파일 평판을 이용한 알려지지 않은 악성 코드 탐지

조윤진
고려대학교 컴퓨터 정보통신공학과
e-mail : yunjin_cho@korea.ac.kr

Unknown Malware Detection Using File Reputation

Yun-Jin Cho
*Dept. of Computer Science, Korea University

요 약

최근 발생한 다양한 해킹 사건에서와 같이, 신규 또는 알려지지 않은 악성코드를 이용한 지능형 지속 공격이 점차 증가하고 있다. 기존의 악성 코드가 해커의 단순한 호기심이나 해커 자신의 능력을 과시하기 위해 제작되어 불특정 다수를 공격했다면, 최근의 해킹 사건에서 사용된 악성코드는 오직 특정 대상만을 목표로 하여 제작, 사용되고 있는 것이 특징이다. 현재 대다수의 악성코드 탐지 방식인 블랙리스트 기반의 시그니처에 의한 탐지방식에서는 악성코드의 일부분이라도 변경이 되면 해당 악성코드를 탐지할 수가 없기 때문에 신규 악성 코드를 탐지하고 대응하는 것이 어렵다. 그러므로 지능형 지속 공격에 대응하기 위해서는 새로운 형태의 파일 탐지 기술이 필요하다. 이에 본 논문에서는 파일의 다양한 속성 및 사용자 분포에 따른 평판점수를 통해 신규 악성코드를 탐지하는 기법을 제안한다.

1. 서론

초기 사이버 공격은 호기심 및 자기 과시를 위한 목적으로 일부 해커에 의해 제작 되어 불특정 다수를 목표로 공격을 하였다. 하지만 최근 발생한 사이버 침해 사고를 보면, 특정 공격 대상 기업을 지정하여 해당 기업의 보안 상황을 연구하고, 지속적으로 공격을 시도하여 해킹에 성공한 것을 볼 수 있다.

이러한 공격을 APT(Advanced Persistent Threat) 즉, 지능형 지속 공격이라고 말한다. APT 의 대표적인 공격 방식에는 웹을 이용한 워터링홀 공격과 이메일을 이용한 스피어피싱이 있다. 그런데 이 두 가지 공격방식은, 대다수의 안티바이러스 제품들에서 사용하고 있는 블랙리스트 방식의 시그니처 기반을 통한 악성코드 탐지 기법을 우회하기 위해 알려지지 않은 파일을 이용한다는 것에는 동일하다. 블랙리스트방식의 시그니처 기반 악성코드 탐지 기술은 치명적인 단점이 세 가지 있다.

첫 번째는 파일의 일부분에서라도 변화가 발생하면 그것을 다른 파일로 인식하고 탐지를 못한다는 것이며, 두 번째는 점점 늘어나고 있는 악성코드를 모두 탐지하기 위해서는 시그니처를 저장하고 있는 파일의 크기가 계속 증가해야 한다는 것이다. 마지막 세 번째는 악성파일이 발견되지 않았으며, 해당 파일에 대한 분석이 완료되지 않았다면 해당 악성코드는 탐지할 수 없다는 것이다. 그러므로 시그니처에 의한 파일 탐지기술은 APT 에서 사용되는 신규 악성코드를 탐지하는 것에는 한계가 있다.

하지만 파일은 해시값과 같이 고유한 값으로 각 파일을 구별할 수 있으며, 파일의 다양한 속성정보와 각 파일들의 분포 현황을 이용하여 파일을 점수화하면, 평판 기반의 파일 탐지를 할 수 있다. 따라서 본 논문에서는 기존의 악성코드 탐지방식의 한계점을 보안하기 위해, 파일의 다양한 속성 및 사용자 분포에 따른 다양한 결과를 기반으로 알려지지 않은 악성 파일에 대한 대응 방법을 설명한다.

2. 관련 연구

각종 보안 장비들에서 악성코드를 탐지하는 기법 중 가장 보편적으로 사용하는 기술은 시그니처에 의한 악성코드 탐지 방법이다. 시그니처 기반의 악성코드 탐지 기법은 알고 있는 악성코드의 바이너리 정보와의 비교를 통해 파일을 탐지하는 정적 분석 방식이다. 하지만 이러한 대응방식은 알려지지 않은 악성코드 대응에 대해 몇 가지 취약점을 가지고 있기 때문에 새로운 탐지 기술의 필요성 및 새로운 파일 탐지 기법을 설명하고자 한다. 본 논문에서 이야기하는 파일은 오직 실행 가능한 파일만을 대

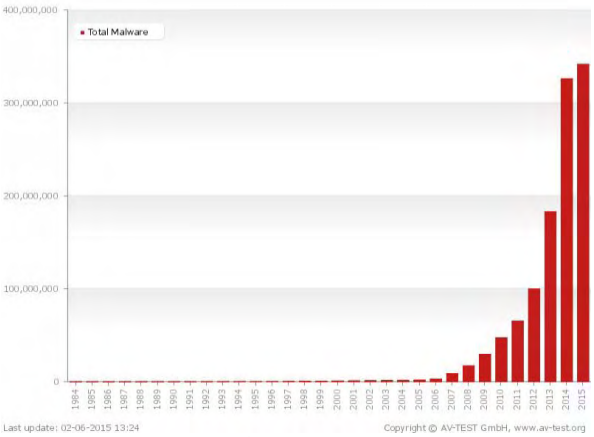
상으로 하며 문서 파일과 같이 개인이 만든 파일은 연구 대상에서 제외한다.

2.1 시그니처 파일 탐지의 문제점

시그니처 기반의 악성코드 탐지 기술은 정확한 파일 탐지를 위해 파일의 바이너리값과 비교 방식을 통한 파일 탐지 방법을 사용한다. 그러한 탐지 기법은 탐지해야 하는 파일의 수량이 증가하게 되면 탐지해야 하는 파일의 바이너리 값도 증가할 수밖에 없으며, 이는 악성코드 탐지를 위한 검사 시간에까지 영향을 미치게 된다.

$$\text{Scanning Time} = \frac{\text{Number of viruses} \times \text{Number and size of the file objects} \times \text{Number of Detection Methods}}{\text{Speed of Processor}}$$

아래 (그림 1)과 같이 최근 10 년동안 악성코드의 수량이 기하급수적으로 증가하고 있기 때문에 악성코드 검사 시간이 크게 증가할 수밖에 없는 문제가 생기는 것이다.



(그림 1) AV-TEST 에서 발표한 최근 10 년간의 악성코드 수량

또한, 시그니처 방식에 의한 파일 탐지 방식은 <표 1>과 같이 분석이 진행된 파일의 탐지 바이러리에 의한 시그니처가 제작완료된 이후부터 탐지가 가능하다. 이는 신규 악성코드 발견부터 탐지까지 시간이 필요하다는 것으로 볼 수 있으며, 또한 악성코드 탐지를 위해서는 해당 파일을 수집을 해야만 탐지가 가능하다는 것으로 볼 수 있다.

악성파일 발견
↓
파일 수집
↓
분석 요청
↓
파일 분석
↓
시그니처 제작
↓
배포 및 적용

<표 1> 악성코드 탐지 시그니처 제작 단계

2.2 알려지지 않은 파일

Anti-Virus 회사 입장에서 파일은 알려진 파일과 알려지지 않은 파일로 구분할 수가 있으며, 이는 다시 알려진 좋은 파일, 알려진 나쁜 파일 그리고 알려지지 않은 파일로 구분할 수가 있다. 알려진 나쁜 파일, 즉 알려진 악성코드는 시그니처 기반의 파일탐지 기술을 이용하여 탐지 가능하지만 알려지지 않은 파일들 중에서의 악성코드는 시그니처 기반의 탐지엔진을 이용해서는 탐지를 할 수가 없다. 그런데 최근 발생한 침해 사고에서 사용된 악성코드는 알려지지 않은 악성코드를 이용한 공격이었다. 즉 이제는 알려진 악성 코드 탐지가 아닌, 알려지지 않은 악성코드 탐지부분이 더욱 중요하게 된 것이다.

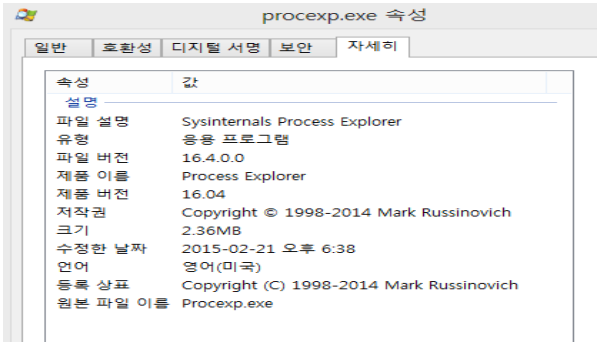
3. 탐지 방법 연구

3.1 파일의 무결성 정보

MD5 나 SHA 와 같은 알고리즘을 이용한 파일 정보는 파일 전송 시 전송된 파일과 수신된 파일의 변조유무 확인 시 많이 사용하고 있다. 이러한 무결성 정보를 이용하면 파일의 유일한 값을 확인할 수 있기 때문에 파일들을 구분 할 수 있게 된다. 즉, 각 파일들을 데이터베이스화할 때 Primary Key 로 사용할 수 있는 값이 되는 것이다.

3.2 파일의 속성 정보(Metadata)

파일 속성 정보는 파일의 속성과 특징 그리고 다른 자원과의 관계를 기술하여 파일의 활용성을 높이고 파일에 대한 제어와 관리를 돕는 구조화된 데이터를 제공한다. Windows 환경에서는 속성을 수집하는 GetFileAttributes 와 GetFileAttributesEx 함수를 제공하고 있으며, (그림 3)과 같이 GUI 를 통해서도 확인할 수 있다.

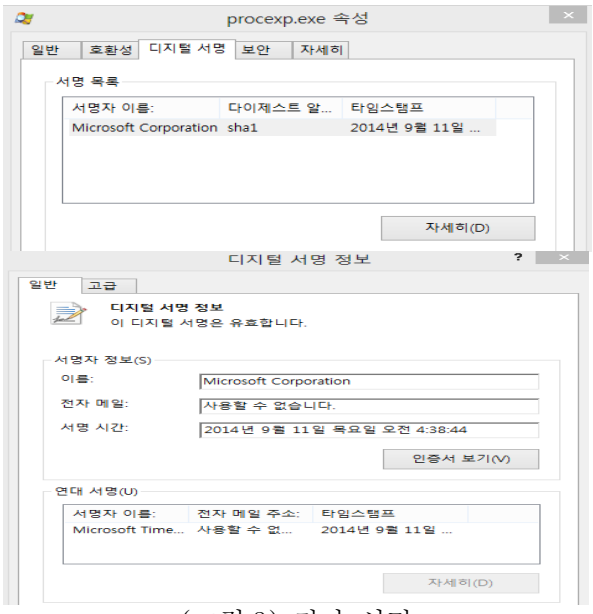


(그림 2) GUI 를 통한 파일의 속성정보

3.3 전자 서명(Digital Signature)

전자 서명은 파일을 제작한 사람의 신원을 증명하기 위해 디지털 형태로 파일에 첨부된

정보이다. 이는 파일 생성자에 대한 정보 확인 및 해당 파일을 개발한 기업 또는 개인이 자신들이 개발한 파일이 위변조 되지 않았음을 전자 서명을 통해 증명하는 용도로 사용할 수 있다. 즉, 전자 서명 정보는 파일의 제작자 확인을 통해 그 파일의 신뢰성 여부를 판단할 때 사용할 수 있는 정보이다.



(그림 3) 전자 서명

3.4 파일의 분포

하나의 시스템에서 사용되고 있는 파일은 다른 누군가의 시스템에서도 동일하게 사용될 수 있다. 즉, 파일의 고유한 값을 토대로 얼마나 많은 사용자가 해당 파일을 가지고 있는지 파일에 대한 분포를 확인할 수 있다. APT 공격 시 사용되는 알려지지 않은 악성 파일은 일부 사용자를 위해 제작되어 배포되므로 많은 사용자가 사용할 수 없다. 그러므로 이러한 파일 분포 정보는 해당 파일 원본을 수집하여 분석하지는 못했지만 여러 사용자가 동일하게 사용함으로 인해 파일에 대한 평판점수 적용 시 이용할 수 있는 정보이다.

3.5 파일 평판 정보표

파일의 다양한 정보를 기반으로 아래 <표 3>과 같이 파일에 대한 평판 정보 표를 만들 수 있다.

파일 평판 정보 표에는 수집된 파일의 속성 정보, 전자서명 여부, 동일 파일 사용자 수, 그리고 마지막으로 해당파일을 처음 발견된 날짜와 같은 정보가 필요하다. 파일이 오랜 기간 동안 많은 사용자들이 사용한 것이고 신뢰성 있는 기관에서 생성된 것이라는 점이 인증되어 있다면 파일에 대한 평판이 높아질 것이기 때문이다.

파일 해시값	92e04bcf92cf588f434393d0b3b6bca2
실행 파일 이름	procexp.exe
날짜	2015.02.10 00:34:31
파일 경로	c:₩windows₩download₩procexp.exe
해시 알고리즘	MD5
파일 제작 회사	Microsoft Corporation
파일 버전	16.04
파일 사이즈	13934040
웹 도메인	technet.microsoft.com
다운로드 사이트	https://technet.microsoft.com/en-us/sysinternals/bb896653
다운로드를 시도한 프로세스	c:₩program files₩internet explorer₩iexplore.exe
동일 파일 사용자 수	100000 명 이상
처음 발견된 날짜	2010.02.12
해당 파일을 사용한 OS 정보	Windows 7 Home Premium Edition
전자 서명 여부	Signed file, verified signature

<표 3> 파일 평판 정보 표

3.6 파일 평판 평가표

파일 평판 정보 표의 각 사항들을 통해 파일에 대한 평가를 할 수 있으며, 이러한 평가내용을 기반으로 각 파일에 점수를 부여할 수 있다. 아래 <표 4>는 예시문이다.

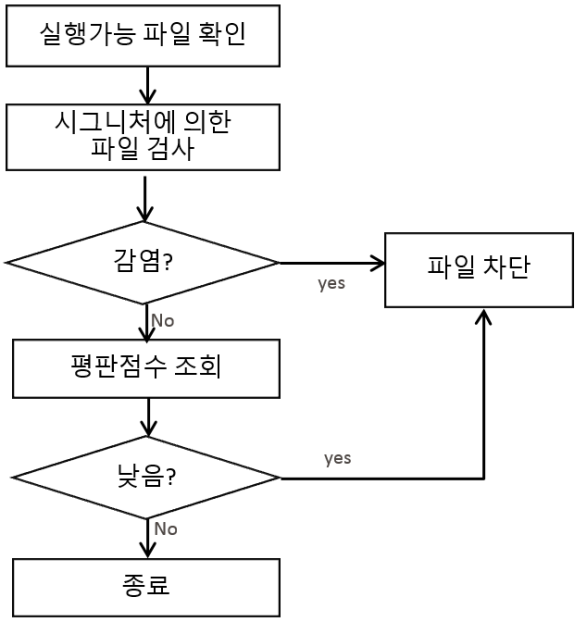
확인 사항	평가 사항
파일 생성일 확인	최근 한달 이내에 생성되거나 수정 되었는가?
버전 명시 확인	버전은 명시되어 있는가?
파일 사이즈 확인	파일 사이즈는 10KB 에서 500KB 사이인가?
사용자 수 확인	해당 파일을 사용하는 사용자는 많은가?
전자 서명 여부 확인	전자 서명은 유효한가?

<표 4> 파일 평판 정보 평가표

3.7 평판 정보를 이용한 파일 탐지 기법

시스템 상에서 파일 검사 시 시그니처를 이용하여 알려진 악성코드 탐지 후 평판 기법을 이용하여 파일을 추가적으로 분석할 수 있다. 예를 들어, 사용자 또는 실행파일에 의한 파일 다운로드 시 해당 파일은 Temp 폴더에 임시다운로드가 되며, 시그니처 기반의 파일 탐지 엔진이 해당 파일을 검사하게 된다. 이때 해당 파일이 알려진 악성코드라면 바로 탐지되지만 알려지지 않은 악성코드 파일이라면 시그니처 기반에서는 탐지가 불가능하다. 하지만 평판 정보를 이용하게 되면 시그니처 기반 탐지 이후

파일의 고유한 값인 해시값을 토대로 파일의 평판 점수를 확인하여 파일을 추가로 탐지할 수 있게 된다.



(그림 4) 시그니처기반과 평판정보를 이용한 악성파일 탐지 방법

4. 결론

시그니처 기반의 악성코드 탐지 방법은 APT 공격 시 사용되는 신규 악성코드를 탐지하는 것에는 한계가 있다. 그러므로 본 논문에서 제시한 파일 평판 정보를 이용한 파일 탐지 기법을 사용하게 되면 신규 악성코드 탐지율을 높일 수 있을 것으로 보인다.

참고문헌

[1]av-test, http://www.av-test.org/en/statistics/malware/
[2] Umakant Mishra "Contradictions in improving speed of virus scanning"
[3] TaeGuen Kim, In-Kyoung Kim, Eul Gyu Im "Malware Detection Method via Major Block Comparison" 보안공학연구논문지 제 9 권 제 5 호 P401-416 2012 년
[4] Jaeho Lee, Sangjin Lee "A Study on Unknown Malware Detection using Digital Forensic Techniques" 정보보호학회 P107 -122 2014 년
[5] Hee Jun Kwon, Eul Gyu Im "Researches about Malware Detectoion Using N-gram"
[6] Lara Srour, Ayman Kayssi, Ali Chehab "Reputation-Based Algorithm for Managing Trust in File Sharing Networks"
[7]유홍렬, 정성미, 권태경 " 새롭게 진화하는 위협의 패러다임 - 지능형지속 위협(APT)" 전자공학회지 P304-318 2014 년
[8]임설화, 김종수,양준근, 임채호 "APT 현황과 신종 악성코드 대응방안" 정보보호 학회지 제 24 권 제 2 호 P64-72 2014 년
[9] MoonGoo Lee, Chunsock Bae "A Study for the Principle Cases of Advanced Persistent Threat Attacks" 대한전자공학회 P939-942 2013 년
[10] Vinod P, V.Laxmi,M.S.Gaur "Survey on Malware Detection Methods" Hack.in 2009, 3rd Hackers' Workshop on Computer and Internet Security P74-79, 2009
[11] Daesung Moon, Hansung Lee, Ikkyun Kim "Host based Feature Description Method for Detecting APT Attack" 정보보호학회논문 제 24 권 제 5 호 P839-850, 2014
[12] Kevin Hoffman, David Zage, Cristina Nita-Rotaru "A Survey of Attack and Defense Techniques for Reputation Systems"