리얼 브라우저 기반 웹 크롤러를 이용한 개선된 악성 웹사이트 탐지 기법

조호묵*, 이경석*, 최상용**, 김용민***

*카이스트 사이버보안연구센터, **한국폴리텍대학 서울강서캠퍼스 정보보안학과,
***전남대학교 전자상거래전공

An Improved Detecting Scheme of Malicious Website using Real Browser-based Web Crawler

Ho-Mook Cho*, Kyeong-Seok Lee*, Sang-Yong Choi**, Yong-Min Kim***

*KAIST Cyber Security Research Center, **Dept. of Information Security, Seoul Kangseo Campus, Korea Polytechnics, ***Dept. of Electronic Commerce, Chonnam Nat'l Univ.

요 약

최근 인터넷의 발전과 동시에 인터넷을 이용한 악성코드 유포는 가장 심각한 사이버 위협 중 하나이며, 탐지 우회 기법이 적용된 악성코드 유포 기술 또한 발전하고 있어 기존의 악성코드 유포 네트워크 분석 방법으로는 효과적으로 대응하지 못한다. 본 논문에서는 기존의 알려진 웹 크롤러 기반의 분석 및 탐지 도구보다 지능화된 악성코드 유포 네트워크를 효과적으로 분석 및 탐지할 수 있도록 리얼 웹 크롤러를 제안한다. 제안하는 방법은 기존의 웹 크롤러와 제안하는 리얼 웹 크롤러의 분석성능 측정 및 기능 구현 한계점을 비교하여 지능화된 악성코드 유포 네트워크를 효과적으로 분석 및 탐지할 수 있음을 실험하여 보였다.

I. 서론

최근 ICT 기술의 비약적인 발전으로 인해 인터넷을 이용한 실생활은 편리해진 반면, 인터넷 사용자들을 대상으로 하는 사이버 위협은 증가하였다. 이러한 사이버 위협 중 대표적인 위협의 하나로 Drive-by download 공격이 있다. 이 공격 기법은 사용자 모르게 악성코드를 유포하는 사이트로 유도하여 취약한 애플리케이션을 사용하는 컴퓨터에 악성코드를 감염시키는 공격으로 최근 몇 년간 인터넷상의 가장 심각한 위협으로 간주되고 있다[1].

Drive-by download 공격에 대응하기 위한 연구 방법은 크게 웹 페이지 정적분석, 실행기반 동적분석 등으로 분류할 수 있는데, 정적분석은 AV(Anti-Virus) 엔진과 같이 시그니처를 사용하거나, 이미 탐지되었던 데이터와의 유사도를 비교하여 공격에 대응하며, 시그니처 및 탐지 데이터의 지속적인 업데이트가 필요하다. 또한 유포 스크립트 난독화와 오탐율이 높다는 한계점이 있다. 동적분석은 가상머신이나 에뮬레이터를 이용하여 분석 대상 웹사이트에 접속 후 OS 시스템 수준의 변화를 분석하여 공격에 대응하며, 공격자가 분석환경을 알아낼 수 있고 행위 모니터링 프로세스를 우회할 수 있는 한계점이 있다[2,3].

본 논문에서는 이러한 제한사항을 극복하고 Drive-by download 기법을 이용하여 악성코드를 유포하는 악성 웹사이트를 탐지하기 위해 웹사이트의 콘텐츠 분석 및 링크 추출, 행위분석 에뮬레이팅이 필요한 웹 크롤러인 멀티레벨 에뮬레이션(MULEM, Multi-Level Emulation)과 본 논문에서 제안하는 링크 추출 및 에뮬레이팅이 필요하지 않는 리얼 웹 크롤러(RWC, Real Web Crawler)의 분석 성능과 기능을 비교하여 제안하는 방법이 악성 웹사이트를 효과적으로 분류할 수 있음을 검증하였다.

II. 관련연구

2.1 악성코드 유포 네트워크

공격자는 PC에 악성코드를 감염시키기 위해 수 개의 웹사이트를 논리적으로 연결시켜 악성코드 유포지로 자동으로 연결되는 네트워크를 만든다. 이를 MDN(Malware Distribution Network)라 한다[3]. 공격자는 악성코드 유포 네트워크를 만들 때 사용자의 행위 없이 자동으로 연결되도록 하기 위해 javascript, iframe 등을 이용한다. 또한, 분석을 어렵게 하기 위해 삽입한 링크 정보를 난독화 한다[4]. 이를 통해 취약한 PC가 MDN에 접속하는 것만으로 악성코드에 감염된다.

2.2 악성코드 유포 네트워크 분석 방법

악성코드 유포지 네트워크를 분석 및 탐지하기 위한 연구는 다양한 측면에서 이루어지고 있다. 대표적인 연구의 분야는 정적분석과 동적분석이 있다[2,3].

정적분석은 웹사이트에 포함된 비정상 콘텐츠를 분석하는 시그니처 기반 분석 방법과 웹사이트의 메타정보를 통계적으로 분석하여 과거에 탐지된 웹사이트의 메타정보와 비교하는 방법이 있다. 정적분석은 특징상 난독화된 콘텐츠를 복호화 해야 하는 과정이 필요하며, 콘텐츠를 직접 분석하기 때문에 분석속도가 빠를 수 있으나, 신규 난독화 방법에 효과적으로 대응하기에 한계가 있다.

동적분석은 가상머신 또는 에뮬레이터를 사용하여 분석 대상 웹사이트에 직접 방문한 후 OS 시스템 수준의 변화를 분석하는 방법이다. 동적분석은 실제와 유사한 분석환경을 사용하여 정적분석의 한계점인 난독화에 대한 복호화 과정이 필요 없다는 장점이 있으나 하드웨어 감지, 실행환경 감지, 동작행위 감지 등 분석환경을 회피하는 기술이 적용된 악성코드 유포 네트워크에 대한 효율적 대응에 한계가 있으며, 분석환경이 악성코드에 감염될 수 있는 위험성 등이 상존한다[3].

2.3 멀티레벨 에뮬레이션 주요기능 및 한계점

악성코드 유포 네트워크를 탐지하기 위한

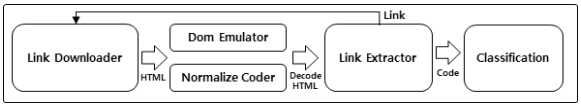

그림 1. MULEM 동작 구조

MULEM은 웹 사이트를 자동으로 탐색하여 악성코드 유포 네트워크를 탐지하는 소프트웨어로서 그림 1과 같은 구조로 동작한다[5].

Link Downloader는 Http request/response 메시지를 이용하여 분석 대상 링크의 HTML 코드를 다운로드하고, Normalize Coder와 Link Extractor에 의해 코드 내 MDN 구현을 위한 javascript, iframe 등 악성행위 관련 코드를 탐색한다. 난독화, 코드분할 등 분석회피 기술이 적용된 스크립트를 복호화 하기 위해 DOM Emulator에서 스크립트를 실행하고 MDN과 관련된 코드를 찾아 악성코드 유포 네트워크를 탐지하며, 악성코드 유포 네트워크 탐지를 위한 탐지 패턴에는 CVE(Common Vulnerabilities & Exposures) 패턴, Generic 패턴, 실행 파일 압축 패턴 등이 사용된다. 이 탐지 패턴은 지속적으로 업데이트가 되어야 신규 난독화 방법에 효과적으로 대응할 수 있으며, 난독화 및 분석이 필요한 스크립트가 많아질수록 분석 속도가 느려지는 한계가 있다. 따라서 본 논문에서는 탐지 패턴을 사용하지 않고 분석 속도가 유지되는 RWC를 제안하고자 한다.

III. 리얼 웹 크롤러

본 논문에서 제안한 RWC는 Real IE(Internet Explorer)를 이용하여 구현되며, 웹사이트에 접속하는 기능, javascript, iframe 등으로 자동으로 연결되는 웹 페이지 모니터링 기능, PE 파일 및 Process 모니터링 기능, IE 및 시스템 초기화 기능이 있다.

최근 연구에 따르면, 악성코드 유포 웹사이트는 TDS(Traffic Direction System)를 이용한 탐지 우회 기법을 통해 일반적인 웹 브라우징이 아닌 웹 크롤러나 가상환경에서 접속을 시도 할 경우 정상 사이트를 보여지는 거짓 응답을 보내 분석 및 탐지를 방해하고, 접속된 PC의 OS(Operation System) 및 브라우저의 버전과 취약점을 인지하여 감염률을 높이는 등 악

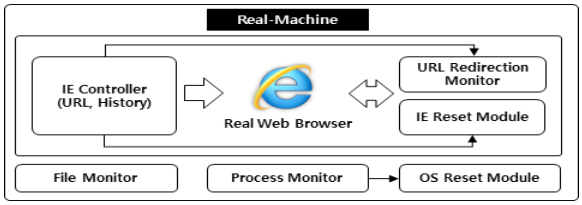

그림 2. RWC 시스템 구성도

성코드 유포 네트워크 기법이 지속적으로 지능화 되고 있다[4].

따라서 본 논문에서 제안하는 RWC는 악성코드 유포 네트워크에 쉽게 노출되도록 의도적으로 취약한 실제 윈도우 시스템과 보안 수준이 낮은 IE 환경에서 동작하며, RWC 시스템 구성도는 그림 2와 같다.

3.1 IE Controller & IE Reset Module

IE Controller는 윈도우즈 시스템에 내장되어 있는 Shdocvw.dll, Mshtml.dll 라이브러리 파일을 참조하여 구현되며, Shdocvw.dll 라이브러리는 웹 브라우징 시의 네비게이션이나 히스토리 등과 같은 기능을 제어하며, Mshtml.dll 라이브러리는 HTML 내용을 처리하는 기능을 제공하여 개발자 도구 수준에서 IE 제어가 가능하다.

IE Controller는 분석할 웹사이트 URL 목록과 접속 히스토리를 관리하며, 자동으로 이동되어지는 URL을 모니터링 할 수 있도록 URL Redirection Monitor에 인터페이스를 제공한다. 또한 웹페이지 접속 종료 상황을 인지하여 IE Reset Module에 신호를 보내 변경된 IE 설정을 초기화할 수 있도록 한다.

3.2 URL Redirection Monitor

공격자는 악성코드 유포 네트워크를 만들 때 사용자의 행위 없이 자동으로 연결되도록 javascript, iframe 등을 이용하여 웹 페이지를 논리적으로 연결하고 사용자 모르게 악성코드 유포지로 이동하여 PC가 감염되도록 한다. 따라서 URL Redirection Monitor는 사용자의 행위 없이 자동으로 웹 페이지의 이동을 모니터링 하여 악성코드 유포 네트워크를 탐지한다.

3.3 File Monitor

악성 웹사이트에서 악성코드 감염 절차는 사용자가 악성코드 유포 네트워크에 접속할 경우 IE 또는 OS 취약점을 이용하여 사용자 행위 없이 PC에 악성코드가 다운로드 되고 자동으로 실행되어 감염되는 것이 일반적이다. 따라서 File Monitor는 OS 수준에서 생성 및 변형되는 PE 파일을 모니터링 하여 URL Redirection Monitor와 복합적으로 악성코드 유포 네트워크를 탐지한다.

3.4 Process Monitor & OS Reset Module

분석환경 회피기술이 적용된 악성코드 유포 네트워크 분석 및 탐지하기 위해 의도적으로 보안을 취약하게 설정한 리얼 머신 기반으로 구성되기 때문에 악성코드 감염에 취약할 수밖에 없다. 따라서 Process Monitor에 의해 의심되는 프로세스가 감지되면 OS를 초기화한다.

IV. 실험 및 분석

4.1 실험 방법

실험은 악성코드 유포 네트워크 탐지를 위해 MULEM과 본 논문에서 제안하는 RWC의 분석 성능 및 기능을 비교하는 방법으로 실험하였다. 분석 성능을 측정하기 위해 동일한 사양의 리얼 머신에 각각 MULEM과 RWC를 설치하고 실제 웹 사이트를 대상으로 성능을 측정하였다. 실제 웹 사이트는 단순 텍스트와 링크 구조로 구현된 550개 링크로 구성된 웹사이트 A와 자바스크립트, 레이어 등 웹사이트 A보다 복잡한 4,890개 링크로 구성된 중·소규모의 웹사이트 B, 자바스크립트, 로그인, ActiveX 등 15,200개 링크로 구성된 대규모 웹사이트 C에 대해 분석 성능을 측정하였다. 이때 Drive-by download 공격에 대한 탐지 여부를 판별하기 위해 iframe을 이용하여 악성코드가 다운로드 및 실행되는 논리적으로 연결된 한 세트의 웹 페이지를 구현하고, 분석대상 URL 목록에 포함하여 Drive-by download 공격을 각각의 크롤러에서 탐지하는지 실험하였다. 또한 동일한 조건 하에서 분석 성능을 측정하기 위해 웹 사이트 C의 로그인 및 ActiveX 설치가 되어야 접근하는 웹 페이지에 대해서는 분석 대상에서 제외하였다.

4.2 실험 결과 및 분석

본 논문에서는 각각 550개, 4,890개, 15,200개 링크로 구성된 실제 웹 사이트를 대상으로

표 1. MULEM와 RWC 성능 및 기능 비교 (△ : 기능 추가 시 문제 해결)

구분	분석 능력	탐지 우회 기법			비개방형			스크립트 에뮬레이터	감염	크롤링 탐지/차단
		난독화	명령 분할	TDS	플러그인	로그인	ActiveX			
MULEM	저하	△	△	△	X	X	X	필요	X	○
RWC	유지	○	○	○	△	△	△	필요 없음	△	X

MULEM과 제안하는 RWC를 이용하여 분석 성능을 측정하였고, 기능 구현 한계점을 비교하여 제안하는 방법이 악성코드 유포 웹사이트를 효과적으로 분류할 수 있음을 검증하였다. MULEM과 제안하는 RWC의 성능 및 기능 비교는 표 1과 같다.

MULEM과 RWC의 성능을 측정한 결과를 살펴보면 그림 3과 같이 단순 웹 페이지인 웹사이트 A 경우 MULEM은 6분의 분석 시간이, RWC는 12분의 분석 시간이 소요되어 두 배의 성능 차이의 보였다. 반면 MULEM은 분석 요소가 증가되면 될수록 분석 성능이 떨어지는 것을 확인할 수 있다. 반면 제안한 RWC의 분석 성능은 일정하게 유지되는 것을 확인할 수 있고 javascript, flash 등 분석 요소가 많은 대규모 웹사이트 C의 경우에 MULEM 보다 빠른 분석 성능을 확인 할 수 있다.

기능을 비교한 결과를 살펴보면 스크립트 난독화 및 명령 분할, TDS 등 탐지 우회 기법을 적용한 악성코드 유포 웹사이트를 분석하기 위해서 MULEM은 분석 에뮬레이터 또는 모듈 추가하여야 하나 RWC는 어떠한 분석도구도 추가하지 않고 분석이 가능하다. 비개방형 웹페이지에 대해서는 MULEM은 접근이 불가능하지만 RWC는 사용자 행위 개입 기술을 추가하면 접근과 분석이 가능하다. MULEM에 반해 RWC는 실제 컴퓨팅 환경을 사용하기 때문에 쉽게 악성코드에 감염되는 문제점은 있으나 IE Reset 및 OS Reset Module에 의해 시스템 초기화에 필요한 시간 소요가 발생하지만 복구가 가능하다.

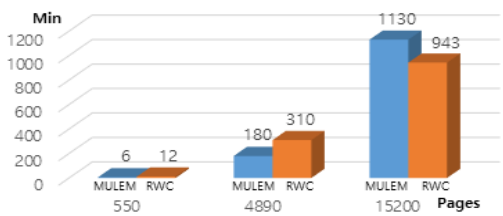

그림 3. 웹사이트 별 분석 성능 비교

V. 결론 및 향후 연구

악성코드 유포 웹사이트를 효과적으로 분석 및 탐지하기 위해서 MULEM은 지속적인 시그니처 및 탐지 데이터를 업데이트해 주어야 하며, 스크립트 분석을 위해 추가적인 에뮬레이터가 필요하고 비개방형 웹페이지에 대해 분석이 불가능 하다. 뿐만 아니라 스크립트가 많은 웹사이트를 분석할 경우 본 논문에서 제안한 RWC 보다 분석 성능이 떨어진다. 따라서 제안하는 RWC는 지속적으로 지능화되는 악성코드 유포 네트워크를 효과적으로 분석 및 탐지하기 위한 도구로 활용 될 수 있을 것이다. 향후 비개방형 웹페이지의 사용자 행위 개입 기술을 개발하고, IE 및 OS가 악성코드에 의해 감염되지 않도록 시스템 개선하는 연구를 지속할 계획이다.

[참고문헌]

[1] ENISA "Threat Landscape 2014", "https://www.enisa.europa.eu/publications/etl2015", 2015, Sep. 9.

[2] Shindo, Yasutaka, et al. "Lightweight Approach to Detect Drive-by Download Attacks Based on File Type Transition." Proceedings of the 2014 CoNEXT on Student Workshop. ACM, 2014.

[3] Egele, Manuel, et al. "A survey on automated dynamic malware-analysis techniques and tools." ACM Computing Surveys (CSUR) 44.2 (2012): 6.

[4] Kim, Byung-Ik, Chae-Tae Im, and Hyun-Chul Jung. "Suspicious malicious web site detection with strength analysis of a javascript obfuscation." International Journal of Advanced Science and Technology 26 (2011): 19-32.

[5] 최상용, 강익선 등, "악성코드 유포 네트워크 분석을 위한 멀티레벨 에뮬레이션", 정보보호학회논문지(Journal of The Korea Institute of Information Security & Cryptology), 23(6), pp.1121-1129, 2013년 12월.