정규표현식을 이용한 시스템 로그 분석

김홍경*, 이경현**
*부경대학교 일반대학원 정보보호학협동과정
khk009@kamco.or.kr, khrhee@pknu.ac.kr

An Analysis of System Log using Regular Expressions

Hong-Kyung Kim*, Kyung-Hyune Rhee**
*Dept. of Information Security, Pukyong-National University

요 약

보안업무를 수행하는 담당자로서 사이버 피해 여부를 파악하기 위한 가장 중요한 업무 중의 하나는 피해를 입은 시스템과 서비스에서 발생되는 다양한 로그들을 정확하게 분석하는 것이다. 그러나 해당 기관이 보안로그를 전문적으로 분석하는 SIEM(Security Information and Event Management)과 같은 솔루션이 없을 경우, 보안업무 담당자가 피해 시스템에서 추출된 로그만 가지고 직접 분석하여 공격여부를 판단하기는 쉽지 않다. 따라서 본 논문에서는 정규표현식을 이용하여 다양한 시스템의 로그를 쉽고 정확하게 분석하는 방법을 제시한다.

1. 서론

해커가 시스템의 취약점을 공격하기 위해서 가장 쉽게 접근하는 경로는 인터넷과 접점을 두고 있는 웹서버라 할 수 있다. 웹서버의 사이버 침해 여부를 확인하기 위해서는 웹서버의 로그파일을 정확하게 분석하여야 한다. 기업에서 운영하는 웹서버의 운영체제는 다양하지만, 본 연구에서는 현재 가장 많이 사용하고 있는 리눅스 운영체제 시스템과 Apache 웹서버에 대한 로그를 분석하고자 한다.

본 논문에서는 기업용 로그 분석솔루션이 없는 경우 보안업무 담당자가 직접 웹서버의 access_log 파일을 분석하는 과정에서 공격패턴을 정확하게 확인하기 위하여 정규표현식(Regular Expressions)을 이용하는 것에 목적이 있다.

이를 위하여 웹서버의 사이버 침해 여부를 확인하기 위한 웹서버의 로그파일을 선별하여 추출하고 그 로그파일의 내용에서 특정 문자열을 대상으로 조건을 지정하여 검색/치환/검사를 실행하는 방법[1]으로 정규표현식을 활용하여 웹서버 로그파일의 분석결과를 정확하게 추출하고자 한다.

2. 로그파일 분석 요구사항

2-1. 로그파일의 수집

본 논문에서는 시스템에서 생산되는 다양한 로그들이 있지만 해커의 침해위협에 대비하기 위하여 보안업무 담당자가 우선 수집하여 분석해야 할 로그 유형들을 <표 1>과 같이 몇 가지 선별하여 정의한다.

<표 1> 사이버 침해 분석에 필요한 로그 유형

로그파일명	저장데이터
access_log	웹서버 접속정보
error_log	웹서버 error정보
messages	콘솔 화면에 출력되는 메시지 정보
secure	사용자 인증에 관련된 정보

물론 <표 1> 이외의 로그 유형도 많이 있고 침해사고 유형에 따라 분석해야 하는 로그파일은 다를 수 있다.

2-2. 정규표현식 문법

보안업무 담당자는 침해사고가 발생한 시점의 로그파일을 수집하여 정규표현식을 통하여 분석하고자 할 경우 정규표현식의 문법규칙을 이해해야 한다.

그 이유는 로그유형마다 문자열 집합이 다를 수 있고 로그 내용에서 꼭 확인해야 하는 정보만 선택적으로 분류해서 재 정의하고 분석결과를 추출해야 할 필요도 있기 때문이다.

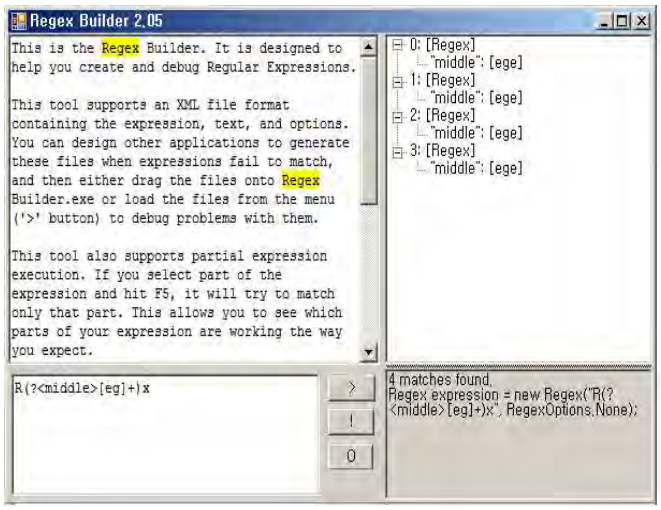
이를 위하여 사이버 침해 분석에 필요한 로그유형에서 선별한 로그파일을 분석하기 위한 Symbol[2]을 <표 2>와 같이 선별하여 정의한다.

<표 2> 정규표현식 Symbol 과 Meaning

Symbol	Meaning
₩d	숫자만 매치(문자·특수문자 제외)
₩w	영숫자 문자나 밑줄과 일치(특수문자 제외)
₩s	공백 문자와 일치
+	문자 하나 이상 찾기
*	문자가 없거나 하나 이상 연속문자 찾기
₩	문자열로 사용할 때 사용
()	찾은 값 중 원하는 값을 캡처
[]	대괄호 안의 문자를 or조건으로 찾음
^	문자열의 첫시작과 일치
$	문자열을 끝냄
(?<alias명>)	캡처한 값에 alias명을 부여
값?	? 앞에 지정한 값의 유무의 조건 지정
(?:)	괄호 안에 내용은 캡처 안함

2-3. 분석 프로그램

본 논문에서는 Apache 웹서버의 access_log를 정규표현식으로 분석하기 위한 프로그램인 Regex Builder[3]을 활용한다. Regex Builder 프로그램은 C#으로 작성된 WinForm 응용프로그램으로써 개발자가 정규표현식을 쉽고 빠르게 작성하고 테스트 할 수 있도록 돕는 프로그램이다.[4]



(그림 1) Regex Builder 프로그램

3. 로그파일 분석

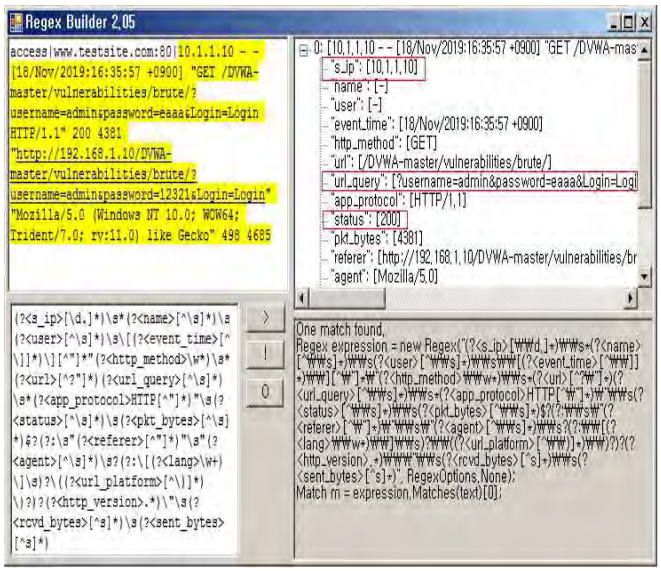
아래의 (그림 2)는 Apache 웹서버의 원본 access_log 에서 사이버 위협이 의심되는 로그내용만 추출하기 위하여 정규표현식 문법규칙을 적용하여 치환한 결과이다.

```
======원본로그======
access|www.testsite.com:80|10.1.1.10 - - [18/Nov/2019:16:35:57 +0900] "GET
/DVWA-master/vulnerabilities/brute/?username=admin&password=eaaa&Login=Login
HTTP/1.1" 200 4381
"http://192.168.1.10/DVWA-master/vulnerabilities/brute/?username=admin&password
=12321&Login=Login" "Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0)
like Gecko" 498 4685

======정규표현식======
(?<s_ip>[\d.]*)\s*(?<name>[^\s]*)\s(?<user>[^\s]*)\s\[(?<event_time>[^\]]*)\][^"]*"(?<ht
tp_method>\w*)\s*(?<url>[^?"]*)(?<url_query>[^\s]*)\s*(?<app_protocol>HTTP[^"]*)"\s(
?<status>[^\s]*)\s(?<pkt_bytes>[^\s]*)$?(?:\s"(?<referer>[^"]*)"\s"(?<agent>[^\s]*)\s?(?:
\[(?<lang>\w+)\]\s)?\((?<url_platform>[^\)]*)\)?)?(?<http_version>.*)\"\s(?<rcvd_bytes>
[^s]*)\s(?<sent_bytes>[^s]*)
```

(그림 2) Apache access_log 정규표현식 치환

아래의 (그림 3)은 Apache 웹서버의 원본 access_log 와 정규표현식을 Regex Builder 프로그램을 이용해서 분석한 결과를 나타낸다.



(그림 3) Regex Builder 프로그램 분석결과

위의 Regex Builder 분석 결과를 통해 해커의 ip 주소는 "s_ip":10.1.1.10, 공격은 "url_query"로 username=admin, password=eaaa를 전송하여 관리자 권한으로 로그인을 시도하였고 "status": 200을 볼 때 공격이 성공한 것으로 판단할 수 있다.

4. 결론

본 논문에서는 Apache 웹서버의 access_log 파일에서 일부분의 로그를 정규표현식 문법규칙을 활

용하여 피해를 정확하게 분석하였다. 하지만 개선된 방안으로, 시스템의 Log Format을 공통로그형식으로 설정한다면 표준 형식의 로그파일로 별도 생산할 수 있고, 웹 페이지 형태의 정규표현식 검사기 애플리케이션을 개발[5]한다면 표준 형식의 로그파일을 웹 페이지에 업로드 하여 로그파일 분석결과를 자동으로 얻을 수 있다.

참고문헌

[1] Young-Bo Kim, "Regular Expression with JavaScript", ITC, 2010, Page 1.

[2] Junghoo Cho and Sridhar Rajagopalan, "A Fast Regular Expression Indexing Engine", IEEE, San Jose in USA , 2002, Page 3.

[3] Regex Builder Program Download Path “http://www.sourceforge.net/projects/regexbuilder/”

[4] Dimitar and George Totkov, "Visual Parser Builder", RANLP , 2005, Page 4.

[5] Joonseon Ahn and Yeong-Min Kim and Jang-Wu Jo, "Development of a String Injection Vulnerability Analyzer for Web Application Programs", KIPS Transactions:PartA, Volume 15A Issue 3, 2008, Page 187