지능형 악성코드 분석을 위한 시나리오 기반의 수집 및 모니터링 플랫폼 구현

Implementation of a Scenario-based Collection and Monitoring Platform for Intelligent Malware Analysis

저자
(Authors)

조호묵, 이경석, 김용민
Ho-Mook Cho, Kyeong-Seok Lee, Young-Min Kim

출처
(Source)

한국정보과학회 학술발표논문집 , 2016.6, 1093-1095 (3 pages)

발행처
(Publisher)

한국정보과학회
KOREA INFORMATION SCIENCE SOCIETY

URL

http://www.dbpia.co.kr/Article/NODE07017747

APA Style

조호묵, 이경석, 김용민 (2016). 지능형 악성코드 분석을 위한 시나리오 기반의 수집 및 모니터링 플랫폼 구현. 한국정보과학회 학술발표논문집, 1093-1095.

지능형 악성코드 분석을 위한 시나리오 기반의 수집 및 모니터링 플랫폼 구현*

조호묵[O1], 이경석[1], 김용민[2]
[1]카이스트 사이버보안연구센터, [2]전남대학교 전자상거래전공
chmook79@kaist.ac.kr, harvist@kaist.ac.kr, ymkim@chonnam.ac.kr

Implementation of a Scenario-based Collection and Monitoring Platform for Intelligent Malware Analysis

Ho-Mook Cho[O1], Kyeong-Seok Lee[1], Young-Min Kim[2]
[1]KAIST Cyber Security Research Center, [2]Chonnam National University Dept. of Electronic Commerce

요 약

최근 인터넷의 발전과 동시에 개인정보 유출, DDoS, APT 공격 등 사이버 위협 또한 지속적으로 급증하고 있다. 이 사이버 위협의 중심에는 악성코드가 있으며, 악성코드는 사이버상의 가장 심각한 위협으로 분류된다. 뿐만 아니라 최근 악성코드는 분석을 회피하는 기술이 적용되어 효과적으로 대응하지 못한다. 본 논문에서는 기존의 알려진 수집 및 분석 환경보다 지능형 악성코드를 효과적으로 분석할 수 있도록 시나리오 기반의 행위 정보 수집 및 모니터링 플랫폼에 대한 구현 방안과 활용 방법을 제안한다.

1. 서 론

최근 인터넷의 비약적인 발전과 동시에 개인정보 유출, DDoS, APT 공격 등 사이버 위협도 지속적으로 급증하고 있다. 사이버 공격의 형태는 다양하지만 모든 공격의 중심에는 악성코드가 있으며, 악성코드는 사이버상의 가장 심각한 위협으로 분류된다. 특히 AV-TEST 연구소가 발행한 악성코드 유포 통계 자료에 따르면 최근 10년간 신규 악성코드가 지속적으로 증가되는 것을 알 수 있고, 최근 몇 년 사이 폭발적으로 증가되는 것을 볼 수 있다[1]. 뿐만 아니라 최근 악성코드는 분석을 어렵게 하기 위해 복잡하고 정교해 지고 있어 악성코드의 50% 이상이 안티 바이러스 제품에서 탐지되지 않아 위협적인 상태로 유지된다[2]. 이러한 지능형 악성코드를 효과적으로 대응하기 위해 다양한 분석 환경이 연구되고 있지만 분석을 어렵게 하는 기술 또한 지속적으로 정교해 지고 있어 폭발적으로 증가하는 신규 악성코드를 효과적으로 분석하기에 한계가 있다.

본 논문에서는 악성코드를 분석하는 전통적인 방법과 악성코드에 적용되어 분석을 어렵게 하는 기술을 소개하고, 지능형 악성코드를 효과적으로 분석을 위한 시나리오 기반의 수집 및 모니터링 플랫폼(SCMP : Scenario-based Collection and Monitoring Platform)의 구현 방안에 대해 설명하고 SCMP의 활용 방법에 대해 제안하고자 한다.

2. 관련 연구

2.1 악성코드 분석 기술

전통적으로 악성코드를 분석하는 방법은 크게 두 가지로 나눌 수 있다. 첫 번째는 디버거나 역공학(Reverse Engineering) 기법을 사용하여 악성코드를 직접 실행하지 않고 분석하는 정적 분석(Static Analysis)이다. 정적 분석은 실행 조건 없이 악성코드의 구조와 동작 특징을 분석 할 수 있는 장점이 있다. 반면 분석 기능을 자동화하기 어렵고, 실행 코드 암호화나 패킹(Packing) 등 은닉기술이 적용되어 분석에 상당한 시간과 노력이 필요하다[3,4]. 두 번째는 악성코드를 에뮬레이터나 가상머신 환경에서 실행시켜 악성행위 수행 여부를 분석하는 동적 분석(Dynamic Analysis)이다[5]. 동적 분석은 행위 정보를 기반으로 분석하기 때문에 신규 악성코드에 대한 탐지 가능성이 높고, 분석 기능의 자동화가 가능하여 분석 시간을 단축할 수 있으나, 악성코드가 동작하기 위한 특별한 환경이나 조건이 필요하고 악성코드가 분석 환경을 인지하여 회피가 가능하다는 단점이 있다[6].

최근 연구에 따르면, 신규 악성코드의 수와 종류는 지속적으로 증가하고 있는데, 그 이유는 기존의 악성코드와 자동 제작 도구를 이용하여 신종 및 변종을 쉽게 만들어 내기 때문이다. 이러한 악성코드의 감염에 의한 사용자 피해 역시 급속도로 증가하고 있어 이를 효과적으로 분석하기 위해 최근 동적 분석에 대한 연구가 중요한 이슈가 되고 있다[5].

2.2 분석환경 회피기술

폭발적으로 증가하는 새로운 악성코드를 효과적으로 분석하기 위해서는 가상 환경(Virtual Environment)을 이용한 동적 분석 방법이 정적분석 방법에 비해 많은 이점을 제공한다. 하지만 최근 지능화된 악성코드는 실행되는 환경을 인지하기 위해 분석환경 여부를 검사하여

* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발 사업의 일환으로 수행함(13-921-06-002).

분석을 위한 의도된 환경일 경우 악성행위를 동작 시키지 않거나 실행을 멈추는 등의 분석환경 회피기술이 적용되고 있어 효과적으로 악성코드를 식별하는데 한계가 나타나고 있다[7]. 악성코드에 적용된 분석환경을 인지하는 원리는 크게 4가지이다[5].

- 하드웨어 탐지: 가상머신의 디바이스는 쉽게 식별되는데, 대표적으로 VMWare의 네트워크 인터페이스가 정의된 "pcnet32"를 인지하는 것
- 실행환경 탐지: 실행되는 환경이 디버거와 같이 프로세스를 모니터링 할 수 있는 상태인지 확인하는 것
- 외부 어플리케이션: 악성코드가 실행되는 환경에 Process Monitor와 같이 알려진 모니터링 애플리케이션이 동작하고 있는지 확인하는 것
- 동작 행위: 특정 권한이 있는 명령어의 실행되는 시간 차이를 이용하여 악성코드의 초기 동작 시간을 지연시키는 것

3. SCMP 구현 및 활용

본 논문에서 제안한 SCMP는 Real IE(Internet Explorer)를 이용하여 악성코드를 수집하는 기능, Web Downloader를 이용하여 악성코드 수집하는 기능과 수집된 악성코드에 대해 시나리오 기반의 행위 수집 및 모니터링 기능이 있다.

최근 연구에 따르면, 약 80% 이상의 악성코드는 실행 압축, 난독화, Anti-VM 등의 분석을 어렵게 하는 기술이 사용되고 있어 기존의 알려진 탐지 기술 및 분석 환경으로는 효과적으로 악성코드를 식별하기에 한계가 있다[5]. SCMP의 기능별 시스템 구성도는 그림 1과 같다.

3.1 Real IE를 이용한 악성코드 수집 기능

최근 악성코드를 쉽고 빠르게 퍼뜨릴 수 있도록 취약한 웹 클라이언트를 이용한 드라이브 바이 다운로드(Drive-by download) 공격이 가장 널리 사용되고 있다[8]. 따라서 Real IE는 의도적으로 취약한 윈도우 시스템과 IE 보안 조건을 낮추어 드라이브 바이 다운로드 공격에 쉽게 노출될 수 있는 상태에서 미리 선정된 대상 웹 페이지에 접속하여 PE(Portable Executable) 파일을 수집한다. PE 파일 수집은 파일 모니터가 시스템을 실시간으로 감시하다가 웹 페이지 접속 후 PE 파일이 생성되면 즉시 악성 유무를 판단하기 위해 VirusTotal에 질의하고 악성일 경우 File Server의 PE 파일들과 중복 검사 후 파일을 저장한다. 이때 윈도우 시스템이 악성코드에 의해 감염되는 것을 방지하기 위해 프로세스 모니터에 의해 화이트 리스트(White List) 제외한 프로세스가 감지되면 시스템을 초기화 하는 작업을 수행한다.

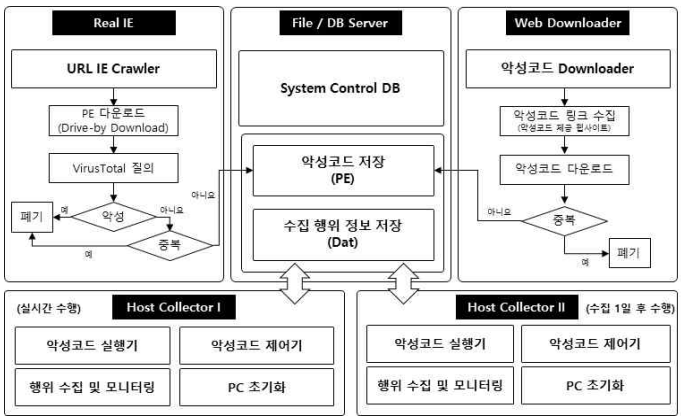

그림 1. SCMP 기능별 시스템 구성도

3.2 Web Downloader를 이용한 악성코드 수집 기능

악성코드를 효과적으로 분석 및 탐지하기 위해서는 다양한 악성코드의 행위정보의 데이터 셋 확보가 선행되어져야 한다. 따라서 Web Downloader는 다량의 악성코드를 모아 놓은 웹 사이트[9,10]에 접속하고 HTML 태그를 분석하여 악성코드를 다운로드 한다. HTML 태그를 분석할 때 악성코드를 실제 다운로드할 수 있는지 여부와 윈도우에서 실행되는 PE 파일 여부를 선행 검증하기 위해 악성코드의 4Byte를 Http Web Stream 방식으로 다운받아 검증한다. 다운이 가능하고 PE 파일일 경우 악성코드를 완전히 다운받아 File Server의 PE 파일들과 중복 검사 후 파일을 저장한다. 만약 악성코드 수집 대상 웹사이트에서 필요 이상으로 많은 악성코드가 실시간 업데이트되어 부하가 발생하면 해쉬 유사도 비교 툴인 SSDeep을 이용하여 최근 다운받은 악성코드와 유사도를 평가하여 선별적으로 악성코드를 다운받을 수 있다.

3.3 Host Collector를 이용한 악성코드 행위 수집 기능

악성코드 행위정보 수집을 위해 악성코드의 정상적인 실행과 완전한 동작이 선행되어져야 하지만, 최근 지능형 악성코드는 분석을 어렵게 하기 위해 분석환경 회피 기술을 대부분 적용한다. 따라서 Host Collector는 리얼 머신(Real Machine) 기반에 의도적으로 보안을 취약하게 설정한 윈도우 시스템을 이용하고 Process Monitor 또는 디버거 등의 알려진 모니터링 애플리케이션을 전혀 설치 않았다. 이는 악성코드에 적용된 분석 환경을 인지하는 원리 중 "하드웨어 탐지", "실행 환경 탐지", "외부 어플리케이션"에 대응이 가능하다. 뿐만 아니라 악성코드 제어기에 의해 실행되는 악성코드를 감시하다가 버튼 클릭, 인스톨, 스크롤 업·다운 등 사용자의 행위가 개입되어야 하는 상황을 자동으로 인지하여 적절한 사용자 행위를 삽입한다. 또한 표 1과 같이 악성코드가 좀 더 완

표 1. 악성코드 동작을 위한 사용자 행위

Event	User Interaction
E1	Click start button
E2	Click show desktop
E3	Click quick tray icon
E4	① Click start button ② Click start menu ③ Click all Programs button
E5	① Click start button ② Enter characters to run textbox
E6	① Click start button ② Enter characters to run textbox ③ Run windows utility package
E7	① Click start button ② Enter characters to run textbox ③ Run notepad ④ Enter characters to notepad

전한 동작을 위해 윈도우 사용자가 기본적으로 많이 행하는 패턴을 7가지로 분류하여 악성코드 실행 시 30초간 랜덤하게 5개의 사용자 행위를 삽입한다. 행위정보 수집 및 모니터링은 여섯 가지 기능으로 구성되어지는데 악성코드를 종합적으로 완전하게 분석하기 위한 기능으로써 첫 번째 레지스트리(Registry) 변화를 수집 및 모니터링 하여야 한다. 악성코드는 악성행위 및 자동 실행을 위해 레지스트리 값을 수정하는 등의 작업을 수행하기 때문이다. 두 번째는 파일 변화를 수집 및 모니터링 하여야 한다. 자신을 숨기거나 활동을 확대하기 위해 시스템의 임의의 위치에 자신을 복사하거나 파일을 삭제하기 때문이다. 세 번째는 프로세스(Process) 변화를 인데 새로운 프로세스를 생성하거나 시스템의 프로세스를 강제 종료시키는 작업을 수행한다. 네 번째는 악성코드 동작에 사용되는 메모리(Memory)로 자신을 탐지하지 못하도록 파일로 저장되는 것이 아니라 메모리에 상주하며 악성행위를 수행한다. 다섯 번째는 네트워크 변화로 악성행위 수행을 위해 C&C(Command & Control) 서버와 통신하거나 추가적인 악성코드를 다운받는다. 마지막으로 악성코드가 악성행위를 위해 호출하는 API를 수집 및 모니터링 하여야 한다. 이렇게 여섯 가지 수집 및 모니터링 기능은 악성코드의 행위정보를 수집하여 각 기능별 Dat 형식으로 File Server에 저장된다. 즉, 악성코드 하나당 6개의 Dat 파일이 생성된다. 악성코드에 의해 수집 Dat 파일이 감염 또는 손상을 최소화하기 위해 File Server의 파일 복사 스케줄러를 통해 Host Collector의 수집 행위가 종료되는 시점에 맞춰 별도의 저장 공간으로 파일을 옮긴다.

Host Collector는 실시간 수행 및 수집 1일 후 수행 형태로 나눠져 수행하는데 이는 타이머에 의해 동작하는 악성코드에 대해 대응하는 측면과 시간에 변화에 따라 달라지는 행위정보를 수집하기 위함이다. Host Collector의 수집 시기를 다양화하여 구축하면 타이머에 의해 동작하는 악성코드에 좀 더 효과적으로 대응할 수 있다.

악성코드 대다수가 실행 후 10초 이내에 악성행위를 위한 동작을 수행하지만 혹시 지연 시간을 두고 동작하는 악성코드 등의 행위정보 수집을 위해 Host Collector는 2분간 행위 수집 및 모니터링을 수행하고 1분간 PC를 재부팅하면서 감염된 PC를 초기화한다. 따라서 Host Collector는 하루 최대 480개의 악성코드 행위정보를 수집 및 모니터링 할 수 있다.

SCMP를 이용하여 최근 3개월 동안 수집된 악성코드 샘플과 악성 행위정보 수집 데이터 현황은 표2와 같다. 이 수집 결과는 지능형 악성코드를 효과적으로 분석 및 탐지하는데 중요한 판단 지표가 될 데이터 셋으로 활용될 수 있다.

표 2. SCMP를 이용한 악성코드 및 악성행위 수집 현황

(가) 악성코드 수집 현황 (단위 : 개)

구분		'16.1	'16.2	'16.3	합계
Real IE (대상 : 43만개)		29	18	24	71
Web Downloader	site1	30,642	7,590	1,944	40,176
	site2	32,860	1,592	4,251	38,703
합계		63,531	9,200	6,219	78,950

(나) 악성행위 수집 현황 (단위 : GB)

구분	'16.1	'16.2	'16.3	합계
Host Collector1 * 4	105.4	198.2	268.8	572.4
Host Collector2 * 4	108.5	123.0	167.5	399
합계	213.9	321.2	436.3	971.4

4. 결론 및 향후 연구

지능형 악성코드를 효과적으로 분석 및 탐지하기 위해서는 판단 지표가 될 다양한 악성코드 샘플과 악성 행위정보의 데이터 셋 확보가 선행되어져야 한다. 본 논문에서는 다양한 악성코드와 악성 행위정보를 보다 완전하게 확보할 수 있도록 SCMP를 제안하였다. 제안하는 SCMP는 지속적으로 증가하는 신규 지능형 악성코드에 효과적으로 대응할 수 있는 분석 지원 도구로 활용 될 수 있을 것이다. 향후 수집되는 악성 행위정보 데이터 양을 경량화 할 수 있도록 수집 및 모니터링 알고리즘을 개선하고, 제안한 SCMP가 수집한 악성 행위를 기반으로 악성코드를 분석 및 탐지할 수 있도록 분석 엔진을 개발하는 연구를 지속할 계획이다.

참 고 문 헌

[1] https://www.av-test.org/en/statistics/malware/

[2] European Union Agency for Network and Information Security (ENISA): ENISA Threat Landscape 2015, https://www.enisa.europa.eu/publications/etl2015, Dec 2015

[3] A. Moser, C. Krügel, and E. Kirda, "Exploring multiple execution paths for malware analysis," IEEE Security and Privacy, pp. 231-245, May. 2007.

[4] Nwokedi Idika and Aditya P. Mathur, "A Survey of Malware Detection Techniques," Department of Computer Science, Purdue University, Feb. 2007.

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

[6] V. Thomas and P Ramagopal, "The rise of autorun-based malware," McAfee, 2009.

[7] Raffetseder, T., Kru¨gel, C., and Kirda, E. Detecting system emulators. In 10th International Conference on Information Security (ISC). 1-18. 2007.

[8] 최상용, 강익선, 김대혁, 노봉남, 김용민, "악성코드 유포 네트워크 분석을 위한 멀티레벨 에뮬레이션", 정보보호학회지, VOL.23, NO.6, Dec. 2013.

[9] http://malshare.com/

[10] http://malc0de.com/database/