논문 2009-1-3

검색엔진을 이용한 취약점 분석 시스템 개발

Development of Vulnerability Scanner using Search Engine

주복규*, 민병우**, 장문석**, 안창균**, 양동혁**

Bok-Gyu Joo, Beung-Woo Min, Moon-Suk Chang, Chang-Kyum Ahn and Dong-Hyuk Yang

요 약 컴퓨터 사용이 보편화되고 언제 어디서나 인터넷 연결이 가능하게 되어 보안에 대한 위협은 급속히 증가하고 있다. 우리는 사용자가 쉽게 보안 취약점을 점검할 수 있는 보안 스캐너를 개발하였다. 기존의 보안 스캐너는 새로운 취약점이 알려지면 보안 서버에 그 내용을 갱신하고 나서 고객 서버를 점검하는 체제인데 반해, 우리 시스템은 검색엔진을 이용하여 취약점들을 자동으로 모아서 보안 서버에 색인해 두고, 사이트 관리자가 별도의 갱신 없이 보안 서버를 이용하여 사이트의 보안 취약점을 점검할 수 있게 해준다.

Abstract In these days, security threat is ever increasing as computer systems and networking is everywhere. This paper is on the development of security scanner using search engine, with which site managers can easily check security vulnerability on their systems. Our security server automatically collects security-related information on the Internet, and indexes them in the database. To check the vulnerability of a customer server, the client system collects various system-specific information, and sends necessary queries to our security server for vulnerability checking. Up-to-date and site-specific vulnerability information is retrieved through the viewer, which allows the customer effectively to check and respond to security threat on client systems.

Key Words : *System Vulnerability, Security Server, Vulnerability Scanner*

Ⅰ. 서 론

현대는 회사 업무뿐만 아니라 대부분의 개인 활동이 온라인으로 이루어지는 유비쿼터스(ubiquitous) 시대이다. 이에 따라 보안 위협은 증가하고 보안 시스템은 일상에서 중요한 도구가 될 것이다.

2008년에는 컴퓨터 보안과 관련하여 대형 사건들이 일어났으며, 그 중 옥션 해킹사건은 가장 큰 사건으로 주목을 받았다[1]. 옥션에 대한 공격은 해외 크래커에 의해 악성코드가 옥션 직원을 대상으로 무작위로 뿌려졌고 이에 감염된 관리자 계정이 크래커에 의해 탈취되면서 발생된 사건으로 알려져 있다. 그 외 국민은행 해킹 사건, 월드오브워크래프트 게임 해킹사건 등 대형 사건들이 자주 발생하게 되면서 인터넷이 생활화 된 국민들에게 개인 정보 유출은 심각한 문제가 되었다. 그림 1은 우리가 일상적으로 사용하는 웹이라는 공간이 얼마나 취약한지를 보여주고 있다[2].

이 논문은 이러한 컴퓨터 내부의 정보 유출과 웹 취약점에 대한 공격에 대비할 수 있는 보안 시스템의 개발에 관한 것이다. 우리가 개발한 시스템은 크게 취약점 정보들을 검색하는 검색기 부분과 이 정보를 수집하여 서버 내로 가지고 오는 수집기 부분이 있으며, 이를 위해 검색엔진을 이용하여 개발하였다.

본 논문의 구성은 다음과 같다. 2장에서 검색엔진 기술과 우리가 사용하는 루신 라이브러리를 소개하고, 3장

*정회원, 홍익대학교 컴퓨터정보통신공학과
**준회원, 홍익대학교 컴퓨터정보통신공학과 HUST
접수일자 2008.12.3, 수정완료 2009.2.05

에서는 관련 제품을 비교한다. 4장에서는 본 연구에서 개발한 보안 스캐너의 설계와 구현, 그리고 실행 결과들을 소개한다. 마지막장은 결론과 향후 계획이다.

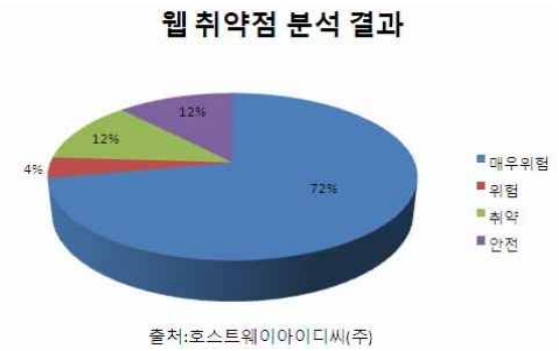


그림 1. 2008년 웹사이트 취약점 분석 결과
Fig 1. Vulnerability Analysis in 2008

Ⅱ. 검색 엔진 기술

검색엔진이란 인터넷에서 정보를 찾기 위한 필수 도구이다. 일반인들은 보통 실제 '검색엔진'과 '디렉토리 서비스'를 구분하지 않고 사용하고 있다. 두 방법은 리스트를 어떻게 컴파일 하느냐에 따라 검색 방법에 차이가 있는데 우리는 디렉토리가 아닌 검색엔진의 원리를 이용하는 것이다[3].

검색엔진의 동작은 스파이더 기능과 색인 기능이 결합되어 이루어진다. 스파이더 기능은 스파이더가 웹페이지를 방문하고 내용을 읽은 후에 사이트 내의 또 다른 페이지 링크를 파고 들어가는 기능이다. 검색엔진의 다른 주요 기능인 색인 기능은 스파이더가 찾아낸 모든 웹페이지 정보를 담아놓고 웹페이지의 내용이 바뀌면 색인 부분의 내용도 새롭게 갱신되도록 하는 것이다.

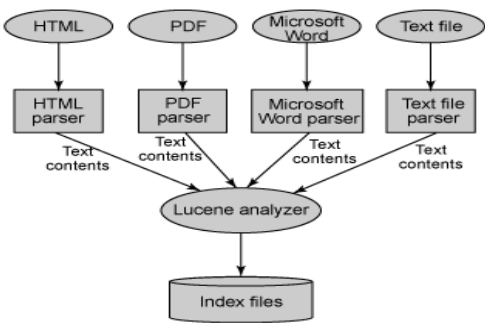


그림 2. 루신의 구조
Fig 2. Structure of Lucene

루씬(Lucene)은 Apache 재단에서 수행하는 오픈 소스 프로젝트의 일환이며 라이브러리 형태로 제공 된다. 기본적으로 자바 언어로 개발되어 있으나 Delphi, Perl, C#, C++, Ruby, PHP같은 다양한 언어로도 제공되고 있다[4].

루씬 자체는 HTML, PDF, Doc등의 문서를 읽어 내거나 데이터베이스에 접근 할 수 없다. 그림 2에서 보는 것처럼 각각의 콘텐츠를 파싱하여 텍스트 형식으로 읽을 수 있도록 게이트를 만들고, 라이브러리를 이용하여 해당 문서를 분석하고, 개발자가 제공한 가중치를 부여하여 색인을 한다. 색인된 데이터를 기준으로 검색하여 해당 문서의 특정 단어나 구문이 포함된 컨텐츠를 찾게 해준다. 우리는 스파이더를 통해 얻어진 다양한 보안 취약점 정보들을 루씬 라이브러리를 이용하여 효율적으로 색인하고 색인된 데이터를 통해 검색기능을 구현 하는데 루씬을 이용 하였다.

Ⅲ. 관련 제품 비교

기존의 취약점 점검 스캐너라고 불리는 프로그램들은 웹과 네트워크의 취약점을 점검해주는 프로그램으로서 대상 웹 사이트나 서버 네트워크의 취약점을 공개된 공격방법을 이용하여 외부에서 점검해 준다. 새로운 공격방법이 알려지면, 기업 보안 서버의 데이터를 업데이트 한 뒤, 클라이언트 측 프로그램의 데이터를 체크하여 업데이트 시키는 과정을 거친다. 이 방식은, 만약 기업 서버의 업데이트가 늦어지면 그만큼 클라이언트들이 위험에 노출된다. 또한, 점검하려는 대상이 자신의 웹 사이트이면 점검이 되지만 대상이 다른 사람의 웹 사이트라면 그것은 공격이 된다[5].

대표적인 취약점 점검 스캐너로는 Acunetix (www.acunetix.com)의 Web Vulnerability Scanner와 Tenable Network Security(www.nessus.org)의 Nessus가 있고, 구글의 취약점 검색 결과를 이용하여 점검을 수행하는 유명 해커 그룹인 CDC(Cult of the Dead Cow)의 Goolag 스캐너(www.goolag.org)도 많은 관심 대상이다. 국내에서도 여러 회사에서 웹 방화벽, 취약점 점검 스캐너 등을 개발하고 있으며, 패닉 시큐리티사 (www.panicsecurity.com)의 웹 취약점 점검 스캐너가 Active X, Flash 등을 많이 사용하는 국내 웹 환경에 맞

게 개발되어 많은 곳에서 사용되고 있다.

우리가 개발한 스캐너(H-Scanner)가 기존 스캐너와 가장 큰 차이점은 검색엔진 기술을 이용하였다는 점과 서버의 외부가 아닌 내부에서 취약점을 점검한다는 것이다. 기존의 스캐너들은 새로운 공격방법이 나오고 그 정보가 업데이트되기 전까지는 사용자가 점검을 하여도 그 취약점에 대해 알 수 없다. 하지만, H-Scanner는 검색엔진 기술을 이용하여 새로운 취약점 정보에 대한 업데이트 시간과 노력을 최소화한다.

H-Scanner는 크롤러(crawler)라는 봇을 이용하여 웹상의 취약점 정보들을 수시로 수집하고 자동적으로 검색서버의 데이터베이스를 갱신한다. 고객은 자신의 서버환경에 맞는 검색질의어를 만들고, 중앙 서버의 색인된 취약점 정보들을 검색하여 자신의 서버를 점검한다. 이 방식은 서버 내부의 취약점을 점검하는 것이기 때문에 내부에서 어떠한 공격도 이루어지지 않게 한다. 사용자 서버를 점검하기 위해서는 검색 서버의 인증을 받아야하기 때문에 다른 사용자가 자신의 서버를 점검하거나 자신이 다른 서버를 점검하는 것은 불가능하고 다른 목적으로는 사용할 수 없다.

IV. H-Scanner

이 장에서는 우리가 개발한 H-Scanner의 구조 및 동작 원리 그리고 실제 동작을 예를 들어 설명한다. H-Scanner는 자바 언어로 개발하였고, 루씬 검색 엔진을 이용하여 완성하였다[6]-[7].

1. 구조 및 동작 원리

H-Scanner의 검색 서버는 그림 3에서와 같이 인터넷의 웹 사이트 들에서 취약점 정보를 수집해서 루신 검색엔진 라이브러리를 이용하여 데이터베이스화 한다. 취약점 점검은 점검 대상 시스템에 대한 정보에 따라 만들어진 질의어 형태로 소켓 통신을 통해 전달받아 데이터베이스에 색인되어있는 알맞은 취약점 정보를 검색하여 서버 관리자에게 통보하고 서버관리자는 검색결과를 이용하여 해당 서버의 보안을 하게 된다.

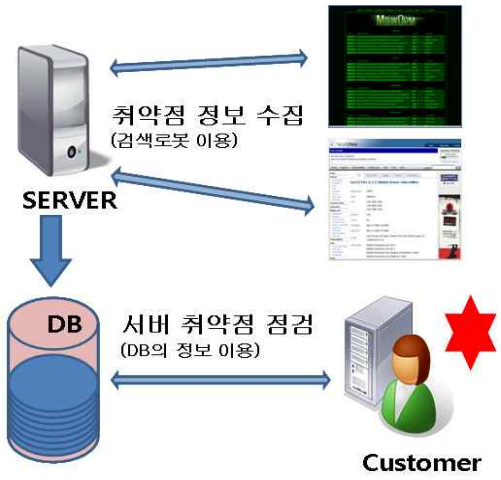


그림 3. H-Scanner 개념도
Fig 3. Concept of H-Scanner

검색 서버에는 취약점 정보를 수집하는 기능과 찾아낸 취약점을 색인하는 기능이 있다. 수집기능은 자체 제작된 스파이더를 이용하여 웹 사이트나 메일링 리스트, 각종 문서 등을 파싱하여 하이퍼 링크된 주소 부분이나 참조할 문서들을 찾은 뒤, 연결된 다른 문서를 대상으로 다시 수집과정을 거친다. 이러한 방식으로 최대한 많은 취약점 정보를 자동으로 수집하고 이를 루씬 라이브러리를 이용하여 색인한다[8]-[9].

취약점 정보를 검색하기 위하여 먼저 점검 대상 서버에서 시스템 정보를 수집한다. 즉 대상 시스템의 운영체제와 CPU의 종류, 메모리 상태, 파티션 정보, 네트워크 설정 정보 등을 수집한다. 이를 바탕으로 실행 중인 프로세스를 통해 검색 대상 어플리케이션을 찾는다. 예를 들어 httpd라는 데몬이 있을 경우 Apache 웹서버가 실행되고 있으므로 이를 검색 대상에 포함시킨다. 다음은 열려있는 포트를 검색하여 현재 실행되고 있는 어플리케이션을 찾는다. 또한 해당 취약점이 특정 어플리케이션 버전에서만 존재할 수 있으므로 버전 정보도 수집한다.

이렇게 얻어진 시스템에 관한 정보를 바탕으로 질의어를 생성 후 검색 서버로 보내어 색인된 데이터베이스에서 질의어에 맞는 취약점을 뷰어로 반환하여 관리자가 이를 볼 수 있도록 한다.

2. 검색 대상 시스템의 정보 수집

이 절에서는 H-Scanner의 동작을 실제 사용 예를 통하여 설명한다. 우선, 대상 시스템의 기본적인 정보를 수집한다. 즉, 운영체제 정보, CPU 정보, 메모리 할당 정보,

파티션 정보, 스왑 파티션 정보, 스왑 메모리 할당 정보, 네트워크 설정 정보 등이다.

그림 4. 실행중인 프로세스
Fig 4. Running Processes

다음은 그림 4에서 보는 바와 같이 대상 서버에 실행중인 프로세스 정보를 수집한다. httpd 프로세스가 실행중이므로 현재 Apache 웹 서버가 설치되어 운영 중인 것을 알 수 있고, mysqld_safe 프로세스를 보고 MySQL 데이터베이스가 운영 중인 것을 알 수 있다. 따라서 이 두 소프트웨어를 검색 대상에 포함시킨다.

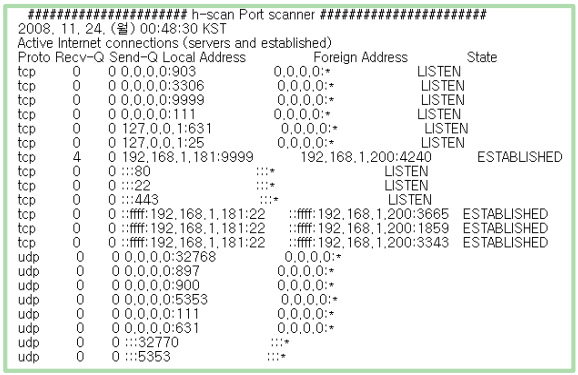


그림 5. 포트 스캔 정보
Fig 5. Port Scan Information

다음에는 포트 스캔 정보를 수집한다. 그림 5는 H-Scanner의 포트 스캔 정보를 보인다. 이전 과정에서 알아내었던 웹 서버의 존재를 80번 포트가 LISTEN상태인 것을 통하여 다시 한 번 확인할 수 있고, 3306번 포트가 LISTEN상태인 것을 통하여 MySQL 데이터베이스가 운영 중임을 확인할 수 있다. 또한, 25번 포트를 통하여 Sendmail, 22번 포트를 통하여 SSH가 설치되고 운영되는 것을 알 수 있으므로 이들을 검색 대상에 포함시킨다.

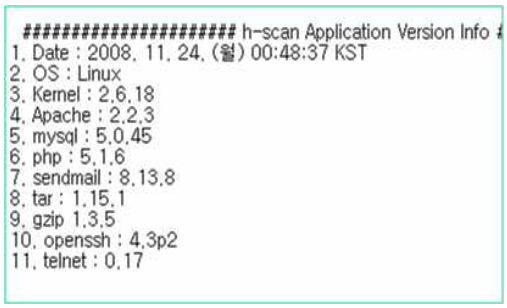


그림 6. 어플리케이션 버전 정보
Fig 6. Application Version Information

다음은 어플리케이션 버전 정보를 수집한다. 그림 6에서 보는바와 같이, 기 수집된 정보를 통하여 알 수 있었던 커널, 프로세스와 어플리케이션의 버전을 확인함으로써, 좀 더 정확한 질의어 생성이 가능하고 해당 버전에 대한 정확한 취약점 정보를 검색할 수 있게 된다.

3. 정보 분석

수집된 정보를 바탕으로 로그분석, 트래픽분석과 취약점 분석을 수행한다. 그림 7은 로그분석 결과를 보여준다. 오픈 소스인 logwatch를 이용하여 유닉스 서버의 로그를 분석한 것으로 분석된 로그를 통하여 관리자는 침해 여부를 판단할 수 있다.

그림 7. 로그 분석
Fig 7. Log Analysis

다음은, 오픈소스인 Mrtg를 이용하여 대상 서버의 시간대별 트래픽 확인이 가능하며 시간대별 CPU/메모리 사용량 확인을 가능하게 하였다.

4. 취약점 분석

위에서 수집한 대상 시스템에 대한 상세한 정보를 이용하여 검색 서버로부터 취약점 정보를 찾아서 점검을 한 뒤, 그림 8과 같이 뷰어를 통하여 그 결과를 확인할 수 있다. 취약점을 점검하고 심각성, 대응방법을 제시함으로써 관리자가 사전에 공격에 대응할 수 있도록 한다.

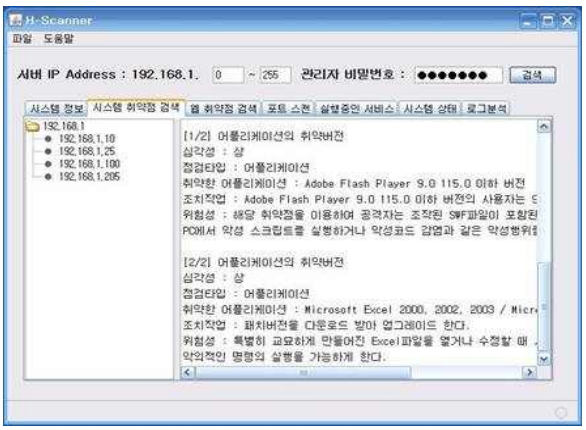


그림 8. 뷰어를 통한 취약점 점검 결과 확인
Fig. 8. Vulnerability Check Result on Viewer

5. 시스템 평가

우리가 개발한 H-Scanner가 실제 시스템에 잘 적용되는지 어떤 방식으로 작동될지 확인해보기 위해 주변 서버관리자들을 대상으로 시스템을 사용하게 해보았다.

우리가 개발한 보안 서버를 사용하여 취약점을 탐색해 본 결과 접속하는 것으로 취약점이 자동으로 분석되고 그 결과를 알기 쉽게 뷰어를 통하여 확인할 수 있다는 것이 시스템의 장점이라고 공통적으로 평가해 주었다. 그에 반해서 취약점을 분석하는데 시스템을 점검해야 하는 만큼 시간이 오래 걸린다는 문제점이 제기 되었고, 또한 관리자 비밀번호를 사용하게 하더라도 위험에 노출될 수 있지 않을까 하는 우려를 지적해 주었다.

V. 결 론

기존의 취약점과 관련한 시스템은 시스템을 사용하는 사람이 주기적으로 업데이트해야 하는 번거로움이 있었다. 신속한 업데이트가 없으면 새로운 취약점에 약점을 노출하는 일이 빈번히 발생하였다. 본 연구에서는 그러한 점을 해결하고자 보안 서버가 검색엔진의 원리를 이용해 수시로 인터넷에서 웹 사이트의 취약점을 분석하고 수집하여 DB화하고 다른 서버관리자가 이용함으로써 자동적으로 서버의 취약점을 분석해주는 시스템을 개발하였다.

우리가 개발한 스캐너의 가장 큰 장점은, 검색서버가 인터넷에 있는 취약점 정보를 수시로 수집하고 색인하여 저장해 두기 때문에, 고객 시스템이 취약점에 노출될 확률이 낮아진다는 것이다. 또한 고객 사이트 관리자가 취약점 정보를 찾을 때, 자신의 시스템에 최적인 질의어를 사용할 수 있으므로 가장 정확한 취약점 정보를 알아낼 수 있다.

H-Scanner를 널리 사용되도록 하기 위해서는 먼저 사용자 평가에서 나타난 단점들을 개선하여 더 편리하게 만들어야 한다. 또한 현재의 시스템은 유닉스나 리눅스 시스템을 사용하는 서버에 한정되어 있다. 앞으로 윈도우등 다른 OS환경에도 호환이 가능하도록 하여 더 많은 종류의 시스템 취약점을 분석해낼 수 있도록 개선이 필요하다.

참 고 문 헌

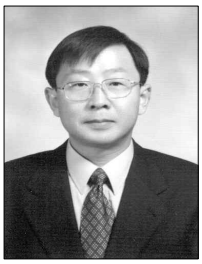
[1] http://www.inews24.com/bizmeka/itinfo/news_view.php?g_serial=324577&g_menu=022600
[2] http://hostway.co.kr/company/pr/news_view.html?type=news&page=2&number=1183
[3] http://www.webpromo.co.kr/searchTip/tips/working.html
[4] http://lucene.apache.org/java/docs/index.html
[5] 대피드 스투타드, 마커스 핀토, 웹 해킹 & 보안 완벽 가이드-웹 애플리케이션 보안 취약점을 겨냥한 공격과 방어, 에이콘 해킹.보안 시리즈, 에이콘 출판, 2008.
[6] 이영현, 웹과 DB연동을 이용한 검색 엔진의 설계 및 구현, 한국교원대학교 석사학위논문, 1998. 2
[7] 김광미, 정보 흐름 그래프를 이용한 웹 검색 성능 향상에 관한 연구, 조선대학교 석사학위논문, 2005. 2

[8] http://www.joinc.co.kr/modules/moniwiki/wiki.php/JCvs/Search/Document/Lucene
[9] 에릭 해처, 오티스 고스 포드네티츠, 루씬 인 액션, 에이콘출판, 2005.

※ "이 논문은 2006학년도 홍익대학교 학술연구진흥비에 의하여 지원되었음"

저자 소개

주 복 규(정회원)



- 1997년 서울대학교 계산통계학과 학사 졸업.
- 1980년 한국과학원 전산학과 석사 졸업.
- 1990년 메릴랜드대학교 전산학과 박사 졸업.
- 1990년~1998년 삼성전자 중앙연구소 수석연구원
- 1998년~2000년 (주)동양시스템즈 연구소장.
- 2001년~현재 홍익대학교 컴퓨터정보통신공학과 교수.
- 2004년~2007년 아시아태평양지역 첨단망협회 학술위원회 위원장.

<주관심분야 : 네트워크 보안, 모바일 네트워크, 소프트웨어 재사용>

장 문 석(준회원)

- 2009년 2월 홍익대학교 컴퓨터 정보통신공학과 학사 졸업.
- 2007년~2009년 HUST 회원.

<주관심분야 : 네트워크 보안>

안 창 균(준회원)

- 2009년 2월 홍익대학교 컴퓨터 정보통신공학과 학사 졸업.
- 200년~2008년 HUST 회원.
- 2009년 2월~현재 (주)클리포드 사원

<주관심분야 : 네트워크 보안>

민 병 우(준회원)

- 2009년 2월 홍익대학교 컴퓨터 정보통신공학과 학사졸업.
- 2002년~2008년 HUST 회원.
- 2007년~2008년 HUST 회장.
- 2008년 12월~현재 케이티하이텔 사원.

<주관심분야 : 네트워크 보안>

양 동 혁(준회원)

- 2009년 2월 홍익대학교 컴퓨터 정보통신공학과 학사 졸업.
- 2006년~2008년 HUST 회원.
- 2009년 1월~현재 티맥스소프트 사원

<주관심분야 : 네트워크 보안>