침입탐지형 로그 분석기의 설계 및 구현

김도형[0], 김성준, 이원구, 이희규, 이재광
한남대학교 컴퓨터공학과
{dhkim, sjkim, wglee, june, jklee}@netwk.hannam.ac.kr

Design and Implement the Log Analysis Agent

Do-hyung Kim[0], Sung-jun Kim, Won-goo Lee, hee-kyu Lee, Jae-kwang Lee
Dept. of Computer Engineering, Hannam University

요 약

사용자가 웹사이트를 이용하면 이에 대한 기록이 로그라는 형태로 흔적이 남는다. 로그분석이란 데이터를 기반으로 위에서 말한 다양한 정보를 추출해 내는 것이라 할 수 있다. 리눅스 시스템은 사용자 로그인, 메일 등 모든 시스템 활동에 대한 로그를 기록하고 이를 가지고 시스템의 문제에 대해서 분석할 수 있다. 현재 로그 파일을 분석하는 대부분의 프로그램들은 Web로그에 초점을 맞추고 있으므로, 웹 이외의 다른 서비스에 대한 지원이 부족한 상태이다. 많은 국내외 제품들이 존재하지만 대부분의 프로그램이 Web로그에 치중하고 있다. 본 논문에서는 Web 로그 파일에 대한 분석뿐만 아니라 ftp, telnet, mail 서비스에 대한 로그 파일 분석을 통합적으로 수행하여 기존의 상용화 제품과는 차별화 된 로그 분석 도구를 개발하였다.

1. 서 론

일반적인 유닉스 또는 리눅스 서버는 기본적으로 서버에서 일어나는 모든 사항들을 파일 형태로 남기는 기능을 가지고 있다. 이때 저장되는 파일을 Log data라고 하며 이러한 data는 특별한 형태의 기준에 따라 숫자와 기호 등으로 기록된다. 이 기록을 통해 서버 관리자는 관리를 하는데 있어서 필요로 하는 유용한 정보를 생성 할 수 있다. 이러한 행위를 로그분석 이라고 한다. 로그 분석은 여러 가지 측면에서 상당히 중요한 의미를 갖는다. 광고주에게는 광고 효과 측정 및 광고 통계에 따른 광고 전략 수립에 큰 영향을 끼칠 것이며, 웹사이트 운영자에게는 이용자의 관심도, 시간대별 사용량 혹은 링크 에러 등을 측정할 수 있는 주요 지표로써, 쇼핑몰 운영자에게는 구매자의 형태 분석 및 개개인의 성향 분석에 대한 중요한 자료로써 활용이 될 수 있기 때문이다. 그리고 이러한 분석 결과는 웹사이트 운영 전략 수립에 있어서 가장 중요한 요소로 작용한다. 이처럼 중요한 로그 파일을 시스템 관리자가 소홀히 여긴다면 크나큰 문제를 발생시킬 수 있다. 시스템은 사용자 로그인, 메일 등 모든 시스템 활동에 대한 로그를 기록하고 이를 가지고 시스템의 문제에 대해서 분석할 수 있다. 시스템의 로그가 어떤 식으로 기록되고 어떤 의미를 가지고 있는지 이를 어떻게 활용해야 할 것인지를 관리자는 모두 알고 있어야한다. 대부분의 서비스가 로그 파일에 정보를 기록 유지한다. 규모가 큰 시스템일수록 로그파일을 접근하고 갱신하는데 많은 시스템의 자원을 소비한다. 즉, 관리자가 로그 기록에 신경을 쓰지 않는다면 대규모 서비스를 제공하면서 큰 문제를 야기할 수 있다.

이러한 시점에서 현재 우리는 리눅스 시스템을 이용하여 BBS, 학과 서버, 리눅스 서버 운영 중 사용자의 취향을 분석하여 서버운영의 질을 향상시키기 위해 시스템 로그에 대한 관리가 절실히 필요하다고 인식하여 로그 파일을 분석하여 사용자의 서버사용에 대한 정보와 보안에 대한 문제들을 해결하기 위해 각종 로그를 통합적으로 관리하고, 분석하여 관리자에게 다양한 정보를 제공하는 로그 분석기의 필요성을 느끼게 되었다. 그러나 상용화된 거의 대부분의 로그 파일 분석 프로그램이 웹에 초점을 맞추고 있기 때문에 웹 이외의 다른 서비스에 대한 지원이 부족한 상태이다. 많은 국내외 제품들이 존재하지만 대부분의 프로그램이 웹 로그에 치중하고 있다. 또한, 로그 파일은 침입을 탐지하기 위한 가장 중요하면서도, 기본적인 자료가 된다. 이러한 로그 파일을 한 시스템에서 분석하여 자신의 시스템에서만 활용하는 것보다는 다른 시스템에서도 활용할 수 있게 설계하고 구현한다면 보다 효율적인 로그 분석 도구가 될 것이고, 외국 기술에 의존하는 것보다는 국내 실정에 맞는 로그 분석 프로그램을 개발하고 사용하는 것이 바람직하다. 본 연구에서는 웹 로그 파일에 대한 분석뿐만 아니라 FTP, 텔넷, 메일 서비스에 대한 로그 파일 분석을 수행하여 기존 제품들과 차별화 한다. 그리고 리눅스 시스템 관리자에게 방대한 양의 시스템 로그파일을 분석하고, 이를 기반으로 다양한 리포트 기능을 제공하고, 시스템 해킹을 사전에 방지하는 시스템 개발을 목표로 한다.

2. 관련연구

2.1 로그분석 시스템

정보시스템에서는 보유하고 있는 정보 및 자원의 불법 유출을 방지하고 사용 원칙에 위배되는 불법 행위의 추적을 위한 감사 능력이 제공되어야 하며 나아가 시스템 관리 및 운영자의 책임을 명확히 하고 사용자의 행위를 명확히 구분할 수 있는 감사 증적(Audit Trail) 메커니즘이 요구된다. 감사 증적은 언제, 누가, 어떤 자원을, 어떻게 이용하는가 하는 자료를 기초로 하여 아래와 같은 용도로 이용될 수 있다. [1]

- 백업, 통계유지 등 시스템 사용 현황 파악 및 시스템 증설 기초 자료로 이용
- 사용료 청구 등 회계측면의 기초 자료로 이용
- 시스템 자원 사용에 대한 모니터링 및 로깅에 의한 추적자료로 이용

2.2 침입탐지 시스템

인터넷의 확장에 따라서 네트워크를 통한 침입의 가능성이 증가되었고, 이에 따라 시스템이나 네트워크 침입을 즉각적으로 탐지하고 대처할 능력이 있는 기술이 필요하게 되었고, 이러한 기술을 이용하여 자동으로 침입을 탐지, 보고, 조치하는 자동화된 시스템이 필요하게 되었다. [2]

침입탐지 시스템의 목표는 아래와 같이 크게 두 가지로 설명할 수 있다.

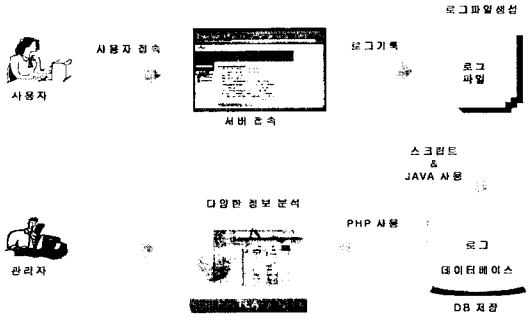
- 침입자에 의한 불법적인 사용을 탐지하는 것 (Anomaly Detection)
- 합법적인 사용자에 의한 오용이나 남용을 탐지하는 것 (Misuse Detection)

2.3 침입탐지형 로그 분석기 설계

2.3.1 침입탐지형 로그 분석기 구조

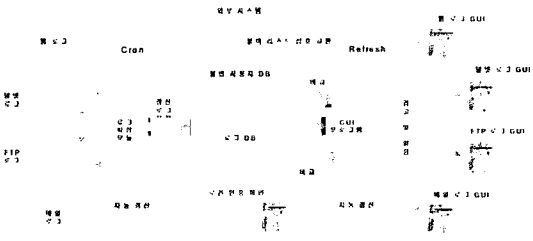
가. 전체구조

로그 분석기의 로그파일 분석 흐름의 [그림 1]에서와 같이, 서버에 접속한 로그를 로그 데이터베이스에 저장한 후, 모든 사용자 및 관리자에게 유용한 로그 정보를 실시간으로 웹 프로그램을 통해 다양한 GUI 형태로 보여주게 된다. 그리고, 통합 로그 분석기의 기본적 구성은 각 로그별로 '로그 파싱 모듈'에 의해 파싱되어 로그 데이터베이스에 저장된다.



[그림 1] 로그 분석 흐름도

이 과정은 주기적으로 로그 파일을 분석하여 데이터베이스에 저장하도록, 일정 시간마다 수행되어야 하는 작업들을 관리해 주는 "/etc" 디렉토리에 존재하는 cron을 이용한다. 저장된 로그 데이터베이스는 관리자와 사용자가 웹 상에서 요구하는 각 로그별 분석 결과를 관리자와 사용자에게 보여준다. 이 과정에서도 주기적으로 변동하는 로그 분석 내용을 실시간으로 관리자와 사용자에게 보여 주기 위해 HTML 태그를 이용하여 주기적으로 현재 페이지를 갱신한다.



[그림 2] 로그 분석기의 전체 구조도

[그림 2]에서는 관리자와 사용자에게 보여줄 로그 분석 절차와 로그 분석 결과의 전체적인 구조를 보여주고 있다.

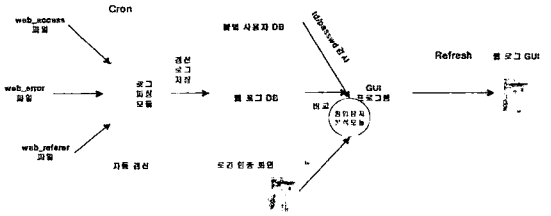
나. 모듈별 구조

1) 웹 로그 분석 모듈

웹서버에서도 시스템을 어느 사이트로부터 접속하였으며, 어느 파일이 다운로드 되었는지에 대한 기록이 access_log 파일에 기록되고, 존재하지 않는 파일에 대한 접근 등의 에러에 대해서는 error_log에 기록된다. 웹서버에 대한 공격은 주로 CGI 프로그램에 집중되고 있는데 취약한 CGI 프로그램에 대한 공격도 이들 로그 파일에 기록된다.

최근에 FTP를 이용하여 해킹 도구를 설치하는 경우도 있지만, 웹을 이용하여 해킹 도구를 다운로드 받는 경우가 늘고 있어 웹 로그의 분석도 중요시되고 있다. 이처럼 웹 로그는 웹서버에 얼마나 많은 사용자가 접속해 있고, 어는 시스템으로부터 접속을 시도했으며, 어느 파일에 대한 접근이 가장 빈번했는지 등의 웹 통계를 내기 위한 용도뿐만 아니라 보안 목적에서도 분석이 이루어져야 한다.

웹 로그 분석 모듈은 [그림 3]에서와 같이 아파치 홈 디렉토리의 'logs' 디렉토리에 위치하는 access_log, error_log, referer_log 파일을 '로그 파싱 모듈'을 이용하여 필요한 데이터를 선별, 각각의 데이터베이스에 저장하여 관리자와 사용자에게 로그 분석 프로그램을 통하여 로그 분석 정보를 제공한다.



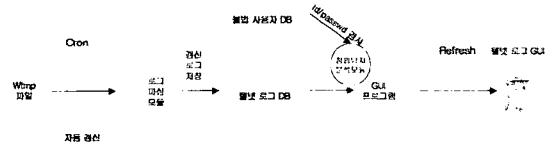
[그림 3] 웹 로그 분석 모듈 구조도

2) 텔넷 로그 분석 모듈

wtmp 파일은 사용자들의 로그정보를 가지고 있다. utmp 파일과 마찬가지로 바이너리 형태이며, 자료구조도 역시 utmp 구조를 사용한다. utmp 파일이 현재 로그인한 사용자에 대한 기록이라고 하면, wtmp는 지금까지 사용자들의 로그인, 로그아웃 히스토리를 모두 가지고 있고, 시스템의 shutdown. booting 히스토리까지 포함하고 있어, 해킹 피해시스템 분석에서 대단히 중요한 로그라고 할 수 있다. wtmp 파일에서 사고분석을 위해 주의 깊게 살펴봐야 할 부분은 다음과 같다.

- 접속시간이 정상적인가? 일반적으로 국외에서 공격을 받았을 경우 우리나라와 시간대역이 틀려, 국내에는 새벽시간대에 침입한 것으로 흔적이 남는 경우가 많다.
- 접속출처가 정상적인 위치인가? 접속하는 IP가 아닌 곳에서 접속하였거나, 특히, 국외 IP 주소에서 접속한 경우는 의심할 필요가 있다.

또한, last 명령을 통해서 원격에서 접속한 호스트를 확인했는데 도메인 네임이 전부 화면에 나타나지 않는 경우가 있다. 즉, last 명령에서는 공격자 추적이 중요한 자료인 원격 호스트명이 16문자까지만 화면에 보여지는데, 이를 넘는 도메인 네임의 경우 어디에서 접속했는지 알 수 없는 경우가 흔히 있다. 리눅스 시스템에서는 secure 파일에 인증 관련 접속로그가 텍스트 형태로 기록되는데, 여기서는 도메인 네임의 문자수에 관계없이 기록이 된다. 따라서, 로그 파싱 모듈을 통해 파싱하기 전에 이러한 파일을 참조할 필요가 있다.



[그림 4] 텔넷 로그 분석 모듈 구조도

그리고, 모든 로그파일이 그렇지만 wtmp 파일도 일정시간을 주기로 갱신(rotate)된다. 이전의 wtmp 파일은 wtmp1 파일에 저장되는데 이 파일에서도 접속 로그를 확인할 필요성이 있다.

텔넷 로그 분석 모듈은 [그림 4]에서와 같이 '/var/log/wtmp' 파일을

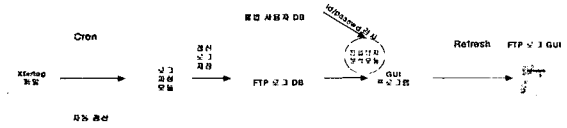
'로그 파싱 모듈'을 이용하여 필요한 데이터를 선별, 데이터베이스에 저장하여 관리자와 사용자에게 로그 분석 프로그램을 통하여 로그 분석 정보를 제공한다.

3) FTP 로그 분석 모듈

xferlog는 ftp 데몬을 통하여 송수신되는 모든 파일에 대한 기록을 제공한다. xferlog 파일에서 접속시간과 원격지 시스템의 적정성, 그리고 로그인 사용자 등을 살펴보아야 할 것이다. 그리고, xferlog에서는 송수신한 파일이 해킹 도구나 주요 자료인지 여부도 주의 깊게 보아야 한다. xferlog 파일에는 다음의 정보가 저장된다.

- 송수신 자료와 시간
- 송수신을 수행한 원격 호스트
- 송수신된 파일의 크기
- 송수신된 파일의 이름
- 파일의 송수신 모드
- 특수 행위 플래그
- 전송 방향
- 로그인한 사용자의 종류

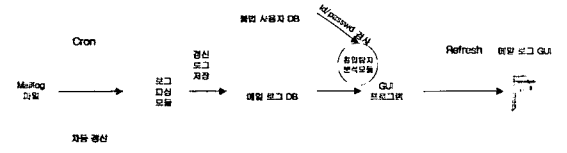
FTP 로그 분석 모듈은 [그림 5]에서와 같이 '/var/log/xferlog' 파일을 '로그 파싱 모듈'을 이용하여 필요한 데이터를 선별, 데이터베이스에 저장하여 관리자와 사용자에게 로그 분석 프로그램을 통하여 로그 분석 정보를 제공한다.



[그림 5] FTP 로그 분석 모듈 구조도

4) 메일 로그 분석 모듈

메일 로그 분석 모듈은 [그림 6]에서와 같이 '/var/log/maillog' 파일을 '로그 파싱 모듈'을 이용하여 필요한 데이터를 선별, 데이터베이스에 저장하여 관리자와 사용자에게 로그 분석 프로그램을 통하여 로그 분석 정보를 제공한다.



[그림 6] 메일 로그 분석 모듈 구조도

3. 결론

국내 및 국외에서도 대부분의 로그분석 프로그램이 웹에 기준을 두고 있다. 본 시스템은 Web, ftp, telnet, mail 서비스에 대한 로그 파일 분석을 수행하여 웹 브라우저를 통하여 모든 기능을 제어하여 다양한 보고서 기능을 제공하는 로그 분석기이다.

현재 리눅스 시스템은 사용자 로그인, 메일 등 모든 시스템 활동에 대한 방대한 양의 로그를 기록하고 이를 사용해 시스템의 문제를 파악할 수 있다. 그러므로 관리자는 로그가 어떤 방식으로 기록되고 어떤 의미를 가지고 있는지, 이를 어떻게 활용할 것이지를 모두 알고 있어야 한다. 로그 분석기는 방대한 양의 로그 파일을 실시간으로 빠르게 분석하여, 보고서 기능을 제공하고, 각종 해킹 시도에 대한 정보를 제공하여 관리자가 서버를 운영할 방향과 개선책, 시스템 보안에 관한 부담이 크게 감소할 것으로 기대된다.

참고문헌

[1] 한국정보보호 진흥원 "로그파일 위•변조 방지 기술"
http://www.kisa.or.kr/technology/sub3/logfile.html
[2] 한국정보보호 진흥원 "실시간 침입탐지기술"
http://www.kisa.or.kr/technology/sub3/ids.htm
[3] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee "Hypertext Transfer Protocol – HTTP/1.1", IETF RFC 2616, June 1999
[4] H. Nielsen, P. Leach, S. Lawrence "HTTP Extension Framework", IETF RFC2774, February 2000
[5] J. Postel "SIMPLE MAIL TRANSFER PROTOCOL", IETF RFC821, August 1982
[6] J. Postel, J. Reynolds "TELNET PROTOCOL SPECIFICATION", IETF RFC854, May 1983
[7] J. Postel "FILE TRANSFER PROTOCOL", IETF RFC765, June 1980
[8] 이찬섭, 박용문, 최의인 "클라이언트/서버 데이터베이스 시스템에서 역방향 로그분석을 이용한 로그 관리", 한국통신학회논문지 Vol.22 No.2 pp.1119-1122, 2000.11