악성코드 사이트 탐지 분석기 구현

강동호, 최양서, 최병철, 한승완, 서동일

한국전자통신연구원, 사이버테러기술분석팀

Implementation of Malicious code site feeler analyzer

Dong-ho Kang, Yang-seo Choi, Byong-cheol Choi, Seung-wan Han, Dong-il Seo

Anti-Cyber Terror Team, ETRI

요 약

광대역 네트워크의 보급과 더불어, 많은 인터넷 사용자의 꾸준한 증가를 통해 국내의 인터넷 환경은 세계적으로도 상당한 규모가 되었다. 이러한 급격한 인터넷의 증가와 비례해서, 이의 결함을 이용한 해킹 건수가 급격히 증가하고 있다. 과거 서버환경에 대한 해킹시도에서 초고속 네트워크환경의 구성원인 개인 클라이언트 시스템에 대한 바이러스, 해킹 등을 이용한 불법적인 공격이 증가하고 있다. 또한, 상대적으로 보안에 무지한 일반 개인 사용자들을 위협하는 비정상적인 악성코드들이 쉽게 접촉할 수 있는 인터넷 환경에 유포되고 있고, 날로 전문화되고, 지능화된 악성코드 기법이 개발되고 있으나, 이에 대한 보안기술이 적절하게 대응하지 못하고 있는 실정이다. 따라서, 악성 코드 대응을 위해서 악성코드기법에 대한 계속적인 분석과 이에 적절한 대응이 필요하다.

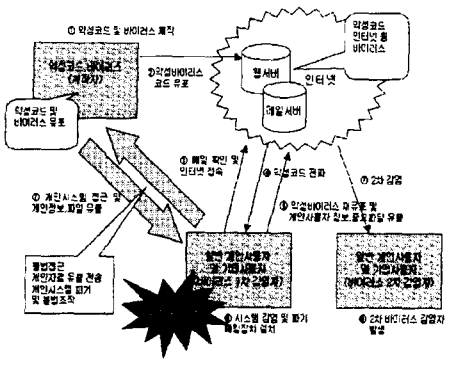
I. 서론

고속 정보화 산업의 육성과 인터넷의 급성장으로 인한 개방형 네트워크의 규모가 방대해지면서 이의 결함을 이용한 해킹 건수가 급격히 증가하고 있으며, 다양한 경로를 통해서 해킹, 바이러스를 이용한 비정상적인 코드들이 인터넷 상에서 유포되고 있다. 또한, 인터넷 표준 프로토콜인 HTTP의 약점을 이용하여 비정상적인 코드를 생성하여 이를 개인 정보의 획득, 악성코드의 실행 등 다양한 목적을 위하여 무분별하게 사용되고 있는 실정이다. 이러한 악성 코드들은 일반인들의 입장에서는 인지하거나 차단하기가 거의 불가능하기 때문에, 이를 탐지하고 차단 할 수 있는 보안 기술이 시급히 요구된다. 많은 보안 전문가들이 이러한 문제점을 끊임없이 제기하고 있으며, 이에 대한 문제점을 제품 벤더에서 보안하여 보안 패치나 어플리케이션 패치를 발표하고 있으나, 그 결함의 위험성이 지대함에도 불구하고 실정상 보안 및 패치 기간이 상당히 소요될 수 밖에 없다. 또한, 날로 전문화되고 지능화된 악성코드 기법이 개발되고 있으나, 이에 대한 보안기술이 적절하게 대응하지 못하고 있으므로, 전문적으로 이에 대한 현실을 파악하기 위한 대응책이 마땅히 연구되어야 할 것이다.

따라서, 우리는 인터넷상에 특정 웹사이트를 통해서 무분별하게 유포되고 있는 악성코드에 대한 탐지를 위해 악성코드 탐지 분석기를 제안하여 구현하였다. 본 논문의 구성은 다음과 같다. 2장에서는 악성 코드에 대한 간략한 소개와 악성코드의 일반적인 전파과정을 설명하고, 3장과 4장에서는 악성코드 탐지 분석기 에 대한 시스템 구조 및 사용자 인터페이스에 대해서 설명하고자 한다.

II. 악성 코드 소개

바이러스와 악성코드는 분류 및 대응 방법에 대해서 그 차이가 분명하다. 바이러스는 프로그램의 변경 및 손상 등, 시스템의 오동작을 유도하는 것이 주목적이지만, 악성코드의 주목적은 전파이다. 즉 "어떻게 악성코드가 만들어지고, 어떻게 치료해야 하는가" 에 대한 연구가 아닌 "어떠한 경로로 전파되고, 그 공격 유형이 어떤 형태를 띄는가" 를 더 중요하게 보고 있기 때문에, 그에 상응하는 분석방법을 가지고 있어야 한다. 또한, 한국정보보호진흥원에서 제공하는 바이러스 경향을 보면 바이러스 유형이 대부분 "I-Worm" 인 것을 알 수 있다. 이것은 악성코드와 바이러스를 결합한 공격방법을 의미한다. 즉 바이러스를 유포하기 위해 기존 방식을 채택하지 않고, 빠르게 전파시킬 수 있는 방법으로 발전하고 있다. 악성코드는 일반적으로 다음과 같은 경로로 전파된다.



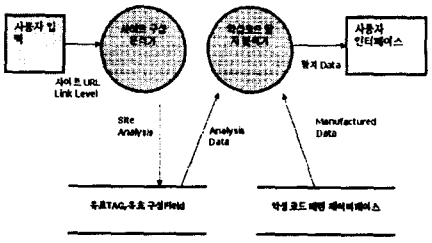
[그림 1] 악성코드 전파 과정 모의 구성도

먼저 제작자가 특정사이트에 악성 프로그램을 등록한다. 배포자는 해당 프로그램을 다운 받아 여러 경로로 배포 시킨다. 트로이 목마 파일이나 이메일의 첨부 파일로서 배포 될 수 있을 것이다. 아무런 의심 없이 해당 파일을 실행시킨 그 순간부터 악성 프로그램은 다른 파일을 감염 시키거나 시스템을 장악한다. 감염된 시스템의 사용자들이 이메일이나 파일 송신을 통해 파일을 배포하게 되고, 이 순환은 계속 반복된다. 악성 인터넷 컨텐츠를 포함한 웹 페이지를 방문하는 순간 사용자의 브라우저는 악성 코드를 다운 받게 된다. 이 코드는 바로 수행 되기도 하지만 미리 프로그램된 특정 이벤트가 발생할 때까지 수면 상태에 들어가기도 한다. 사용자는 자신의 시스템이 감염되었는지도 알지 못한 채 수백 수천의 복사본을 다른 사람에게 보내게 되는 것이다.

악성코드는 운영체제나 응용 프로그램에서 사용자의 편리를 위해 제공되는 기능을 악용하여, 사용자의 동의 없이 자동으로 실행되게끔 하는 코드를 의미한다. 따라서, 이러한 코드는 새로운 개발 언어의 등장과 함께 새롭게 재 작성된다. 악성코드는 변조가 쉽고, 그 종류가 알려지지 않은 코드등을 합치면, 일반 바이러스 수 보다 많은 것으로 보고 되고 있으며, 코드 자체가 매우 간결하고 그다지 놓은 전문 지식을 요하는 기술이 아니기 때문에 그 피해는 더욱 더 확산되고 있는 실정이다.

III. 악성사이트 탐지 분석기

다음 [그림 2]는 악성 사이트 탐지 분석기의 시스템 구조를 보여주고 있다.

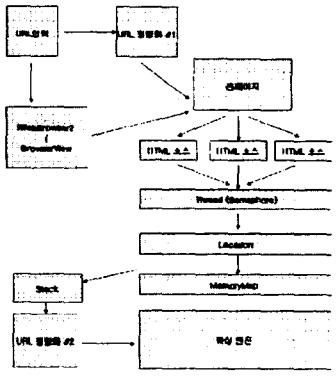


[그림 2] 악성코드 사이트 분석기 구조도

사용자가 방문하고자 하는 사이트 주소를 입력하면 사이트 구성 분석기에서는 URL을 정형화한 후, 해당 페이지에 대한 분석을 실시한다. 사이트 구성 분석기를 통해서 얻어진 유효 TAG, 유효구성 Field는 악성 코드 탐지 분석기 모듈로 넘겨지게 되며, 해당 데이터는 악성 코드 패턴 데이터베이스를 통해 악성 코드 유무를 출력하게 된다.

1. 사이트 구성 분석기

사이트 구성 분석기는 사용자가 정의한 레벨의 범위내에서 원하는 목적 사이트의 내용을 분석기가 설치된 시스템에 저장하여, 이를 악성코드 탐지 엔진에 활용하기 위한 모듈이다. 검색 사이트의 내용 중에서 HTML과 관련된 내용을 제작된 분석기를 기반으로 정형화된 유효 TAG를 분리해 낸다.



[그림 3] 사이트 구성 분석기 구조도

2. 악성코드 탐지 분석기

정규 표현식에 의해 검출된 유효필드 ,태그들을 입력받아, 패턴 데이터베이스와 구축된 악성코드 이진 트리와 비교 검색 후 조건에 일치하면, 악성코드 탐지를 알린다.

이진 트리는 일반적으로 사용하는 이진 트리 방식과는 달리, 우측노드에는 같은 레벨의 속성을 가진 필드들은 링크시키고, 본인의 속성과 관련 있는 유효속성 들은 좌측의 링크를 통해서 연결시키는 구조를 가진다.

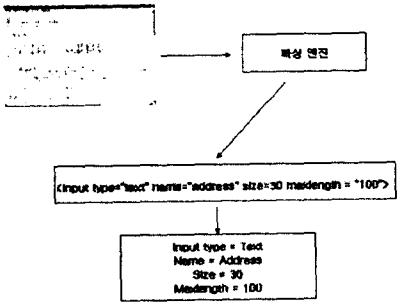
3. 악성코드 탐지 분석

본 악성 사이트 탐지 어플리케이션에서 악성코드를 탐지하기 위하여, 별도의 정규 라이브러리를 사용하였다. 이 라이브러리를 사용하는 목적은 악성코드의 복잡한 패턴을 매칭하기 위해서이다. 이것은 특정 태그를 검색하거나 추출하는 기능 등에 있어서 그 기능을 극대화 시킬 수 있다. 아래는 특정 태그를 검색하거나 추출을 위한 패턴 검색 일부분을 나타내고 있다.

```
<input type-" text" name-" address" size-30 maxlength -
"100" >

// input type을 파싱을 하기 위한 코드
<\s*(textarea)\s*[^>]+>[^<>]*(</(textarea)>)?
// Create our regular expression objectboost
::RegEx
e x p r (
(<\s*(textarea|input|select)\s*[^>]+>[^<>]*(</(select|textare
a)>)?)", TRUE);
```

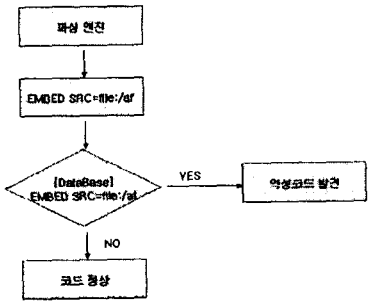
다음 [그림 4]는 악성 패턴 검색을 통해서 얻어지는 결과에 대한 예제도이다.



[그림 4] 악성 코드 패턴 검색 구조도

4. 악성코드 검출

악성코드를 분석하여 일정하게 정해진 패턴을 데이터 베이스화 한 자료와, 파싱 엔진으로부터 검출된 유효 태그와 매칭 시켜 악성코드 유무를 검출한다. 아래 [그림 5]는 악성코드를 검출하기 위한 흐름도이다.

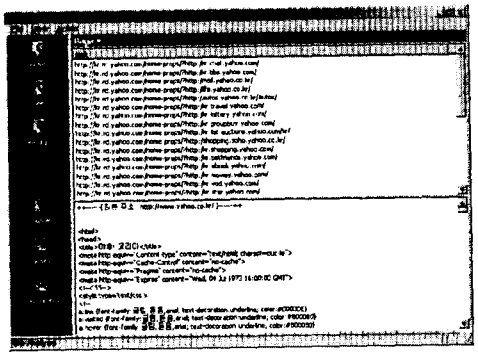


[그림 5] 악성코드 패턴 매칭 흐름도

IV. 악성 코드 탐지 분석기 사용자 인터페이스

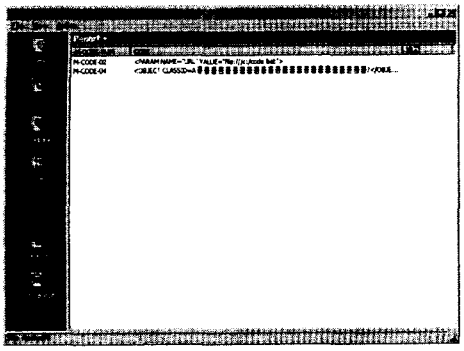
전체적인 인터페이스는 Dialog 기반으로 설계되었으며, 가장 기본적인 Browsing Viewer는 "MS 웹 브라우저" 컴포넌트를 사용하였다.

아래 그림은 전체 인터페이스 중 HTML 파일 정보 인터페이스를 나타내고 있다. 상위에 있는 URL 리스트는 해당 페이지에 링크되어 있는 주소를 파싱 엔진(Parsing Engine)을 통해 얻어진 결과 값이다. 하위 소스는 해당 페이지의 소스를 출력하는 부분이다.



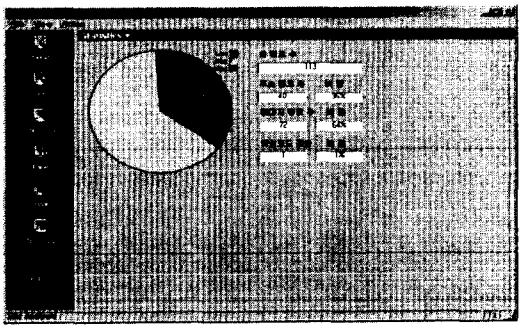
[그림 6] HTML 파일 정보 인터페이스

다음 [그림 7]은 분석 보고 인터페이스를 나타내고 있다. 다운로드 되어진 파일을 근거로, 악성코드가 내포되어 있을 유효 태그를 분리한 결과값과, 악성코드 데이터 베이스와 매칭시킨 결과값을 보여주는 기능을 제공하는데, 이때 악성 코드 데이터 베이스에 이름 지어진 악성 코드 네임과, 해당 코드 모두를 표시해 준다. M-CODE-NUM은 악성코드 데이터베이스에 이름 지어진 악성코드 이름 출력하고, CODE는 악성 코드 데이터베이스와 유효 태그와 매칭 시켜 악성코드로 판명된 유효 태그 출력한다.



[그림 7] 분석 보고 인터페이스

통계 인터페이스는 다운로드 되어진 파일의 주소 링크 수와, 이미지 링크 수, 그리고 악성 코드를 유발할 수 있는 코드의 총 비율을 Pie Chart 형식으로 보여 주는 기능을 제공한다.



[그림 8] 통계 인터페이스

V. 결론 및 향후 연구 방안

악성코드 탐지 분석기를 통해서 우리는 웹사이트에서 유포되는 악성코드를 검색하여 악성코드 유포를 차단하고, 특히, 악성코드 패턴을 데이터베이스화하여 지속적으로 변화하고 있는 악성 코드 패턴을 분석하고, 새로운 악성 코드 패턴을 발견시 데이터베이스를 갱신하여 새로운 악성코드 유형에도 대응하고자 하였다. 하지만, 악성코드는 다양한 형태로 수행된다. 또한, 수많은 스크립트 언어의 출현으로 인해 그 수가 기하급수적으로 증가하고 있는 실정이다. 악성코드 검출은 몇 가지 특정 패턴만을 이용하여 검출할 수 있는 성격이 아니기 때문에, 주기적으로 다양한 스크립트 언어에 대한 기술동향 및 보안 권고문, 사이트 등을 참고, 연구하여 패턴화 하는 것이 우선이며, 이것을 데이터 베이스화 하여 매칭, 검색하는 기술이 핵심이다. 향후 연구과제로는 본 논문에서는 자바 스크립트(Java Script) 및 XML 등의 패턴 및 파서(Parser)를 고려하지 않았다. 하지만 현재 자바스크립트를 이용한 악성코드 해킹 사례가 빈번하게 이루어지는 만큼, 이 분야와 관련한 자료 및 패턴 연구가 필요하다.

참고문헌

[1] 이상구외2, "악성코드 사이트 탐지 분석기 개발에 관한 연구", ETRI, 2002

[2] 한국정보보호센터, "웹 기반 악성 이동 코드 탐지 기술 개발", 1999

[3] http://www.cert.org/tech_tips/malicious_code_FAQ.html

[4] http://www.cert.org/advisories/CA-2000-02.html

[5] Jason Rafail, "Cross-Site Scripting Vulnerabilityes", CERT, 2001