악성코드 자동 분석 시스템의 결과를 이용한 악성코드 분류 및 분석

나재찬, 조영훈, 윤종희
영남대학교 컴퓨터공학과
e-mail : youn@yu.ac.kr

Malware Classification and Analysis of Automated Malware Analysis System

Jaechan Na, Jonghee M. Youn
Dept. of Computer Engineering, Yeungnam University

요 약

쿠쿠 샌드박스(Cuckoo Sandbox)는 가상머신을 이용해 악성코드를 자동으로 동적 분석할 수 있는 도구이다. 우선 악성코드의 MD5값을 이용하여 VirusTotal을 이용해 종류를 분류하고, 쿠쿠 샌드박스로 악성코드 동적을 분석하여 결과파일을 이용해 악성코드에서 호출한 API들에 대한 정보를 추출하고, 다양한 종류별 악성코드 그룹에 대해서 API빈도를 종합하고, 또한 다른 종류군의 악성코드 그룹과 API 빈도를 비교해 특정 종류의 악성코드 그룹에 대한 특징적인 API를 찾아내어 향후 이런 특징 API들을 이용해 악성코드의 종류를 자동으로 판정하기 위한 방법을 제시한다.

1. 서론

쿠쿠 샌드박스(Cuckoo Sandbox)[1]의 시스템은 VMware[5], VirtualBox[3], KVM[4] 등과 같은 가상머신 소프트웨어를 이용하여 악성코드의 악성행위를 동적 분석한다. 가상머신 안에서 악성코드가 동작하였던 내용들을 쿠쿠 샌드박스로 정보를 전송하고, 그 정보로 악성코드의 행위에 대해서 분석할 수 있었다. 분석된 정보는 “report.json”으로 저장이 되고, 그 정보를 이용하면 악성코드가 어떠한 API를 호출하여 동작하는지 알 수 있다. 본 논문은 많은 악성코드를 이용해 특정 종류의 악성코드가 주로 실행하는 API를 찾아보고자 한다. 최종적으로 직접 분석한 API를 확인해보고, 결론으로 끝을 맺는다.

2. 악성코드의 분류

우선 악성코드를 분류하기 위해서는 현재 프로그램의 MD5값을 알아야 한다. 쿠쿠 샌드박스의 결과파일 (report.json)에 포함되어 있으며, MD5값을 VirusTotal[2]에 입력하게 되면 현재 파일의 대략적인 분류를 알 수 있다. 총 180개의 악성코드 파일을 이용하여 분류 하였으며, 그 종류는 Backdoor, Dialer, Downloader, Fropper, Injecter, Rootkit, Spyware, Trojan, Worm 등으로 분류하였다. 분류는 각 안티바이러스 프로그램마다 결과가 달라서, 공통적으로 가장 많이 선택 된 것으로 하였다. 이제 이 분류를 바탕으로 종류별로 가장 많이 호출하는 API를 찾아보고자 한다. 단, 본 논문에서는 지문 관계상 Backdoor와 Downloader에 대해서만 분석하였다.

3. 악성코드의 API 분석

분류된 악성코드를 바탕으로 쿠쿠 샌드박스를 이용하여 악성코드를 분석하였다. 우선은 분류된 종류별로 API호출 횟수를 모두 더해 표1과 같이 각 종류별로 정리하였다.

LdrGetProcedureAddress	4110
RegOpenKeyExW	2301
RegCloseKey	2201
RegQueryValueExW	1813
GetSystemMetrics	978
RegQueryValueExA	938
~	~
~	~
gethostbyname	2
setsockopt	2
OpenSCManagerW	2

표1. Backdoor Total API

위와 같이 다양한 종류의 악성코드 그룹에 대해서 API 리스트를 추출하고 그 호출 횟수에 대한 정보를 생성해 특징 API의 결과를 분석 하였다. 또한 이렇게 추출된 각 그룹별 API 리스트와 호출정보를 활용해서 각 그룹별 상대적 API 호출 횟수(RC)를 구하였다. 상대적 API 호출 횟수는 특정 종류에서는 자주 나오나 다른 종류에서는 자주 나타나지 않는 API들에 대해서 높은 점수를 부여하고 다른 종류에서 흔히 발견되면 음수값을 가지고 같거나 비슷

하게 호출되면 0에 가까운 값을 가지도록 설계하였다. 그 값을 구하는 식은 아래와 같다.

RC = MP - {(SP(1)/N) + (SP(2)/N)+...+(SP(N)/N)}
RC : API 별 상대적 호출 횟수
MP : 특정 종류의 분석 할 악성코드의 API 호출 수
SP : 나머지 종류의 악성코드의 API호출 수
N : 자신을 제외한 나머지 종류의 개수

위와 같은 방법을 사용해 Backdoor와 Downloader에 종류 악성코드에서 API호출수를 확인 해 보면, 표2와 표3과 같이 나타난다. 그리고 우리는 이렇게 호출 상대적 호출 횟수가 상위 10위안에 드는 것들을 각 종류별 특징 API로 선택하였다.

표2. Backdoor API TOP10

NtDeviceIoControlFile	11
VirtualProtectEx	6.05
WriteConsoleW	6
RegSetValueExA	5.95
RegCreateKeyExA	5.25
DeviceIoControl	1.2
bind	0.9
socket	0.9
connect	0.55
closesocket	0.45

표3. Downloader API TOP10

NtWriteFile	66.1
NtReadFile	28.5
RegEnumValueW	21.95
LdrGetProcedureAddress	21.5
RegQueryValueExA	16.05
RegCloseKey	13.8
RegEnumValueA	9.5
RegEnumKeyExW	7.85
VirtualProtectEx	5.15
select	3.4

4. API결과와 분석

악성코드의 종류에는 여러 가지가 있지만, 우선은 위에서 제시한 Backdoor와 Downloader를 기준으로 비교해 본다. Backdoor의 경우 NtDeviceIoControlFile를 이용해 후킹하고, RegSetValueExA와 RegCreateKeyExA를 이용해 레지스트리를 설정하고 생성한다. Downloader의 경우 NtWriteFile를 이용하여 파일에 데이터를 기록한다. 그리고 RegEnumValueW를 이용해 지정한 키가 가지고 있는 모든 값의 이름들을 가져오는 등의 행동을 한다. 간단한 방법이지만 위와 같은 방법으로 추출한 API의 결과가 실제 역할과 비슷하다는 것을 알 수 있다.

5. 결론

본 논문에서 제안한 방법으로 악성코드를 분석하면 가장 자주 호출되는 API를 추출 하고, 이것을 통해 악성코드의 종류를 분류할 수 있게 한다. 지금은 200개정도의 데이터를 사용하여 빈도를 측정하였지만, 보다 많은 데이터를 사용한다면 좀 더 정확하게 종류를 구분 지을 수 있을 것으로 기대된다.

참고문헌

[1] Cuckoo Sandbox, http://www.cuckoosandbox.org
[2] Virus Total, http://www.virustotal.com
[3] Virtual Box, http://www.virtualbox.org
[4] KVM, http://www.linux-kvm.org
[5] Vmware, http://www.vmware.com