Anti-Virus 오진 방지를 위한 정상파일 취합 및 판단기준에 관한 연구

허종오*, 안형봉, 차인환
*안철수 연구소, 조지아텍 정보보안연구소
e-mail : maha96@naver.com, shead620@gmail.com, inhwan.cha@inta.gatech.edu

A Study on Collecting and Determining Criteria of Normal Files for Prevention of Anti-Virus False Positive

Jong-Oh Hur*, Hyung-Bong Ahn, Inhwan Cha
*AhnLab Corporation. Georgia Institute of Technology.

요 약

급격히 증가하는 악성코드로 인해 Anti-Virus 연구소들은 대량의 악성코드를 자동으로 분석 및 진단해야 할 필요성을 느끼게 되었다. 이러한 자동 분석은 오진(False Positive)의 증가라는 부작용을 가져왔다. 주요 파일에 대한 오진단은 Anti-Virus 를 통해 최종 보안을 유지하고 있는 대다수의 시스템을 정지시켜, 국가적인 손실을 유발할 수 있다. 따라서, 관련업체들은 오진을 방지하기 위해 다양한 연구를 진행하고 있다. 하지만, 정상파일을 보유하지 않고 파일정보만으로 오진여부를 판단하는 방법은 근본적으로 오진단을 감소시키는데 한계를 가지고 있다. 따라서, 최근에는 정상파일 셋을 구축하여 Anti-Virus 의 오진여부를 최종적으로 판단하는 연구가 진행되고 있다. 하지만, 수많은 파일 중에서 수집이 필요한 파일에 대한 기준과 수집한 파일이 정상파일 인지를 판단할 기준이 존재하지 않아, Anti-Virus 연구소들은 정상파일 셋 구축에 어려움을 겪고 있다. 따라서, 본 고에서는 정상파일에 대한 개념과 정상파일이 악성코드 파일과 비교되는 7 대 특성을 개발하고, 이를 토대로 기존에 제시된 적이 없는 정상파일 취합기준과 판단기준을 제안하였다. Anti-Virus 연구소는 본 기준들을 통해 우수한 오진방지용 정상파일 셋을 구축하여, 악성코드의 증가로 인해 함께 증가하고 있는 오진을 줄이고, 사용자 시스템의 안정성을 도모함으로써, 오진으로 인한 국가적, 경제적 손실을 방지하는데 큰 기여를 할 것으로 기대된다.

1. 서론

최근에 악성코드 툴을 이용하여 제작된 웜과 트로이목마들이 인터넷을 통해 유포되면서, 악성코드의 수는 비약적으로 증가하였다. 이러한 악성코드의 급격한 증가로 Anti-Virus 의 악성코드 진단수도 급격히 증가하게 되었다. 허니팟, 허니넷 등을 이용한 Anti-Virus 연구소의 악성코드 수집능력도 급성장하여 Anti-Virus 진단수는 지속적으로 증가할 것으로 예상된다.[1-4] 이로 인해 사람이 직접 악성코드를 분석하는 수작업으로는 급속히 증가하는 악성코드에 대해 100% 진단은 요원하게 되었다. 따라서, 최근에는 악성코드를 시스템이 자동으로 분석하여 진단하는 자동분석법이 증가하고 있다. 대규모의 Anti-Virus 연구소에서는 이러한 분석법을 자체 개발하여 대량의 악성코드를 진단 및 치료하고 있다.[5-6] 이러한 자동분석법이 증가하면서 오진단(False Positive)도 함께 증가하는 부작용이 나타나게 되었다. 오진단으로 인한 피해는 작게 보면 일부 시스템 마비 정도로 그칠 수 있으나 주요파일에 대한 오진단은 다수시스템의 마비로 국가적, 경제적 손실을 가져올 수 있다.[7] 오진은 수작업 분석이나 자동 분석 모두 발생할 수 있다. 하지만, 수작업은 사람이 분석하기 때문에, 진단 기준을 분석가의 경험치에 의해서 유연하게 변경할 수 있을 수 있어 정확성이 높으나, 자동 분석은 사람이 분석하지 않으므로, 분석과정에서 사용하는 진단 기준이 잘 못된 경우에는 대량의 오진이 발생할 수 있는 문제점이 있다. 따라서, Anti-Virus 업체들은 오진을 근본적으로 줄이기 위해 진단 기준을 정밀하게 선정하고 주기적으로 점검 하는 등 오진을 막고자 총력을 기울이고 있다. 가장 각광 받는 방법은 이미 알려진 정상파일을 수집하여 정상파일 셋을 구축한 후 Anti-Virus 시그니처를 배포하기 전에 최종적으로 정상파일 셋을 검사하여 오진을 사전에 확인하여 차단하는 방법이 있다. 하지만, 이 방법을 사용하기 위해서는 대량의 정상파일을 수집해야 하는데, 이때 수많은 파일 중에서 어떤 파일을 수집할 지와 어떤 파일을 정상파일로 판단할 지에 대한 기준이 없어 정상파일 셋 구축에 어려움을 겪고 있다. 따라서, 본 연구에서는 기존 Anti-Virus 의 오진단 주요 사례와 오진단을 줄이기 위한 관련 연구를 살펴보고, 정

상파일 취합 및 판단기준을 제안하고자 한다. 이를 통해 Anti-Virus 연구소들이 우수한 정상파일 셋을 구축하여 오진단을 줄임으로써, 오진단으로 인한 경제적 손실을 줄이고, Anti-Virus 에 대한 신뢰성을 향상 시키는데, 기여하고자 한다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련된 기존 연구들에 대해 알아보고, 3 장에서는 정상파일의 개념 및 정상파일의 특성에 대해 도출하였다. 그리고 4 장에서는 정상파일의 특성을 토대로 정상파일을 취합하는 기준과 판단하는 기준을 제시하였다. 5 장에서는 결론을 통해 향후 연구계획에 대해 정리하였다.

2. 주요 오진단 사례 및 관련연구

2.1 주요 오진단 사례

근래에 발생한 주요 오진단사례는 다음과 같다.

- 2007 년 5 월 18 일 시만텍: 중국어 윈도우 XP 서비스 팩 2 파일인 netapp32.dll 과 lasass.exe 를 Backdoor.Haxdoor 로 오진하고 시스템 파일을 삭제해 버려 시스템이 부팅되지 않는 문제 발생.[8]
- 2008 년 7 월 10 일 안철수연구소: 윈도우 XP 서비스 팩 3 의 LSASS.EXE 파일을 악성코드로 진단 및 삭제하여 시스템이 부팅되지 않는 문제 발생. [9]
- 2009 년 2 월 16 일 맥아피: 바이러스스캔에서 모증권 홈트레이딩시스템의 키보드 보안 프로그램을 악성코드로 오진하여 서비스가 정지하는 문제 발생. [10]

이러한 오진단들은 대량 파일을 자동분석시 발생한 것으로, 시스템 또는 서비스에 치명적인 문제를 일으킬 수 있어, 오진단을 줄이기 위해 다양한 연구가 진행되고 있다.

2.2 오진단을 줄이기 위한 관련연구

2.2.1. WhiteList 관리 방안에 대한 연구

오진단을 방지하기 위해 주요파일에 대해서 WhiteList 라는 목록을 만들어 관리 하는 방식이다. 이를 통해 시스템에 치명적인 문제를 일으킬 수 있는 파일의 이름, 설치 위치 등을 목록으로 관리하여, Anti-Virus 가 오진으로 해당 파일을 삭제하지 못하게 하는 방법이다. 손 쉽게 주요 파일을 관리할 수 있다는 장점이 있는 반면에, 파일의 수가 대량일 경우 목록을 유지 관리하는데 어려움이 따르며, 상대적으로 정확성이 떨어진다는 단점이 있다. [11]

2.2.2. NSRL

생성된 모든 파일에 대한 정보를 국제적인 협력을 통해 시그니처(MD5,SHA1,SHA2 등) 방식으로 구축한 DB 로서, 보안소프트웨어 및 범죄분석의 포렌식 툴에 사용하기 위한 정보를 제공한다. WhiteList 의 한계를 극복하는 장점을 가지고 있으나, 정상/악성 파일의 구분이 되어 있지 않고, 파일에 대한 분석정보를 제공하지 않아, Anti-Virus 의 오진단 검출용으로 사용하기에는 어렵다는 문제점이 있다. [12]

2.2.3.MS Catalog

MS 사는 자사 제품의 무결성 정보를 제공하기 위해 자체 보안 알고리즘을 통해 구축한 Catalog 를 제공하고 있다. Catalog 를 통해 간편하게 MS 에서 제공하는 파일에 대한 무결성 변조여부를 확인할 수 있는 장점이 있으나, MS 사에서 만든 OS 관련 파일에 대해서만 정보를 제공하고 있어, 다른 Application 에서는 이 기능을 사용할 수 없다는 한계를 가지고 있다. [5]

2.2.4. 정상파일 셋

OS 파일과 같이 오진단시 치명적인 문제를 일으킬 수 있는 파일과 범용적인 Application 파일 등을 수집하여 정상파일의 셋과 DB 를 구축하여, 실제 Anti-Virus 엔진을 배포하기 이전에 오진단 여부를 최종적으로 확인하는 방식이다. 파일정보뿐 아니라, 실제 파일을 통해 오진단을 판단하는 것으로 가장 정확성이 높다는 장점을 가지고 있으나, 정상파일을 대량으로 수집하고 판단하는 작업이 선행되어야 하는 단점을 가지고 있다. [7]

최근의 Anti-Virus 연구소들은 오진단 검출 정확성이 가장 높은 정상파일 셋 구축 방식을 선호하고 있다. 하지만, 어떤 파일을 취합해야 되는지에 대한 기준이 없고 정상파일에 대한 판단 기준 존재하지 않아, 정상파일 셋 구축에 어려움을 겪고 있다.

3. 정상파일의 개념 및 7 대 특성

본 장에서는 정상파일의 개념을 설명하고, 정상파일이 악성코드 파일과 차별되는 7 대 특성을 제시한다.

3.1. 정상파일의 개념

정상파일은 악의적인 기능을 하지 않고, 제작자가 공개한 기능 및 목적 내에서 본래의 역할을 수행하는 신뢰된 제작자에 의해서 개발 및 배포되는 파일이라고 정의할 수 있다.

3.2. 정상파일의 특성

정상파일의 개념으로만 정상파일을 판단하기에는 어려움이 있다. 따라서, 본 절에서는 정상파일이 가지는 특성을 제시한다.

① 정상파일은 악의적인 기능을 수행하지 않는다.
② 정상파일은 사용자가 승인한 기능만을 한다.
③ 정상파일은 정상적인 라이브러리 파일로 구성된다
④ 정상파일은 시스템이나 다른 파일의 기능을 방해하지 않는다.
⑤ 정상파일은 신뢰된 제작자에 의해 제공된다.
⑥ 정상파일은 적절한 사이즈를 가지고 있다.
⑦ 정상파일은 지정된 위치에 설치된다.

이와 같이 정상파일은 7 대 특성으로 상세 정의할 수 있다. 이 특성은 앞으로 정상파일을 취합 및 판단할 때 기준의 근거가 된다.

4. 정상파일의 취합기준 및 판단기준

본 장에서는 7 대 특성에 기반하여 정상파일을 취합할 때 사용되는 취합기준과 수집된 파일을 최종적으로 정상파일인지 확인하기 위한 판단기준을 제시한다.

정상파일의 특성에 기반하여 수집할 대상을 선정 할 때 적용하는 취합기준은 [표 1]과 같다. 기준을 살펴보면, 운영체제 시스템에서 사용하는 OS 파일은 정상파일의 7 대 특성을 만족하고 있으며, Operation System 구성파일들은 오진단시에 시스템이 정지하는 치명적인 영향을 미치기 때문에 정상파일 취합기준에 해당된다. 주요 수집 유형은 보급률이 높은 MS 계열 OS 파일이 대다수를 차지하고 있다. H/W Driver 파일은 Hardware Device 의 정상적인 기능이 수행하도록 하는 파일로서, 정상파일의 특성을 모두 포함하고 있다. 또한, 오진단시에는 Device 장치가 작동을 멈추는 문제가 발생하기 때문에, 필히 수집되어야 하는 파일이다. Public Application 은 범용적으로 사용되는 모든 응용프로그램을 말하는 것으로서, 해당 응용프로그램을 Anti-Virus 에서 오진시에 사용자의 작업을 방해할 수 있어 수집이 필요하다. 단, 그 유형 및 수가 다양하고 많은 파일이 존재하므로, 응용프로그램 유형별 인기도를 선정하여 취합 대상을 선정한다. Local Application 은 범용 응용프로그램이 아니라, 조직내에서 사용하기 위해 개발된 특수한 응용프로그램을 뜻한다. 오진시에는 해당 조직내 생산 및 업무시스템을 마비시키는 큰 문제를 일으킬 수 있으므로, 반드시 수집이 필요하다. 단, Local Application 은 조직의 자산이므로, 해당 조직의 요청시 수집을 원칙으로 한다.

기준	설명	유형
OS 파일	- 치명적인 시스템 failure 를 방지 목적	MicroSoft, Apple SUN/AIX/HP FreeBSD RedHat Nokia Google
H/W driver	- Device 동작 failure 를 방지 목적 - 3rd party 제공업체 시장 점유율을 기준으로 취합대상 선정	Board CPU Chipset Device
Public Application	- 사용자 작업 장애를 방지 목적 - Application 유형별 인기도를 기준으로 취합 대상을 선정	System utility Graphic Multimedia Publish Game Programming
Local Application	- 사용자 조직의 생산활동 장애방지 목적 - 사용자 조직에서의 개발한 application file - 조직에서 요구시에 대상으로 선정	정보 관리 병원 관리 IT 서비스 금융 서비스 방송 서비스 국가보안 운용 설비제어 프로그램 시스템 펌웨어
		시스템 드라이버
Script File	- 업무장애 방지 목적 - 일괄 작업등에 사용되는 script 언어 파일	Batch script Python script Shell script Excel/VBS script Autoit script
Install file	- S/W 설치 불가 장애 방지 목적	Installer

<표 1> 오진 방지를 위한 정상파일 취합기준

지금까지 정상파일 대상을 수집할 때 사용할 취합기준을 제시하였다. 다음 절차로 수집된 파일이 최종적으로 정상파일인지를 판단시 사용할 판단기준을 [표 2]와 같이 제시한다.

항목	기준	설명
제작 정보	제작사 정보	제작사 신뢰도/인지도를 통한 정상파일의 배포처 여부를 구분
파일 속성	Name	파일의 최상위 기본정보
	Hash Value	-제공된 무결성 정보와 일치도 -MS Catalog, NSRL 등 공인된 무결성 정보와 일치도
	Date Time	배포본과 일치여부 판단 기본정보
	Size	- 악의적 코드에 의한 변경여부 판단 기본정보
	Description	- 배포처 정보와 일치여부를 통한 판별 근거 - 파일의 동작과 일치여부를 통한 판별 근거
	Update Time	- 파일의 변화(patch)여부 판별 근거
	Path	- 파일의 시스템상의 위치에 따른 속성 분류 판단 근거 - 배포본의 규정된 파일의 위치와 다를 경우 조작 여부 판단 근거
	Version	- 배포본과 일치여부 판단 기본 정보 - 정상적인 Build 본인지에 대한 판단 근거 - 취합기준에 따른 취합된 파일의 이력관리 근거
	Type	파일 확장자를 통한 속성 및 행위 판단 근거 (실행/데이터/드라이버/dll)
	Packer	- 배포본의 진위 여부 판단 근거 - 난독화 등을 통한 분석 방해 여부 판단 근거 -unknown packer 판단 여부
무결성 정보	CodeSign	- 파일의 사이닝 정보를 통한 배포처 판별

		- 디지털서명 유효성 검사를 통한 수정여부 판단 근거
	File Header	- 패딩영역(헤더상 빈영역)을 통한 파일 수정여부 판단 근거 - 파일 섹션 손상여부를 통한 수정여부 판단 근거
	MS Catalog	MS Catalog 정보/MS 문자열 정보 매칭을 통한 MicroSoft 배포 파일 여부 판단 근거
History	취합 시기	정상파일 이력관리
	취합 정보	취합경로와 제작사 배포방식 일치 여부를 통하여 배포본에 포함된 파일임을 판별.
Black List	진단 값	보안소프트웨어 진단 값과 동일한 값을 보유하고 있는지를 통해 악성여부 판단.
	Import Function	PE info 상의 import 된 function 종류에 따른 black list 여부 판단 근거
File Action	행위	- 실행 파일인 경우 실행 유/무를 통한 손상여부 판단 근거 - 실행시 증상을 통한 정상/악성여부 판단 근거

<표 2> 오진방지를 위한 정상파일 판단기준

정상파일 여부를 판단하기 위해서는 가장 먼저, 제작 정보를 이용한다. 제작 정보를 통해 해당 제작사가 정상파일을 배포하는 업체로서의 신뢰도와 인지도를 가졌는지를 판단하고, 지속적으로 해당업체의 신뢰도를 평점을 통해 관리한다. 파일 속성은 해당 파일이 가지고 있는 내부적인 정보로서, 해당 파일의 특징을 나타낸다, 따라서, 해당 정보의 정상여부를 판단하는데 가장 많은 기초자료를 제공하는 기준이 된다. Name, Hash Value, Date time, Size, Description 등이 사용된다. 무결성 정보는 해당 파일의 변조여부를 확인하여, 원래 제작사에서 제공한 파일과 동일한지를 판단한다. 이때, CodeSign, File Header, MS Catalog 가 사용된다. History 는 취합시기와 취합경로가 기준이 되며, 취합시기는 해당 파일이 제작사에서 배포한 시기와 동일한지 판단하는 근거가 된다. 취합정보를 이용해 파일을 수집한 경로가 제작사가 공식적으로 배포본을 배포하는 사이트와 동일한 지 확인한다. BlackList 는 해당 파일의 악성여부를 Anti-Virus 진단과 Import Function 정보를 통해 사전에 확인하여, 정상파일을 판단하기 전에 사전 분류작업을 위한 용도로 활용된다. Function Action 은 최종적으로 사용되는 판단기준으로서, 위의 판단기준을 통해 정상파일 여부를 확인하기 어려울 경우에 해당 파일을 실행하여, 정상/악의적인 행위를 하는지를 분석가가 확인하여 최종적으로 정상파일여부를 판단한다.

5. 결론 및 향후 연구계획

본 논문에서는 악성코드의 기하급수적인 증가로 대량의 악성코드 자동분석법이 개발되고 이로 인해 다시 오진이 증가하는 점에 주목하였다. Anti-Virus 는 대량의 악성코드를 진단하면 자연스럽게 오진이 증가할 수 밖에 없는 구조이지만, 주요 파일에 대한 오진은 악성코드에 못지 않은 큰 피해를 일으킬 수 있다. 따라서, 최근에는 오진단 방지를 위해 정상파일 셋을 구축하는 연구가 진행되고 있다. 하지만, 정상파일에 대한 명확한 취합 및 판단 기준이 없어, Anti-Virus 연구소들은 구축에 어려움을 겪고 있다. 따라서, 본 연구에서는 Anti-Virus 의 오진을 줄이기 위한 연구의 시발점으로서, 정상파일의 7 대 특성을 제시하였다. 이 특성을 기반으로 정상파일을 수집시에 사용할 취합기준을 개발하였으며, 취합 후에 최종적으로 정상파일로 확인할 판단기준을 개발하였다. 본 논문에서 제시한 Anti-Virus 오진 방지를 위한 취합기준과 판단기준을 통해 다수의 우수한 정상파일 셋을 구축하여 오진단을 줄임으로써, Anti-Virus 의 신뢰성을 높이고 시스템의 안정성을 도모함으로써, 오진으로 인한 국가적, 경제적 손실을 예방하는데 기여할 것으로 기대된다.

향후 연구계획으로는 취합기준과 판단기준을 바탕으로 정상파일 셋을 구축하는 시스템을 구현하고 지속인 보완을 거쳐, Anti-Virus 오진 방지 구현 모델을 제시하고자 한다.

참고문헌

[1] F. Freiling, T. Holz, and G. Wicherski, “Botnet Tracking-Exploring a Root-Cause Methodology,” ESORICS 2005, LNCS 3679, pp. 319-335, 2005.

[2] D. Barroso, “Botnets-The Silent Threat,” ENISA Position Paper, Nov. 2007.

[3] 이한우, 최현상, 이희조, “DNS 기반의 봇넷 탐지 시스템,” 한국정보처리학회 추계학술발표대회논문집, pp. 13790-1382, 2006 년 10 월

[4] 허종오, “악성코드 수집을 위한 허니팟 구현 모델 연구,” 한국정보처리학회 춘계학술발표대회논문집, 2010 년 4 월

[5] “ASEC Report 2009,” AhnLab Corporation, 2009

[6] 허종오, 조시행 “악성코드 수집을 위한 글로벌 허니팟 시스템 구축에 관한 연구,” 한국정보과학회 KCC2010 논문집, 2010 년 7 월

[7] “ASEC Report 2010,” AhnLab Corporation, 2010

[8] Symantec, Press Release, http://www.symantec.com, May. 2007

[9] AhnLab, Press Release, http://www.ahnlab.com, Aug. 2008

[10] Mcafee, Press Release, http://www.mcafee.com, Feb. 2009

[11] Thomas Parsons, A False Positive Prevention Framework for Non-Heuristic Anti-Virus Signatures, Jan. 2009.

[12] National Institute of Standards and Technology, National Software Reference Library(NSRL), http://www.nsrl.nist.gov/