악성코드 유포 탐지를 위한 웹사이트 변조 분석 방법

최상용°[1], 한기문[1], 김용민[2]

[1] 한국과학기술원°, [2] 전남대학교

A Method of Website Modification Analysis for Detection of Malware Distribution

Sang-Yong Choi°[1], Ki-Moon Han[1], Yong-Min Kim[2]

[1]Cyber Security Research Center, Korea Advanced Institute of Science and Technology

[2]Dept. of Electronic Commerce, Chonnam National University

csyong95@kaist.ac.kr[1], linuzen@kaist.ac.kr[1], ymkim@chonnam.ac.kr[2]

요 약

최근 3 20, 6 25 사이버 공격은 웹을 통한 악성코드 유포 및 해킹의 대표적인 사례이다. 공격자는 대량의 악성코드를 효과적으로 유포하기 위해 정상 웹사이트를 해킹하여 악성코드 유포지로 연결되는 경유지를 생성한다. 즉, 정상 웹사이트는 공격자에 의해 변조되어 악성코드 유포지로 악용된다. 최근 악성코드 유포지로 연결되는 비정상 웹사이트를 탐지하기 위한 다양한 기술들이 연구되고 있지만 이러한 기술은 대부분 점검대상 페이지에 포함된 악성링크를 발견하는데 초점이 있다. 하지만 공격자의 난독화, 지능형 스크립트 등 공격자의 다양한 탐지회피 방법에 능동적으로 대응하기에는 한계가 있다. 본 논문에서는 웹사이트 관리자가 악성코드 유포링크를 분석하지 않고 관리대상 웹사이트의 변조 여부를 실시간으로 확인할 수 있는 분석 방법을 제안한다. 이를 통해 허가되지 않은 변경의 발생을 관리자가 확인할 수 있게 함으로 궁극적으로 관리대상 웹사이트가 악성코드 유포지로 사용되지 않도록 예방할 수 있다.

1. 서론

인터넷 환경의 발전은 사용자에게 편리함을 준 반면 인터넷 사용자의 개인정보 유출 등 위협 또한 증가하고 있다. 최근 사이버 위협은 악성코드로부터 시작된다. 2013 년도에 발생한 3 20[1], 6 25 사이버테러[2]는 모두 사용자의 PC 에 악성코드가 감염되어 발생된 사례이다.

최근 이와 같은 위협을 탐지하여 능동적으로 대응하기 위한 악성코드 유포경유지 분석방법이 다양하게 연구되고 있다[3-5]. 하지만 기존의 연구는 악성코드 유포 스크립트 또는 웹사이트의 행위를 분석하는 것을 중점으로 하고 있다. 이는 분석이 완료된 악성행위에 대해서는 정확한 탐지가 가능하지만, 다양한 탐지회피 방법에 대한 정확한 분석이 이루어지지 않는다면 궁극적으로 악성코드 유포행위를 탐지할 수 없는 한계가 있다.

본 논문에서는 이러한 문제점을 해결하기 위해 보다 근본적으로 웹사이트 변조의 측면에서 접근방법을 제시하고자 한다. 악성코드 유포지와 경유지를 생성하기 위해 공격자가 공격을 하는 행위의 결과로 웹사이트는 변조된다. 관리자의 승인이 없이 변경된 페이지를 정확하게 찾을 수 있다면, 공격자의 다양한 공격방법에 대해서도 능동적으로 대응이 가능하다.

2. 관련연구

공격자는 악성코드를 유포하기 위해 악성코드 유포사이트를 생성한다. 이후, 보다 많은 사용자 PC 를 유포사이트로 유도하기 위해 정상적인 웹사이트를 해킹한 후 유포사이트로 유도되는 링크를 삽입한다[6]. 이 과정에서 공격자는 탐지시스템의 탐지를 회피하기 위해 스크립트를 난독화 한다. 스크립트 난독화에 사용되는 기술은 자바스크립트 함수를 사용하여 문자열을 변경하거나 유니코드의 사용, 압축패커의 사용[7] 등 단순한 패턴매칭으로는 탐지하기 어려운 기술을 사용한다. 또한 이러한 난독화 기법은 정상 웹사이트에서도 소스코드 은닉 등을 위해 사용되는 방법으로, 난독화되었다는 것만으로 악성코드 유포 코드로 판단하기에는 한계가 있다.

악성코드 유포행위 탐지를 위한 연구로는 HTML 페이지의 콘텐츠를 분석하는 정적 분석, 자바스크립트 에뮬레이션, 브라우저 에뮬레이션 등 Low interaction 방법, 가상머신을 사용한 High interaction 등을 포함한 동적 분석 방법이 있다[3-5]. 하지만 기존의 방법들은 공격자의 난독화 기법을 분석하고, 실제 발생되는 행위를 분석한 후 그 기법을 적용시키는 방법을 사용하여 새로운 기법에 대한 능동적 대응이 부족하다. 이와 같은 한계점을 해결하기 위해서는 스크립트 난독화 방법, 다운로드 행위 등에 상관없이

분석할 수 있는 방법이 필요하다.
기술한 공격행위의 또 다른 특징은 대부분 <*iframe*>, <*script*>등 페이지 내 새로운 태그가 삽입되는 형태이다[6]. 이는 정상적인 웹사이트의 구조가 변경되었음을 의미한다. 즉, 악성코드 유포 네트워크 생성을 위해 공격자가 수행하는 공격은 웹사이트 변조로 나타난다. 따라서 HTML 태그의 변화를 정확히 알 수 있다면 기술한 기존 분석방법의 한계점인 새로운 유포방법에 대한 대응을 제공할 수 있다.

3. 웹사이트 변조 탐지 방법

본 논문에서 제안하는 웹사이트 변조 탐지 방법은 그림 1 과 같다. 먼저 메인 페이지의 콘텐츠를 다운로드 한 후 페이지 내의 HTML 태그를 추출하고, DOM 을 생성한다. 다운로드 한 페이지의 이미지는 악성코드 유포지로의 링크가 삽입될 수 있기 때문에 이미지의 변경을 점검한 다음 페이지에 포함된 링크(Link)와 텍스트(Contents)를 추출한다. 추출된 링크는 내부링크인지 외부링크인지에 따라 외부링크인 경우는 방문하지 않는다. 하지만 내부 링크인 경우에는 추적을 위해 다운로드 대상으로 등록한다. DOM 트리를 생성한 후에는 이전 DOM 트리와 현재 DOM 트리가 변경되었는지를 비교한다. 비교를 위해 먼저 DOM 트리의 각 노드에 대해 자식노드를 포함한 해시값인 $HASH_{(노드\ N)}$을 생성하고, 이전 방문시의 해시값과 현재 해시값을 비교한다. $HASH_{(노드\ N)}$은 DOM 트리 내의 N 차수 이하 모든 노드의 해시가 저장되어 있다. 따라서, $HASH_{(노드\ N)}$의 값이 변경되었다면 노드 N 의 하위노드 중 일부가 변경된 것이라 판단하고, 노드 N 의 하위 노드를 점검한다. 이때, 변경되지 않은 노드와 하위노드는 분석하지 않는다.
제안하는 방법의 특징은 첫째, 웹사이트 내의 텍스트의 변화를 점검하지 않는다.

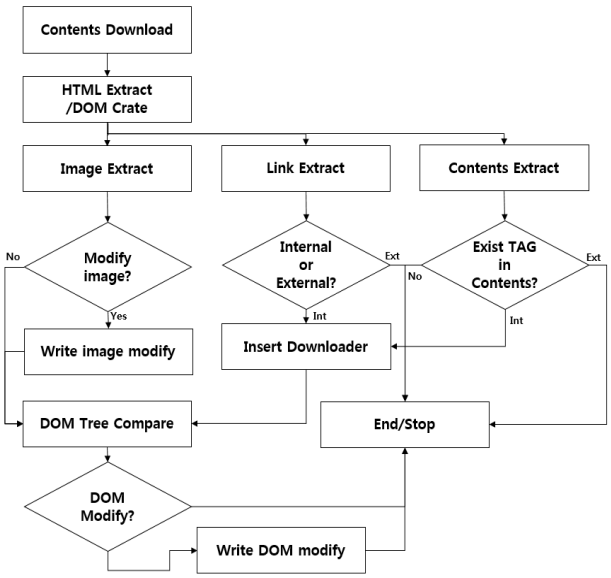

그림 1. 웹사이트 변조 탐지 절차

따라서 게시판의 게시물 변화와 같은 일반적인 변경에 영향을 받지 않는다. 다만, 존재하는 게시물 중 태그가 포함된 게시물에 대해서는 태그의 변경 여부를 점검한다. 즉 새로운 TAG 태그가 삽입되었을 경우 관리자로 하여금 확인할 수 있도록 정보를 제공한다. 둘째, HTML DOM 트리와 해시값을 이용하여 변경된 내용을 점검한다. 이와 같은 방법으로 빠른 시간에 변경된 웹사이트의 엘리멘트를 정확히 추적할 수 있어 악성코드를 유포하기 위한 공격자의 공격 결과인 정상 웹사이트에 삽입된 공격자의 유포지 연결 코드를 식별할 수 있다.

4. 결론

본 논문에서는 웹사이트 변조를 분석할 수 있는 방법을 제안하였다. 제안한 방법은 HTML DOM 구조와 크롤러를 이용하여 관리대상 웹사이트에 포함된 페이지를 능동적으로 추적하여 변조여부를 점검할 수 있다. 이를 통해 관리자에 의해 승인되지 않은 웹사이트 변경 여부를 통지함으로 관리대상 웹사이트가 공격자에 의해 악성코드 유포지 또는 유포지로 연결하기 위한 경유지로 사용되고 있는지 여부를 확인할 수 있다. 제안한 방법은 텍스트 분석 방법을 지양하여 게시물의 단순변화 등으로 발생될 가능성이 있는 오탐의 가능성을 줄였다. 다만, 제안한 방법은 공격자가 HTML DOM 구조를 변경하지 않고 기존 태그 내에 링크를 삽입하는 경우 탐지에 한계가 있을 수 있다. 향후 이에 대한 지속적인 연구와 속도를 높이기 위한 DOM 탐색 알고리즘 등의 지속적인 연구가 필요하다.

5. 참고문헌

[1] Graham Cluley, “DarkSeoul: Sophos-Labs identifies malware used in South Korean internet attack,” *http://nakedsecurity.sophos.com/2013/03/20/south-korea-cyber-attack/*, March, 2013.

[2] ASEC, “6.25 DDoS 공격에 사용된 악성코드 상세분석,” *http://asec.ahnlab.com/949*, June. 2013

[3] G. Wang, et al., “Detection and analysis of drive-by-download attacks and malicious Javascript code,” in *Proc. WWW*, pp. 281-290, Apr. 2010.

[4] N. Provos, et al., “The ghost in the browser: Analysis of web-based malware,” in *Proc. Hotbots*, pp.4-4, Apr. 2007.

[5] G. Wang, et al., “Detecting Malicious Landing Pages in Malware Distribution Networks,” in *Proc. IEEE DSN*, Jun. 2013..

[6] Julianne Pepitone, “NBC hack infects visitors in 'drive by' cyberattack,” *http:// money.cnn.com/2013/02/22/ technology/security/nbc-com-hacked-malware/index.html*, Feb. 2013.

[7] ASEC, “자바스크립트 난독화 이해하기,” *http:// www.ahnlab.com/kr/site/securityinfo/secunews/secuNewsView.do?menu_dist=2&seq=12272*, May. 2008