취약성 DB 자동 갱신 로봇의 설계 및 구현

서혜성[0]*, 최경희*, 박승규*, 정기현**, 이철원***, 이남훈***, 한광택***
*아주대학교정보통신대학 정보및컴퓨터공학부, **아주대학교 전자공학부, ***국가보안기술연구소
(retry[0], khchoi, sparky, khchung)@madang.ajou.ac.kr, (cheolee, nhlee, kthan)@etri.re.kr

Design and Implementation of Robot for Updating Automatically Vulnerability Database

Hae-Sung Seo[0]*, Kyung-Hee Choi*, Seung-kyu Park*, Gi-Hyun Jung**, Cheol-Won Lee***, Nam-Hoon Lee***, Kwang-Taek Han***
* Graduate School of Information and Communication, Ajou University
** Division of Electronical & Electronics Engineering, Ajou University
***National Security Research Institute

요 약

컴퓨터 시스템이나 네트워크의 보안을 강화하는 방안으로 보안상 취약성이 발견되는 보안 취약성을 점검하는 것이 옳다. 그러나 취약성을 파악하기 위하여 국내외 관련 사이트를 수동적 방법으로 검사하는 것은 대단히 어려운 일이다. 따라서 스스로 관련 사이트의 홈 페이지를 검색하고 취약성 정보를 추출한 후 새로운 취약성 관련 정보가 발견되는 대로 이를 취약성 데이터베이스에 기록하는 이른바 취약성 자동 갱신 시스템[1]은 취약성 탐지 시스템의 핵심 기능이다. 본 논문에서 구현한 취약성 자동 갱신 로봇은 웹 페이지 자동 검색프로그램인 스파이더를 활용하여 구현되었으며, ICAT등과 같은 취약점 정보 제공 사이트들로부터 홈페이지를 검색하고 이에 수록된 정보를 수집 및 분석한 후, 취약성 데이터베이스를 자동으로 갱신한다.

1. 서론

취약성(security vulnerability)이란 호스트 컴퓨터나 네트워크에 존재하는 보안 상의 문제점을 지칭한다. H/W와 S/W 기술이 빠르게 발전하면서 취약성의 발견 속도나 그들에 대한 취약성 정보 및 대처방안의 발표 속도 역시 빨라지고 있다. 대부분의 취약성 정보들은 인터넷을 통하여 홈페이지의 형태로 온라인 매체에 의해 배포되고 있어 해당 관련 사이트의 홈페이지를 검색함으로써 취약성과 그들의 대처방안에 대하여 많은 정보의 획득이 가능하다.

그러나 배포 사이트에 따라 문서 형식에 차이가 있어서 각 사이트의 문서들을 수동적으로 수집하고 분석하고 취약성 database에 필요한 항목들을 추출하는 것은 현실적으로 어렵다. 또한 그 속성상 시급히 대처해야 하는 취약성의 특성으로 볼 때 취약성 데이터베이스에 필요한 항목들을 적절한 기준에 따라 자동으로 검색하고 갱신하는 로봇[2]은 취약성 탐지 시스템의 필수적 요소이다. 자동 갱신 로봇의 모듈은 각 사이트의 문서들을 수집, 분석하여 database에 필요한 항목들을 적절한 기준에 따라 갱신 시킨다.

취약성 갱신 로봇은 취약성 관련 웹 문서들이 취약성 관련 정보는 물론 취약성 데이터베이스에 불필요한 기타 정보도 많이 가지고 있기 때문에 이들로부터 필요한 핵심 정보만을 가려내는 기능이 필요하다. 이에, 본 논문에서는 이 모든 과정을 웹 로봇이 웹 문서를 수집하여 데이터베이스를 구성하는 과정과 유사하게, Acme[1] 에서 제공하는 웹 로봇인 java 패키지와 java용 구문 구조 분석기 생성기를 이용하여 제작하였다.

2절에서는 취약성 정보를 제공하는 주요 사이트와 Acme의 패키지, Jlex, java_cup에 대하여 살펴보며 3절에서 로봇의 설계와 구현을 기술하고, 4절에서 결론과 향후과제를 기술한다

2. 관련 연구

2.1 취약성 정보제공 사이트

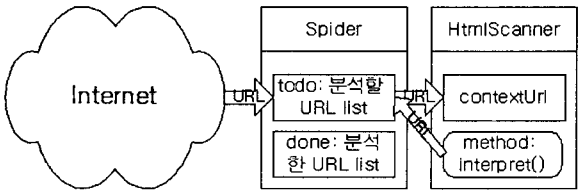
[1] ACME 공개 소프트웨어 제작을 주력으로 하는 연구소
http://www.acme.com

본 연구에서 취약성 정보를 자동으로 얻기 위하여 사이트 CVE(http://cve.mitre.org), ICAT (http://icat.nist.gov), CERT (http://www.cert.org), ISS X-Force(http://xforce.iss.net) 등을 사용한다. 취약성 정보를 제공하는 사이트들이 독자적인 이름 규칙에 의해 취약성 정보를 공개하면 같은 취약성이라 하더라도 정보의 효율적 공유가 어렵다. CVE는 이러한 제한 점을 극복하기 위해 취약성 정보 제공 사이트들이 취약성 이름에 대하여 표준을 제시한다. 갱신 로봇은 CVE ID를 기준으로 취약성 정보 수집, 작성하며 이것은 여러 사이트의 정보를 통합하는데 적절한 기준이 된다.

ICAT은 CVE가 제공하는 표준이름을 토대로 하여 프로그램상에서 분석이 용이한 형태의 원문을 제공한다. 원문은 각 취약성별로 취약성에 대한 설명, 유형, 안정성의 위험한 정도, 보다 자세한 정보를 얻을 수 있는 URL등 정보를 제공한다. ICAT에서 제공하는 URL을 적극적으로 활용할 경우 필요 없는 URL을 분석하지 않아도 되므로 갱신 로봇이 문서 분석에 소비하는 비용을 많이 절감할 수 있다. CERT, ISS X-Force는 각 취약성 정보와 대처방안을 발표하는 주요 사이트들이며 취약성으로 인해 나타나는 증상에 대한 설명이나 취약성으로 인해 시스템이 처할 수 있는 위험요소에 대한 분석과 취약성을 제거할 수 있는 방안을 제공한다.

2.2 Acme Spider 및 HTMLscanner

Acme의 패키지 중에 웹 로봇의 기능을 담당하는 주요 클래스로는 Spider 클래스와 HtmlScanner 클래스가 있다. Spider클래스는 처음에 주어진 URL을 HtmlScanner로 전송하여 새로운 URL 스트링을 감지하여 새 URL을 돌려 받음으로써 URL list를 작성하는 방식으로 웹을 탐색한다.[그림1]은 스파이더의 작동 구조를 보여주고 있다.



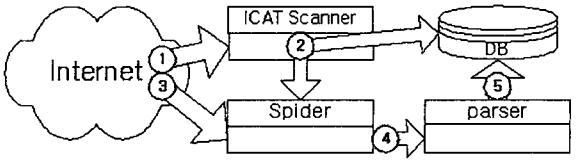
[그림 1] 스파이더 작동 구조

이 패키지는 단순히 URL을 수집하는 기능 이외에도 같은 URL을 중복되게 분석하지 않는 기능, 같은 페이지를 가리키는 URL일 경우 하나의 URL 주소값으로 URL의 표현양식을 고유하게 만드는 기능을 포함하고 있어 취약성 제공 웹 페이지를 손쉽게 수집할 수 있게 하는 장점이 있다.

2.3 Jlex, java_cup

Spider로 입수된 URL의 문서의 형태는 문서를 얻은 사이트에 따라 다르다. 문서 분석기는 문서 내에 포함된 html 태그와 텍스트를 분석하여 DB에 관련된 항목을 수집한다. 각각의 문서형식에 맞추어 일일이 문서 분석 클래스를 생성하는 것은 프로그래머에게 많은 노력이 필요하기 때문에 본 연구에서는 각 문서를 인식할 수 있는 문법구조를 정의하고 Java용으로 제작된 구문구조 분석기 생성기인 Jlex패키지[3]와 구문구조 파서 생성기인 java_cup패키지[4]를 이용하여 이 문법구조를 분석하는 클래스를 자동생성 하도록 하였다.

3. 자동 갱신 로봇의 설계 및 구현

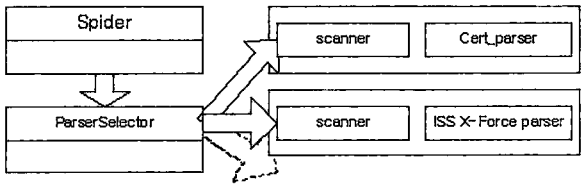


[그림 2] 갱신 로봇의 작동 구조

자동 갱신 로봇의 전체 구조는 [그림 2]와 같다. 갱신 주 목적은 취약성 관련 웹사이트에 액세스하여 취약성 DB에 누락된 새로운 취약성 관련 자료를 검색하여 취약성 DB에 추가하는 것이다. 이를 위해서 1차적으로는 취약성 DB와 가장 유사한 항목을 가지고 있는 ICAT의 문서를 읽은 후에 입력할 항목들을 DB에 저장하며 필요한 URL들을 수집한다. 그 후, 분석해야 할 취약성 대처방안 URL을 Spider객체에 저장한다. Spider객체 내에 저장된 URL들은 URL별로 정의된 분석기(parser)객체에게 넘겨지며, 분석기는 URL들을 분석하여 DB에 필요한 내용들을 추가로 추출하여 갱신 과정을 자동적으로 수행한다.

3.1 자동 검색 엔진

자동 검색은 Spider 클래스와 HtmlScanner클래스를 이용하여 구현하였다. Spider 클래스가 관련 URL을 검색하면, HtmlScanner클래스는 검색되어 넘어온 URL을 각 문서에 맞는 Parser클래스를 ParserSelector 이용하여 선택한다.



[그림 3] 문서 분석기 선택과정

본 논문에서 구현한 ParserSelector클래스는 문서를 제공하는 사이트의 주소와 해당 사이트의 문서를 분석할 수 있는 분석기이다. ParserSelector클래스의 분석기 내에 있는 선택 메소드는 URL에 나타난 각 사이트주소와 클래스 내에 등록된 사이트명을 비교하여 적절한 분석기를 선택하여 실행한다.

3.2 문서 분석기 설계

CERT 웹 문서는 [표1]에 예로 기술된 것과 같은 데이터들이 특정 tag와 문자열의 반복으로 이루어져 있다.

description	<H2>I. Description</H2> The loadmodule(8)..
impact	<H2>II. Impact</H2> Local users can gain ..
solution	<H2>III. Solution</H2> The CERT staff …

[표 1] CERT 웹문서의 구조

분석기는 토큰을 추출하는 scanner클래스와 추출된 토큰으로부터 문법적 구조를 감지하는 parser클래스로 구성되어 있다. scanner클래스는 모든 파서에서 사용할 있는 일반적인 토큰을 추출한다. [[표 2]는 scanner가 추출하는 토큰들의 리스트이다.

토큰	regular expression
STAG	<
ETAG	>
SHTWO	"<H2>"|"<h2>"
EHTWO	"</H2>"|"</h2>"
SHONE	"<H1>"|"<h1>"
EHONE	"</H1>"|"</h1>"
LETTERS	영문자,숫자,42개의 기호 및 특수문자
HTMLTAG	{STAG}({LETTER}|{SPACE}|{ENTER})+{ETAG}

[표 2] 토큰 리스트

3.3 구현 환경

DB 자동 갱신 로봇은 와우 리눅스 7.1에서 J2SE™ v1.3, Oracle 8.1.6환경에서 개발되었으며 [표 3]과 같은 패키지들이 설치되어 있어야 한다. 본 논문에서 구현한 자동 갱신 로봇은 아래 [표3]에 기술된 패키지를 포함하며, 필요한 CLASSPATH 등의 환경변수들은 쉘 스크립트내에서 설정토록 하였다. 프로그램의 실행은 make.sh과 run.sh의 스크립트를 실행함으로써 이루어 진다.

패키지	Ver	출처
Acme		http://www.acme.com
java_cup	0.10j	http://www.cs.princeton.edu /~appel/modern/java/CUP/
JLex	1.2.5	http://www.cs.princeton.edu /~appel/modern/java/JLex/
JDBC	2.0	http://java.sun.com/products/jdbc

[표 3] 로봇을 위해 필요한 JAVA 패키지

4. 결론 및 향후 과제

본 논문에서는 취약성에 대한 정보를 제공하는 웹 사이트로부터 스스로 문서를 수집 및 분석하여 취약성 DB를 자동 갱신하는 로봇을 설계하고 구현하였다. 로봇의 사용을 통하여 최신의 취약성 정보를 빠르고 정확하게 수집할 수 있게 되었다. 향후 과제로 갱신 로봇에 새로운 문서 분석기를 등록할 때 로봇의 소스 수정 없이 하도록 하는 것, 영어 이외의 언어로 작성된 웹 페이지 문서 형태도 파악하는 것, 일괄적으로 실행할 수 있도록 인터페이스를 사용하기 쉽게 강화하는 것 등을 위하여 연구를 계속하고 있다.

5. 참고 문헌

[1] 승현석, 정보검색의 세계, 에이전트를 잡아라, 월간 Internet 4월호
[2] *이강찬*, 로봇 에이전트와 그 활용, 월간 마이크로소프트웨어 10월호, 1996
[3] *Elliot Joel Berk and C. Scott Ananian*, Jlex, http://www.cs.princeton.edu/~appel/modern/java/JLex
[4] *Scott Hudson*, Frank Flannery, C. Scott Ananian, CUP, http://www.cs.princeton.edu/~appel/modern/java/CUP
[5] *Martijn Koster*, The Web Robots FAQ, http://www.robotstxt.org/wc/faq.html
[6] HTML 4.01, http://www.w3.org/TR/html4