흔적이 될 수 있고, 재판에서는 범죄자를 구속하기 위한 법적인 증거자료가 될 수 있다.

본 논문에서는 포렌식으로써의 이미지 로그파일을 제안하고 법적인 증거자료로서 보장받을 수 있는가에 관하여 분석하여 보았다.

논문의 구성은 다음과 같다. 먼저 2 절에서는 관련 연구에 대하여 논한다. 3 절에서는 이미지 로깅 서버의 구현 및 구동방식에 관하여 알아보고 마지막 4 절에서는 결론 및 향후 과제에 대하여 언급한다.

2. 관련 연구

시스템의 구현 방법 기술에 앞서 관련된 지식에 대해 알아보도록 하겠다.

2.1. 포렌식

“포렌식스”의 사전적 의미는 민사 또는 형사 재판에서 과학적이고 기술적인 방법을 사용하여 사건을 조사하고 어떠한 사실을 증명하는 일련의 방법들을 말한다[3].

즉 디지털 포렌식이란 pc, 웹, 휴대전화 등에 내장된 디지털 자료를 근거로 발생한 사실 관계를

증명하는 과정을 말한다. 포렌식에서 수집된 데이터는 법정에서의 증거자료로써, 효력을 발휘하기 위해서는 데이터의 특성을 잘 알아 안전하게 다루어져야 한다.

2.2. 웹로그 파일

클라이언트가 웹 서버에게 정보를 요청하고, 웹서버는 그에 대한 응답을 할 때마다 서버 소프트웨어는 로그파일로 그 기록을 남긴다.

로그란 컴퓨터 시스템에서 이루어지는 일련의 작업 내역을 기록하는 것으로 각각의 작업이나 수행에서 사용된 CPU 시간, 입출력 장치의 사용시간, 수행시킨 명령어, 시작과 종료 시간 등 컴퓨터 시스템의 운용에 대한 모든 기록을 의미한다[4]. 즉, 로그파일이란 웹 서버를 통해 이루어지는 모든 작업들에 대한 기록이라 할 수 있다.

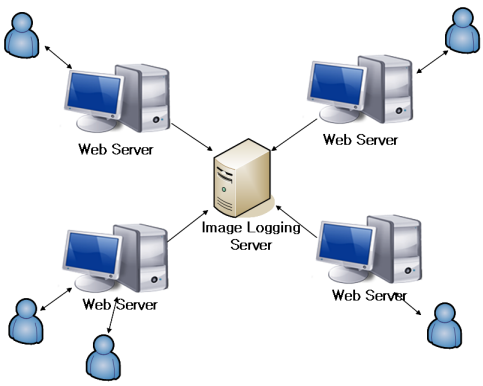
컴퓨터 시스템에 불법으로 침입한 공격자[12]는 흔적을 남기게 되는데 이러한 흔적이 저장되는 곳이 로그 파일이다.

대부분의 웹 로그 파일들은 NCSA(the National Center for Supercomputing Application)와 CERN(Conseil Europeen pour la Recherche Necleaire)에 의해 제정된 CLF(Common Log Format)를 따르고 있다. 이 CLF 는 주로 사용자 IP 주소, 사용자 ID, 접근시간, 요청방식, 접근한 페이지의 URL, 사용한 프로토콜, 에러코드, 전송된 데이터 크기 등의 정보를 포함하고 있다. 이를 통해 다양한 정보를 분석해 낼 수 있다. 일반적으로 웹 로그파일은 웹 서버가 지정한 곳에서 생성되는데 웹 서버 관리자가 웹 서버를 설치할 때 로그파일의 위치와 기록방법 등을 지정하게 되어있다[4].

3. 미니 웹서버와 TCP 서버를 이용한 구현방법

본 시스템은 웹 URL 페이지를 클라이언트에게 제공해주는 웹 서버와 웹 서버로부터 정보를 전달받아 이미지 로깅을 해주는 이미지 로깅 서버로 구현되었다.

(그림 2)는 본 시스템의 동작방식을 보여준다.



(그림 2) 구현한 시스템의 동작방식

동작 순서 및 방법은 아래장에서 언급하도록 하겠다.

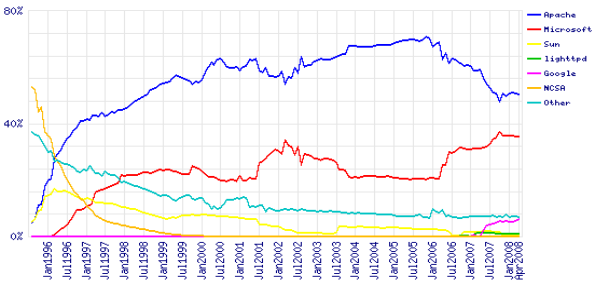
3.1.이미지 로그 파일

현재 웹서버의 로그파일은 확장자가 .txt 인 텍스트파일 형식으로 되어있으며 파일 특성상 조작이 쉬워 보안상의 문제점을 가지고 있다.

이를 보완하기 위하여 본 논문에서는 이미지 로그파일을 저장하는 웹서버를 구현하여 보았다.

3.2. 미니 웹서버

웹 서버는 현재 매우 다양한 종류가 존재하고 있으나 2008 년 4 월 현재 웹 서버의 사용 비율은 (그림 3)을 보면 Apache 웹 서버가 50.42%를 차지하고 있으며 점차 비율이 줄고 있는 추세이다[7]. MS 사의 웹 서버는 35.33%를 차지하고 있으며 점차적으로 사용비율이 증가하고 있다.



(그림 3) 웹 서버 시장 점유율 변화 추이

본 시스템 구현에는 Microsoft(US)사의 미니 웹 서버인 HTTPSvr Version 2.0 을 사용하였다. HTTPSvr 은 오픈소스 웹 서버로써 사용자가 원하는대로 소스를 수정하여 사용할 수 있다.

HTTPSvr 은 기본적으로 CLF(Common Log Format) 형식으로 저장되며, 해당 웹 로그 파일의 각 항목은 다음과 같이 구성되어 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
```

① ② ③ ④

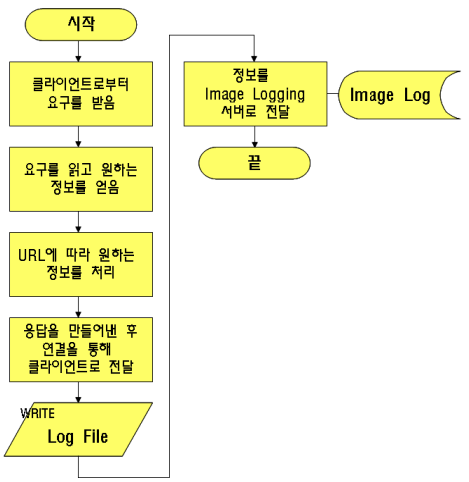
(그림 4) 웹 로그 파일의 일부

(표 1) 웹 로그 파일의 각 항목

번호	항목	설명
1	Status	접속에 대한 성공 여부 확인 코드, 처리상황
2	Host	접근한 사용자의 IP 또는 도메인
3	Time	사용자가 접근한 날짜와 시간
4	Request	요청사항

(그림 5)는 구현된 시스템에서 미니 웹 서버의 동작 순서를 보여준다. 웹 서버는 클라이언트로부터 요청이 들어오면 들어온 요청을 처리한다. 매번 요청이 들어오고 처리할 때마다 로그파일에 그 기록을 남긴다. 본 웹 서버의 로그파일에는 앞절에서 본 바와 같이 클라이언트의 아이피 주소, 접속시간, 요구사항, 서버의 에러원인 등의 정보가 저장된다.

구현된 웹 서버는 로그파일에 이러한 정보들이 기록될 때마다 소켓을 생성하여 이미지 로깅 서버로 클라이언트의 아이피 주소와 요구사항(클라이언트가 방문중인 웹 URL 페이지), 로그파일이 생성된 시간을 보낸다. 이때 웹 서버는 이미지 로깅 서버의 클라이언트가 되는 것이다.



(그림 5) 미니 웹 서버의 처리 순서도

3.3. 이미지 로깅 멀티쓰레드 TCP 서버

이미지 로깅을 하고 이미지 로그파일을 저장하기 위한 서버는 멀티쓰레드 TCP 서버로 구현하였다. 본 서버는 웹 서버에서 받은 정보를 바탕으로 이미지 로그파일을 만들어서 이미지로그 DB에 저장한다.

이미지 로깅 서버 시스템은 운영체제, CPU 등은 미니웹서버의 구현환경과 같으나 구현툴은 Visual Studio .net 을 사용하였다.

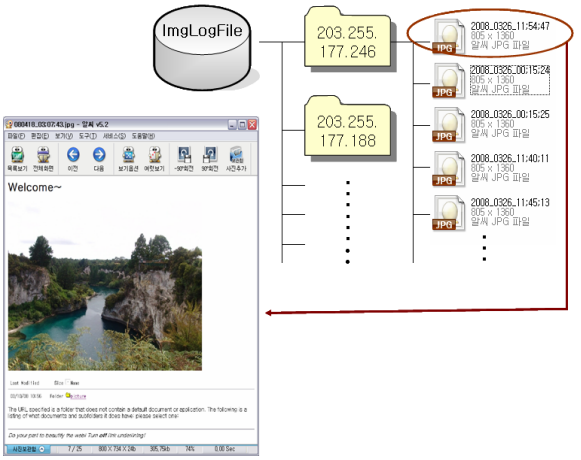
그리고 이미지 로깅 기법에는 Guangming Software(US)사의 HTML SnapShot 이라는 소프트웨어를 사용하였다[13]. 본 소프트웨어는 웹 URL 페이지를 이미지(JPG 파일)로 만들어주는 소프트웨어로써 Visual C++로 구현된 프로그램이다.

구현 시스템에서는 웹 서버에서 받은 정보인 클라이언트의 아이피 주소와 요구사항으로 웹 URL 페이지를 이미지파일로 만들어서 이미지로그 DB 에 저장시킨다.

이미지파일의 이름은 웹 서버로부터 받은 정보인 웹 로그파일 생성시간으로 저장되며 웹 서버 클라이언트의 IP 주소 폴더에 저장된다.

예를 들어 2008 년 1 월 1 일 11 시 26 분 54 초에 아이피 주소가 203.255.177.246 인 클라이언트가 웹 서버에게 요청하여 그 기록이 로그파일에 저장이 되었다면, 웹서버는 동시에 이미지 로깅 서버에 클라이언트의 정보들을 보내고 이미지 로깅 서버는 웹 서버에서 전송받은 정보들을 바탕으로 클라이언트의 IP Address 가 이름인 폴더를 생성하고 그 폴더에 년월일_시간:분:초 형식의 이미지 로그파일을 생성, 저장한다. 이미지 로그파일은 파일이름에서 시간정보를 알 수 있다.

(그림 6)은 디렉토리 구조와 생성된 이미지 로그파일이 저장된 모습을 보여준다.



(그림 6) 이미지 로그파일 디렉토리 구조

4. 디지털 증거로서의 이미지 로그파일

디지털 증거란 이진화되어 저장되거나 전송 중인 데이터로 그 특성상 복제가 쉬울 뿐만 아니라 원본과 사본의 구분이 어렵고, 조작, 변경, 삭제 등이 용이하다. 또한 매체 독립적으로 비가시적이라는 특성을 지니고 있으므로 디지털 증거를 법정에 제출하기 위해서는 가시적인 형태로 변화하여 제출하여야 한다. 위조나 변조 가능성 혹은 그 정보의 출처에 대한 인증의 부재로 인하여 수사시나 법정에서 참고 자료로서의 효력밖에 가지지 못하는 문제점이 있다[1].이에 따라 컴퓨터나 기타 디지털 저장장치로부터 수집된 디지털 증거가 법적 효력을 가지기 위해서는 진정성, 무결성, 신뢰성, 원본성이

보장되어야 한다[9].

증거 데이터의 진정성이란 저장, 수집 과정에서 오류가 없으며, 의도된 결과가 정확히 획득되었고, 그로 인해 생성된 자료임이 인정됨을 뜻한다. 무결성이란 범죄 현장에서 관련된 디지털 저장매체를 수집한 이후로 내부에 저장된 디지털 데이터가 법정에 제출되기까지 변경이나 훼손없이 보호됨을 말하며, 신뢰성이란 증거 데이터의 분석 등 처리 과정에서 디지털 증거가 위조되지 않았고 의도되거나 의도되지 않은 오류를 포함하지 않음을 뜻한다. 디지털 증거의 원본성이란 자체적으로 가시성과 가독성이 없는 디지털 증거를 변환하여 제출하는 과정에서 제출되는 증거 데이터가 원 매체에 있는 데이터와 동일함을 의미한다[8].

이미지 로그파일은 현재의 텍스트파일 형식의 로그파일의 저장 내용인 status, host, time, request 를 모두 포함하고 있고 쉽게 조작할 수 없어 보안상 훨씬 안전하다.

5. 구현된 시스템의 장점

첫째, 시간이 지나 웹 URL 페이지가 삭제된 후에도 그 시간의 웹 페이지를 확인할 수 있다.

둘째, 조작이 어려운 이미지파일인 .jpg 파일 형태로 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.

넷째, 이미지 로그만을 관리하는 TCP 서버 시스템을 구현함으로써 원래의 웹 서버 시스템에는 아무런 영향을 주지 않는다.

6. 결론 및 향후과제

본 논문에서는 웹 포렌식으로 활용할 수 있는 이미지 로깅 서버를 구현하여 보았다.

이미지 로깅 서버는 데이터의 무결성, 신뢰성 등을 확보할 수 있어서 법적 증거로서 매우 확실한 자료가 될 수 있다. HTTP(웹게시판, 블로그 등), FTP, 이메일, Telnet 및 메신저 등 사이버 범죄가 잦은 서버에서 웹 포렌식으로서의 본 이미지 로깅서버는 유용할 것이라 생각한다. 또한 웹서버와 독립적으로 작동하므로 웹서버 시스템에는 아무런 영향을 주지 않는다.

향후에는 DB 구축 및 중복되는 이미지 로그파일을 알고리즘 등을 통하여 분류 할 수 있는 마이닝 기법, 웹포렌식 자료의 추출에 사용되는 웹 포렌식 알고리즘 등의 구체적인 방법을 고안하고 용량 문제 등도 개선점을 찾아야 할 것이다. 그리고 명확한 법적 근거를 토대로 한 웹 포렌식의 발전 방향성이 제시되어야 하며 증거 수집의 무결성 확보 절차는 적법성을 인정하는 법적 근거 또는 사회적 합의가 중요하므로, 웹 포렌식의 지속적인 연구가 필요하다.

참고문헌

[1] 포렌식 법적문제
http://www.cftt.nist.gov.2006.8
[2] Linda Volonino, "Computer forensics and electronic discovery: The new management challenge", Computers & Security, 25(2), pp91-96, 2006.
[3] Warren G.Kruse II, Jay G.Heiser. “COMPUTER FORENSICS : Incident Response Essentials”, Addison Wesley, 2002
[4] http://dic.inuri.com
[5] 박대우, 서정만, "TCP/IP 공격에 대한 보안 방법 연구", 한국 컴퓨터정보학회논문지, 제 10 권 제 5 호, pp217-226, 2005
[6] R. Finlayson, D. Cheriton, "Log Files: An Extended File Service Exploiting Write-Once Storage", ACM, 21(5), pp137-148, 1987
[7] http://news.netcraft.com/
[8] http://www.guangmingsoft.net/
[9] Rahul Bhaskar, "State and local law enforcement is not ready for a cyber Katrina", Communications of the ACM, 49(2), pp81-83. 2006.
[10]고려대학교 산업협력단, "외국판례에 나타난 디지털 증거 수집, 분석, 보존 과정에서의 무결성 논란에 비추어 본 디지털 증거의 활용방안", 대검찰청, 2006
[11]김정옥, "디지털증거의 증거능력 인정 요건-일심회 판결을 중심으로", 디지털 포렌식연구
[12] 김선우, "윈도우 네트워크 프로그래밍 : TCP/IP 소켓 프로그래밍", 한빛미디어, 2004
[13] Mark Reith. Clint Carr. Gregg Gunsch, “An Examination of Digital Forensic Model”, International Journal of Digital Evidence, 1(3), 2002.
[14] Kanellis. p, Kiountouzis. E, Kolokotronis. N, Martakos. D, "Digital crime and forensic science in cyberspace", Idea Group Publishing Hershiey, 2006
[15] H. Custer, Inside Windows NT, Microsoft Press, 1993