# API 기반 악성코드 행위 분류 기법에 대한 연구

A Study on the Method of Malware Behavior Classification based on API

저자
(Authors)

최보민, 강홍구, 이태진
Choi Bo Min, Kang Hong Koo, Lee Tae Jin

출처
(Source)

한국통신학회 학술대회논문집 , 2014.11, 512-513 (2 pages)
Proceedings of Symposium of the Korean Institute of communications and Information Sciences , 2014.11, 512-513 (2 pages)

발행처
(Publisher)

한국통신학회
Korea Institute Of Communication Sciences

URL

http://www.dbpia.co.kr/Article/NODE06078371

APA Style

최보민, 강홍구, 이태진 (2014). API 기반 악성코드 행위 분류 기법에 대한 연구. 한국통신학회 학술대회논문집, 512-513.

# API 기반 악성코드 행위 분류 기법에 대한 연구

최보민, 강홍구, 이태진
한국인터넷진흥원
bmchoi@kisa.or.kr, redball@kisa.or.kr, tjlee@kisa.or.kr

# A Study on the Method of Malware Behavior Classification based on API

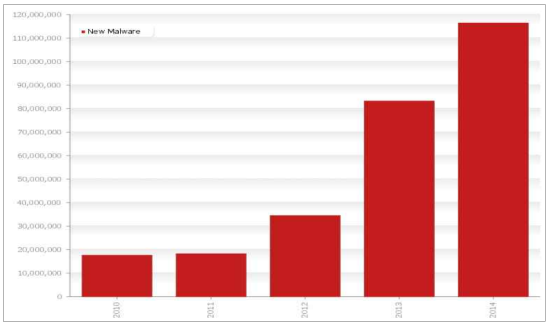
Choi Bo Min, Kang Hong Koo, Lee Tae Jin
Korea Internet & Security Agency

요 약

최근 악성코드 제작 도구의 이용법이 간편화됨에 따라 누구나 쉽게 변종 악성코드를 생성하고 유포할 수 있게 되었다. 이러한 변종 악성코드는 계속해서 급증하고 있어 백신 프로그램에서도 신속히 탐지 및 대응을 하고 있으나, 악성코드 행위에 따른 분류에는 한계가 있다. 이에 효과적 대응을 위한 악성코드 분류기술 연구들이 활발히 수행되고 있다. 본 논문에서는 악성코드가 호출하는 API를 수집하고 분석하여 악성코드 행위를 분류하는 기법을 제안한다. 제안하는 기법은 악성코드가 수행하는 행위를 코드화하고 동일한 코드별로 분류하기 때문에 입력되는 악성코드의 전수 분류가 가능하다는 장점이 있다. 실험을 통해 제안하는 분류 기법의 성능이 우수함을 보였다.

## Ⅰ. 서 론

최근 컴퓨터 및 인터넷의 보급이 정보 위/변조, 유출, 웜이나 바이러스 전파 등과 같은 보안 침해 부작용이 꾸준히 증가하고 있다. 특히, 악성코드 자동 제작 도구의 보편화로 인해 이들 중 대부분은 원형에 일부 기능을 변형하여 보다 정교하고 복잡한 형태인 변종 악성코드들이 손쉽게 제작되고 유포될 수 있게 되었다. (그림 1)은 AV-TEST에서 발표한 최근 5년간 악성코드 증가추이를 보여준다.[1]




〔그림 1〕 최근 5년간 악성코드 증가 추이

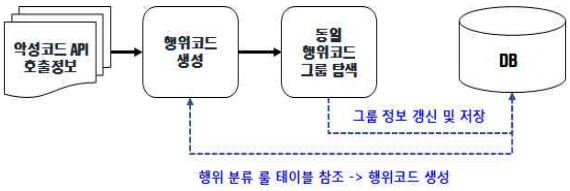
변종 악성코드는 원형만큼 강력한 보안 침해 부작용을 가지고 있어, 백신 프로그램에서도 신속히 탐지 및 대응을 하고 있으나, 악성코드 행위에 따른 분류에는 한계가 있다. 이에 대한 효과적 대응을 위한 악성코드 분류 기술 연구가 활발히 진행되고 있다.[2,3,4] 그러나, 기존의 시그니처 기반 악성코드 분류에 대한 연구들은 웜, 백도어, 루트킷 등의 복합적 행위를 포함한 변종 분류의 한계를 지니고 있다.

본 논문에서는 악성코드가 호출하는 API를 수집하고 분석하여 악성코드 행위를 분류하는 기법을 제안한다. 제안하는 기법은 악성코드가 수행하는 행위를 코드화하고 동일한 코드별로 분류하기 때문에 입력되는 악성코드의 전수 분류가 가능하다는 장점이 있다. 실험을 통해 제안하는 분류 기법의 성능이 우수함을 보였다.

## Ⅱ. 본론

윈도우즈 환경에서 악성코드는 대부분 실행 가능한 파일 형태를 가지고 있다. 악성코드의 행위에는 파일생성, 레지스트리변경, 프로세스생성, 네트워크 등 다양한 행위가 있으나, 이러한 행위는 모두 윈도우즈 시스템 내의 API를 통해 이루어지기 때문에 호출되는 API 분석으로 행위를 판별할 수 있다.

따라서, 본 논문에서는 이러한 API 정보를 기반으로 악성코드 행위 분류 기법을 제안한다. 제안하는 악성코드 행위 분류 기법은 악성코드가 호출하는 API 정보를 추출하는 API 추출 단계, API 추출 정보를 통해 행위별 코드를 조합하는 행위코드 생성 단계, 생성된 행위코드를 이용하여 행위를 기준으로 악성코드를 분류하는 악성코드 그룹핑 단계의 3단계로 수행된다. (그림 2)는 API를 기반으로 악성코드 행위를 분류하는 프로세스를 보여준다.




〔그림 2〕 악성코드 행위 분류 프로세스

(그림 2)의 API 추출 단계에서는 악성코드가 실행되는 과정에서 호출되는 API를 추출하기 위해 윈도우즈 XP/7 환경에서 사용자/커널레벨 API 후킹 모듈을 구현하였다. API 후킹 모듈에서는 수집되는 API들은 180개로 기존 악성코드 분석 보고서 및 샘플 분석을 통해 선정하였다. API 후킹 모듈은 가상머신에서 동작되며, 악성코드를 실행시킨 시간을 중심으로 시스템에서 수행되는 프로세스별 API 정보를 수집한다.

행위코드 생성 단계에서는 API 기반의 악성행위들을 분류하고 행위별 코드를 부여한다. (표 1)은 API 기반 주요 악성행위 리스트를 보여준다.

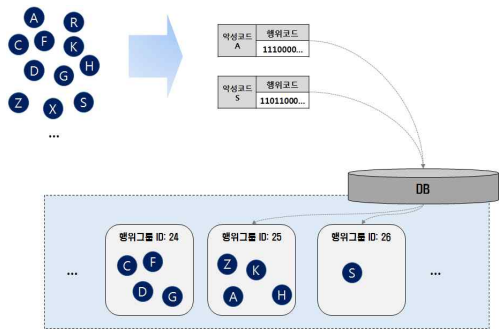
〔표 1〕 API 기반 주요 악성행위 리스트

| 악성행위 | 대상 API |
|---|---|
| 파일 | CreateFiie, ReadFile, WriteFile, CopyFile 등 |
| 다운로드 | URLDownloadToFileA, InternetOpenUrl 등 |
| Drop | FindResourceA, LoadResource 등 |
| 레지스트리 | RegCreateKey, RegOpenKey, RegSetValueEx 등 |
| 프로세스 | CreateProcess, FindProcess, TerminateProcess 등 |
| 네트워크 | WSAStartup, WSASend, Socket, Send, Recv 등 |
| DDos | CreateThread, Connect, listen, accept 등 |
| Time | GetSystemTime, GetLocalTime 등 |
| 디코딩 | GetModuleHandleA, LoadLibraryA 등 |
| 키로깅 | SetWindowsHook, CallNextHookEx 등 |
| 안티디버깅 | IsDebuggerPresent, CheckRemoteDebuggePresent 등 |
| 안티VM | DeviceIoControl, Strstr, CloseHandle 등 |

악성코드에서 추출된 API 정보를 기준으로 악성행위 리스트에 포함되는지 여부에 따라 악성코드의 행위 식별 및 분류를 위한 비트코드(1 or 0 조합)를 생성한다. 악성행위 리스트의 순서는 행위코드의 자리수를 지정하고, 악성행위 리스트에 있는 API 리스트에 추출된 API가 포함되는지 여부는 행위코드 값을 지정한다. 행위코드 값은 추출된 API가 포함되면 1, 그렇지 않은 경우에는 0을 부여한다. 행위코드는 자체로도 악성코드별 행위 정보를 담을 수 있고 유사 행위를 갖는 악성코드 그룹을 식별하는 코드로 사용된다. (그림 3)은 악성코드 행위코드 생성 예제를 보여준다.


〔그림 3〕 악성코드 행위코드 생성 예제

행위코드를 부여 받은 악성코드들은 이를 기준으로 행위가 유사한 악성코드들로 분류하기 위한 그룹핑 작업이 수행된다. 즉, 동일 행위코드를 가진 악성코드는 하나의 그룹으로 그룹핑되고, 이들은 상호 유사한 행위를 갖는 악성코드로 분류된다. (그림 4)는 행위코드를 이용하여 악성코드를 분류하는 예제를 보여준다.




〔그림 4〕 행위코드를 이용한 악성코드 분류 예제

실험에서는 악성코드 자동 생성도구로부터 생성된 악성코드 샘플을 대상으로 백신 프로그램의 진단결과와 제안하는 기법의 분류결과를 비교하였다. 악성코드 샘플은 악성코드 자동 생성도구인 Necro를 이용하여 자동실행 기능을 갖는 것과 갖지 않는 악성 파일을 생성하였다. 백신 프로그램은 Kaspersky를 이용하였다. (표 2)는 백신 프로그램 진단결과와 제안하는 기법의 분류결과에 대한 비교 내용을 보여준다. (표 2)와 같이 백신프로그램은 행위가 서로 다른 악성코드의 진단명이 같지만, 제안하는 기법은 행위에 따라 악성코드를 분류할 수 있었다.

〔표 2〕 백신 프로그램 진단결과 및 제안 기법의 분류결과 비교

| 파일해시 (SHA256) | 백신진단결과 ((Kaspersky)) | 제안하는 기법 | |
|---|---|---|---|
| | | 행위코드 | 주요행위 |
| 12b9f55.. | Trojan-Dropper. Win32.Dorn | 0110000... | 자동실행, 파일생성, ... |
| 6a97980.. | | 1110000... | 파일생성, ... |

## Ⅲ. 결론

본 논문에서는 API 기반 악성코드 행위 분류 기법을 제안하였다. 제안하는 기법은 악성코드가 호출하는 API 정보를 추출하여 악성행위를 식별하고, 이를 코드화하여 유사 행위를 갖는 악성코드를 분류할 수 있다. 특히, 행위코드가 동일한지 여부에 따라 분류되기 때문에 입력되는 악성코드의 전수 분류가 가능하다는 장점이 있다. 실험을 통해 제안하는 분류 기법의 성능이 우수함을 보였다. 향후 API가 갖는 파라미터의 속성까지 고려한 행위 분류 기법의 확장을 통해 분류 정확도를 높일 수 있는 연구를 진행할 계획이다.

## ACKNOWLEDGMENT

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [10044938 , 악성코드 프로파일링 및 대용량 보안이벤트 분석을 통한 공격징후 탐지기술 개발]

## 참 고 문 헌

[1] AV-Test, http://www.av-test.org

[2] H.K. Kang, S.G. Ji, and H.C. Jeong, "A Development of Management System for Malware Groups and Variants," ITS 2011, pp. 72-75, 2011.

[3] 강홍구, 조혜선, 김병익, 이태진, 박해룡, "API 유사도 분석을 통한 악성코드 분류 기법 연구," 한국정보처리학회 추계학술발표대회 논문집, 제20권, 제2호, Nov. 2013.

[4] H.K. Kang, J.S. Kim, B.I. Kim, and H.C. Jeong, "Development of an Automatic Document Malware Analysis System," ICITCS 2012, Vol.1, pp. 3-12, 2013.