시각화 기법을 이용한 악성코드 분석 및 분류 연구*

송인수** · 이동휘** · 김귀남**

요 약

인터넷 기술의 급격한 발전으로 인한 편리함과 더불어 다양한 악성코드들이 제작되고 있다. 악성코드의 발생건수는 날이 갈수록 부지기수로 늘어나고 있으며, 변종 혹은 새로운 악성코드에 대한 유포는 매우 심각하여 악성코드에 대한 분석은 절실히 필요한 시점이다. 악성코드에 대한 판단기준을 설정할 필요가 있으며, 알고리즘을 이용한 악성코드 분류의 단점은 이미 발견된 악성코드에 대한 분류는 효율적이나 새롭게 생긴 악성코드나 변종된 악성코드에 대해서는 새로운 탐지가 어려운 단점이 있다. 이에 본 연구의 목적은 시각화 기법의 장점을 이용하여 기존의 다변량의 악성코드에 대한 측정 및 분석뿐만 아니라, 변종 혹은 새로운 악성코드에 대해서도 새로운 패턴 혹은 형태를 도출하여 새로운 악성코드와 변종들에 대해서 대처하는데 있다. 따라서 본 논문에서는 업체에서 제공되는 악성코드 속성을 시각화하여 분석하는 기법을 제안하고자 한다.

A Study on Malicious Codes Grouping and Analysis Using Visualization

In Soo Song** · Dong Hui Lee** · Kui Nam Kim**

ABSTRACT

The expansion of internet technology has made convenience. On the one hand various malicious code is produced. The number of malicious codes occurrence has dramadically increasing, and new or variant malicious code circulation very serious, So it is time to require analysis about malicious code. About malicious code require set criteria for judgment, malicious code taxonomy using Algorithm of weakness difficult to new or variant malicious code taxonomy but already discovered malicious code taxonomy is effective. Therefore this paper of object is various malicious code analysis besides new or variant malicious code type or form deduction using visualization of strong. Thus this paper proposes a malicious code analysis and grouping method using visualization.

Key words : Malicious Codes, Malicious Codes Regrouping, Visualization, Parallel Coordinates

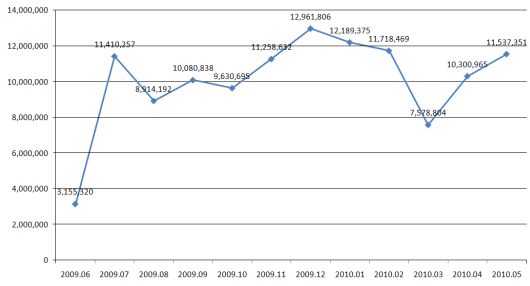
접수일 : 2010년 8월 17일; 채택일 : 2010년 9월 24일

* 본 연구는 지식경제부 지역혁신센터사업인 산업기술보호특화센터 지원으로 수행되었음.

** 경기대학교 산업기술보호특화센터

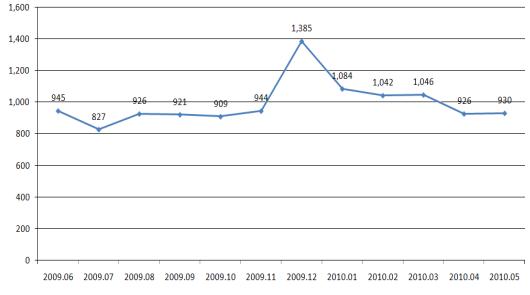
1. 서 론

악성코드는 악의적인 목적을 위해 작성된 실행 가능한 코드를 일컫는다. 바이러스, 웜, 트로이목마 프로그램 등이 모두 여기에 속하며, 이를 악성 프로그램이라고도 칭한다. 컴퓨터 악성코드는 빠르게 진화하고 있으며, 다양한 시스템상의 취약성을 이용하여 악의적인 활동들을 수행하고 있다[1].



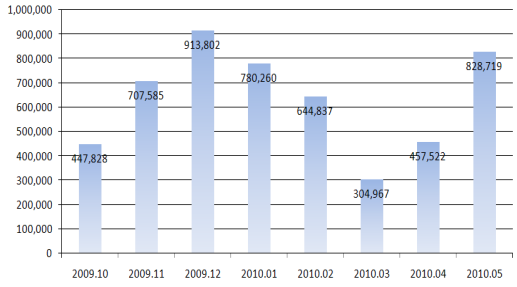
(그림 1) 최근 악성코드 발생건수

(그림 1)에서 안철수 연구소의 보안위협 종합 정보시스템을 통한 악성코드 발생건수 통계치를 살펴보면, 2009년 하반기의 경우 8백만에서 1천 3백만 정도의 악성코드 감염보고가 이뤄지고 있으며, 2010년 상반기의 경우에는 7백만에서 1천 3백만 정도의 감염보고가 이뤄지고 있다[2].



(그림 2) 최근 악성코드 유형

또한 (그림 2)에서는 월별 악성코드 유형은 2009년 상반기에는 큰 변화 없이 지속적으로 증가하여 12월에 1,385건을 기록하였지만, 2010년에 들어서서 다시 일정한 추세를 보이며, 1,000건 정도를 유지하고 있다[2].



(그림 3) 최근 신종 악성코드 감염보고 건수

가장 문제 되는 신종 악성코드 혹은 변형 악성코드에 대한 내용은 (그림 3)에서 나타내고 있다. 2009년 10월부터 살펴보면 447,828건의 신종 악성코드 감염보고가 들어왔으며, 12월에는 913,802건으로 증가 추세를 보였다. 반면 2010년 3월 304,967건으로 감소하였으며, 5월에는 828,719건으로 다시 증가하는 통계를 볼 수 있다. 매월 감염보고 되는 신종 악성코드의 구성을 살펴보면 가장 많이 차지하고 있는 유형은 Trojan으로 매월 감염보고 건수의 50% 이상을 차지하고 있었으며, 그 다음으로 매월 꾸준하게 감염보고 되는 유형으로는 애드웨어, 드로퍼, 웜 순으로 나타내고 있다[2].

이와 같은 통계를 바탕으로 살펴본다면 악성코드에 대한 위협은 양적인 측면과 질적인 측면 모두에서 심각한 수준이다. 특히 질적인 측면에서는 다양한 공격 및 감염 기법들을 통하여 해킹과 악성코드의 경계선이 사라진지 오래이며, 신종 혹은 변종 악성코드에 대한 위협에 대한 대책이 절실히 필요한 시점이다[3]. 현재 국내에서는 악성코드 분류 시스템에 대한 연구는 거의 진행되고 있지 않으며, 대표적인 안티바이러스 업체인 안철수 연구소 와 하우리 등에서 자체 분류 기준을 개발하여 사용하고 있는 실정이다. 해외의 사례로는 소포스, 트래드 마이크로, 카스퍼스키 등에서 분류기준을

만들어 사용 중이다. 지침이나 알고리즘에 의한 분류 방법은 정해진 정책에 벗어나거나 알고리즘에 적용되지 않는 내용에 대해서는 효과적인 분류가 힘든 단점이 있다.

따라서 본 논문에서는 시각화 기법을 이용해 업체에서 분류한 악성코드 내에서의 악성코드 특성과 형태를 시각화를 통해 분석 및 분류하는 기법을 제안한다. 또한 다변량의 악성코드 정보를 직관적으로 이해할 수 있게 돕고, 지침에 따른 알고리즘에 의한 분류, 즉 신종 혹은 변종에 대한 분류가 힘든 단점의 보완하고자 한다.

본 논문의 구성은 총 5장으로 구성된다. 제 2장에서는 관련연구로써 기존의 악성코드 분류 방법과 시각화 기법을 이용한 다변량의 데이터 분석과 악성코드 분석 기법을 설명하고 분석한다. 제 3장에서는 본 논문에서 제안하는 악성코드 속성을 시각화 하기 위한 데이터 선택 방법과 적용 방법, 시각화를 통한 표현 방법에 대해서 설명하고 이를 이용한 악성코드 분석과 분류에 대해서 설명한다. 제 4장에서는 제안하는 방법을 이용한 실험 및 결과에 대해 설명하며, 제 5장에서 결론을 맺는다.

2. 관련 연구

2.1 기존의 악성코드 분류 방법

기존의 악성코드 분류에 관한 연구는 매우 미흡한 상황이다. 이에 안티 바이러스 업체에서 자사의 규칙에 따른 분석 및 분류를 하고 있지만 공개되지 않은 상태이다. 허나 이 필요성은 중요시 되고 있다. 기존의 악성코드 분류 방법론을 살펴보면 악성코드를 총 9개의 그룹으로 나누어 그룹별 클러스터를 이뤄 악성코드 데이터베이스 설계를 하여 적용시킨 후에 점수 계산을 하여 분류한 연구가 있다[4, 5]. 또한 NativeAPI 빈도를 기반으로 퍼지 군집화를 적용시켜 각 업체마다 상이한 악성코드 분류에 대한 재분류하는 기법이 연구되고 있다.

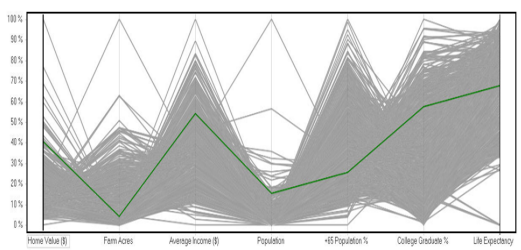
2.2 시각화

2.2.1 개요

정보 시각화 기법의 발전은 대량의 데이터를 저장할 수 있는 컴퓨터 시스템이 가능하기 때문이다. 버클리 대학교의 연구가들은 매년 1 Exabyte의 데이터가 생성된다고 추정한다. 여러 분야에서 다양한 데이터의 효과적인 분석을 하기 위한 연구를 필요로 하고 있다. 다량의 방대한 데이터를 효과적으로 분석하고, 빠르게 인지하고 대응하는 것이 중요하다[7]. 현재 각 분야의 목적과 방식에 따라 다양한 시각화 기법들이 개발되고 있다. 시각화 기법은 수학공식 혹은 통계, 탐지, 분류 알고리즘과 달리 이미지의 패턴을 사용하여 이미지에 의해 우리가 모르고 있던 새로운 패턴을 찾을 수 있는 기법으로도 사용된다. 시각화 기법은 종류가 다양하고 광범위 하여, 제안하는 기법으로 사용하게 될 Parallel Coordinates에 대해서 설명하고, 보안에 활용하고 있는 시각화 기법들과 효과에 대해 설명한다[7].

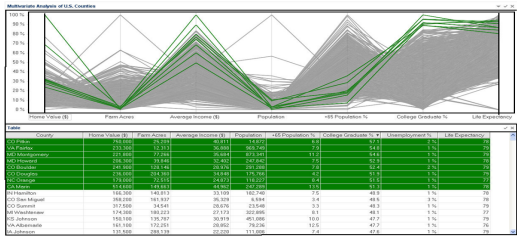
2.2.2 Parallel Coordinates

평행좌표계로 불리는 Parallel Coordinates는 다차원의 데이터를 2차원 평면에 데이터로 표현하여 나타내는 방법으로 동간격으로 놓은 수직선들이 각각 하나의 차원에 해당하게 표현하는 방식이다. 각 차원을 통과하는 선분의 모습을 관찰하여 경향, 패턴, 상관관계를 밝혀 낼 수 있다[8-10].



(그림 4) 평행좌표계를 사용한 시각화 기법의 예

다량의 변수를 Parrallel Coordinates상으로 표현하게 되면, 두 차원의 축을 기준으로 다수의 차원 데이터를 나타낼 수 있다. 따라서 Parallel Coordinates를 사용하게 되면 변형이나 차원 축소가 필요 없이 모두 표현 가능하기 때문에 데이터의 표현과 분석에서 오는 오류를 줄일 수 있는 장점이 있다. (그림 4)에서는 미국 내의 각 도시에 다량의 데이터를 Parallel Coordinates상에 나타내고 있다. Parallel Coordinates의 효과를 살펴본다. 첫째 모든 특징점에 대해 가중치 없이 같은 상태로 표현되는 점이다. 둘째 특징점 수에 대해 제한이 없다는 점을 들 수 있다. 사용자는 자유롭게 다량의 특징점을 추가하고 표현할 수 있다. 셋째 데이터 간의 상관관계와 동향을 표현할 수 있다. 이 중 가장 큰 이점은 특징점 사이에 상관관계를 발견할 수 있는 점이다. (그림 5)에서는 특정한 특징점을 기준으로 가장 상위 그룹을 검색한 결과 매우 유사한 상관관계의 데이터를 발견할 수 있음을 보여주고 있다[9].



(그림 5) 시각화를 통한 상관분석 발견

2.2.3 보안 시각화 기법

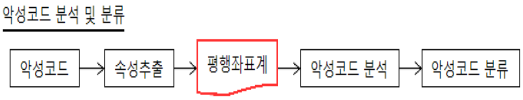
보안 시각화 기술은 매우 다양하다. 표현하는 데이터마다 고유한 방법으로 처리한다. 가장 공통적인 몇 가지 기술을 포함하는 것은 glyphs, color map, parallel coordinate plots, histograms, and scatter plots 등이 있다. 보안 시각화 기술은 표현하는 데이터의 기능에 따라 다른 유형으로 사용하지만 서로 다른 기술을 혼합하여 사용하기도 한다[11, 12].

각 기법을 통하여 네트워크 트래픽 정보 시각화 기술, 보안 경보 시각화 기술, 네트워크 공격 탐지 기술, 패턴분석을 통한 감지 기술 등 보안 시각화 기법의 다양한 연구가 진행되고 있다[11, 12].

3. 제안방법

3.1 개요

본 논문에서 제안하는 악성코드 분석 및 분류 기법은 안티 바이러스 업체에서 제공되는 속성을 추출한다. 확보한 속성을 수치화하여 수치화한 데이터를 Parallel Coordinates 상에 적용해 나타난 악성코드의 형태의 타입을 통해 악성코드 분석 및 재분류를 하는 기법을 제안한다.



(그림 6) 악성코드 분석/분류 순서도

(그림 6)과 같은 과정으로 안티 바이러스 업체에서 정의해 둔 악성코드 정보를 이용하여 각 업체마다의 필수적이며 공통적 항목을 속성으로 추출한다. 추출한 악성코드 속성을 평행좌표계로 표현하여 악성코드 분석을 통해 악성코드 타입을 정의한다.

3.2 데이터

3.2.1 악성코드의 속성 추출

악성코드는 안티 바이러스 업체에서 제공하는 악성코드 정보에서 추출한다. 악성코드 속성 추출은 각 업체 마다 공통적으로 제공하는 속성을 파악하여, 안철수 연구소에서 제공되는 악성코드에서 추출한다. 추출한 악성코드를 수치화 하는 속성 대상은 <표 2>에 대한 내용과 같다.

〈표 1〉 악성코드 속성 표현 기호

기 호	설 명
$M_{(x1,\ x2,\ \cdots,\ xn)}$	악성코드 식별 값 {x_n : 악성코드 속성}
x1	악성코드 종류
x2	활동 플랫폼
x3	감염/설치경로
x4	파일형태
x5	발견일
x6	대표적 증상 1
x7	대표적 증상 2

<표 2>의 기호는 제안하는 방법에 사용될 악성코드 속성을 나타내는데 사용된다. M{x1, x2, …, x7)는 악성코드 하나의 정보를 나타낸다. 하나의 악성코드 정보를 적용한다면 M은 수집하는 악성코드를 순차적으로 부여하는 고유한 index, 업체에서 명명한 악성코드 이름 순서로 조합으로 구성한다. 예를 들면 안철수연구소에서 제공되는 Win-adware/Rugo.233472에 앞에 index번호를 두어 1.Win-adware/Rugo.23347으로 표기한다. 이와 같이 표기한 악성코드에 따른 속성은 x1, x2, …으로 표기하는데 이는 각 악성코드의 속성의 내용을 수치화하여 저장할 때 사용하게 된다. 속성의 종류로는 <표 2>에서 나타내고 있는 악성코드 종류, 활동 플랫폼, 감염/설치 경로, 생성파일명, 파일형태, Registry, 발견일, 대표적 증상이다. 모든 업체에서 공통적으로 사용하고 있는 항목이며, 이 항목으로 악성코드의 내용을 파악할 수 있다.

〈표 2〉 각 악성코드 기호의 내용과 수치화

종 류	내 용	수치화
악성코드 종류	Trojan, Downloder, FILE VIRUS, WORM, DROPPER, KEYLOGGER, BOT, ADWARE/SPYWARE	순차적 값 부여
활동 플랫폼	DOS, Window, Linux, Unix, Palm, FreeBSD …	순차적 값 부여
감염/설치 경로	파일실행, 다운로드, 네트워크, 보안 취약성, 메일, 부팅, 다른악성코드, …	순차적 이진값 변환 후 중복값 표현
파일형태	exe, tmp, dll, bak, exebnb, dnt, sys, ini, ico, lin …	순차적 이진값 변환 후 중복값 표현
발견일	수집된 악성코드 최초 발견일~마지막 발견	순차적 값 부여
대표적 증상1	하드디스크관련, 파일관련, 시스템관련, 네트워크관련, 기타	순차적 이진값 변환 후 중복값 표현
대표적 증상2	보안상위험, 사생활침해가능, 사용자정보유출, 자동실행, 네트워크트래픽발생 …	순차적 이진값 변환 후 중복값 표현

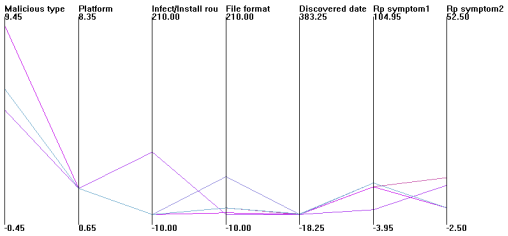
3.2.2 수치화

각 M마다 포함하고 있는 악성코드 속성에는 <표 3>의 내용을 나타내고 있다. 각 속성에 대한 수치화 방식은 다음과 같다.

<표 2>에서 나타내고 있는 악성코드 종류로는 x1에 적용되는 정보로서 mitre에서 수집하는 cve [13]에서 분류하는 악성코드 종류 9가지로 구분한다. Trojan, Downloder, FILE VIRUS, WORM, DROPPER, KEYLOGGER, BOT, ADWARE/SPYWARE를 나타내는데 사용된다. 각 값을 구분하기 위해서 각 종류에 따라 순차적인 수로 표현한다. x2 속성은 활동 플랫폼을 나타내는 속성으로 DOS를 비롯한 업체에서 구분하는 모든 플랫폼을 표현하는데 적용시킨다. 순차적 값을 부여하는 방식과 달리 이진 값으로 표현하는 감염 및 설치경로와 대표적 증상은 중복되는 내용이 있을 수 있기 때문에 이진 값으로 표현하여 중복 표현을 가능 하게 한다. 종류대로 순차적으로 이진값으로 표현한 후에 2^n으로 표현하면 중복되는 값도 표현할 수 있다. 예를 들면 대표적 증상으로 메일발송, 개인정보유출, 네트워크 속도 저하 등이 있다고 하면, 순차적으로 이진값을 부여한다. 메일 발송은 이진수 1값을 주고 개인정보유출은 이진수 10값을 부여하고 네트워크 속도 저하는 이진수 100을 순차적으로 부여한다. 메일 발송과 개인정보유출 증상이 같이 나타나는 악성코드라면 011값으로 표현한다. 이와 같이 표현한다면 다수의 중복된 수치도 효과적으로 표현할 수 있다.

3.3 Parallel Coordinates 적용 방법

Parallel Coordinates는 x축과 y축으로 구성된다. x축은 악성코드의 속성을 표현한다. y축은 x축에 적용된 속성의 수치를 나타내는데 사용한다.



(그림 7) Parallel Coordinates 구성요소 적용

x축은 수집한 악성코드 속성인 악성코드 종류, 활동 플랫폼, 감염/설치 경로, 파일형태, 발견일, 대표적증상1, 대표적증상2로 구성하며, y축은 수집한 악성코드 속성을 수치화 한 값을 x축에서 표현하고 있는 속성에 맞추어 표현한다. 구성하고 있는 각 속성이 가지고 있는 수치를 표현하는데 표현되며, 각 속성이 표현할 최소값과 최대값 영역을 확보한다.

3.4 악성코드 분석 및 분류

안티바이러스 업체에서 제공되는 악성코드 속성을 수집하여 수치화한 값을 Parallel Coordinates상에 표현한 내용을 기반으로 악성코드 간의 유사도를 분석한다. 또한 나타나는 뚜렷한 형태에 따른 악성코드 형태를 분류하고 시간에 따른 변화에 대한 분석과 함께 동향을 파악한다.

4. 실험 결과

4.1 실험 방법

본 논문에서 제안하는 Parallel Coodinates를 이용한 악성코드 분석 및 분류 방식은 안티 바이러스 업체인 안철수 연구소에서 제공되는 악성코드 속성을 수집하였다. 수집한 속성은 모든 안티 바이러스 업체에서 공통적으로 구분해 둔 속성으로 악성코드의 내용을 파악하기에 충분하다. 수집한 악성코드 속성을 각 속성에 맞게 수치화하여 확보하였다. 확보한 값을 다변량 데이터 분석에 효과적인 시각화 기법 중 하나인 Parallel Coordinates 상에 표현하였다.

4.2 실험 데이터 추출 결과

본 논문은 객관적인 자료 확보를 위해 국내의 대

〈표 3〉 악성코드

악성코드	M	P	I	F	D	R1	R2
1. Win-Trojan/ Backdoor.35281	1	2	17	11	2	26	7
2. Win-Spyware/ CyToday.81408	9	2	1	8	3	16	2
3. Win-Adware/Fast BSearch.732672	9	2	1	8	3	16	10
4. Win-Adware/You Porn.253952	9	2	1	1	3	16	2
5. Win-Adware/BHO. Cnsmin.61440	9	2	1	8	3	16	2
6. Win-Downloader/ Adload.201216	2	2	1	1	3	18	2
7. Win-Dropper/ Guidesh.764356	6	2	1	41	3	18	2
8. Win-Dropper/ Guidesh.741914	6	2	1	41	3	18	2
9. Win-Spyware/ Gever.249856	9	2	1	0	3	16	2
10. Win-Adware/ Lastlog.350720	9	2	1	192	3	16	2
11. Win-Spyware/ Gever.418304	9	2	1	0	3	16	2
12. Win32/IRCBot. worm.81408.I	5	2	67	1	1	4	8
13. Win-Trojan/ Agent43746.B	1	2	67	65	4	36	0
14. Win-Dropper/ ColorSoft.201140	6	2	1	8	4	18	2
15. Win-Adware/ ColorSoft.106496.B	9	2	1	0	4	16	2

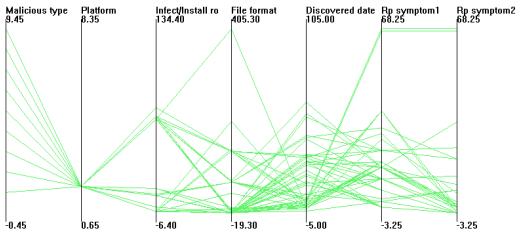
표적인 안티 바이러스 업체인 안철수 연구소에서 제공되는 악성코드 속성 정보를 이용하여 실험하였다. 실험 데이터는 제안한 방법을 통한 수치화를 하였다. 이는 악성코드 속성 정보를 단순히 수치화 하는 전처리 작업으로 악성코드의 속성이 그대로 표현하기에 충분하다. 확보한 실험 데이터의 일부를 <표 3>에서 나타내고 있다.

- M : Malicious type
- P : Platform
- I : Infect/Install route
- F : File format
- D : Discovered date
- R1 : Rp symptom1
- R2 : Rp symptom2

<표 3> 데이터를 근거로 제안된 XmdvTool을 통해 Parallel Coordinates로 표현을 한다.

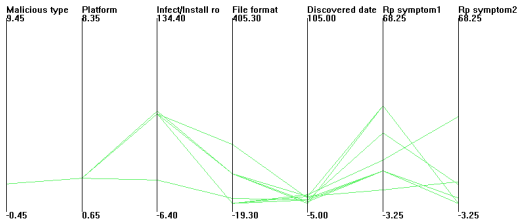
4.3 실험 결과 분석

수치화를 통한 데이터를 기반으로 Parallel Coordinates에 나타내었다. Parallel Coordinates는 XmdvTool 7.1을 이용하여 표현하였으며, 나타낸 결과는 다음과 같다.



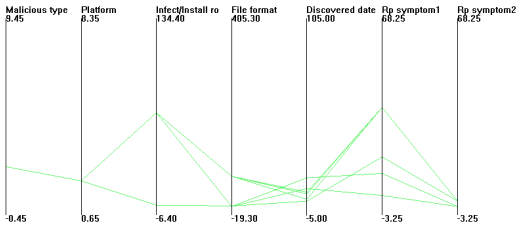
(그림 8) 악성코드 속성 시각화 실험 결과

(그림 8)은 Trojan, Downloder, FILE VIRUS, WORM, DROPPER, KEYLOGGER, BOT, ADWARE/SPYWARE 8가지 대한 악성코드와 아직 정해지지 않은 속성을 모두 나타낸 악성코드 전체적으로 나타낸 그래프이다. 전체적으로 살펴 볼 경우 한 눈에 알 수 있는 내용은 악성코드 종류와 활동 플랫폼은 한 가지라는 내용을 알 수 있다. 그 이외에는 정확하게 알 수 있는 내용이 없기에 각 악성코드만 나타내어 봤다.



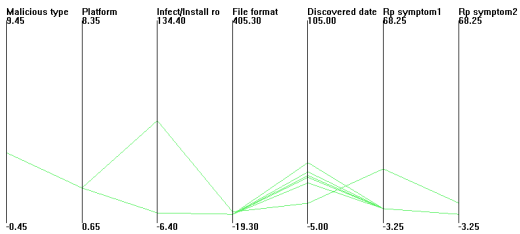
(그림 9) 트로이목마 시각화 그래프

트로이목마를 나타낸 그래프에서는 증상에 있어서 다양한 형태를 표현하고 있는 것을 알 수 있다. 다양한 형태를 가지고 있다는 것은 트로이목마의 악성코드가 많다는 것을 알 수 있었다.



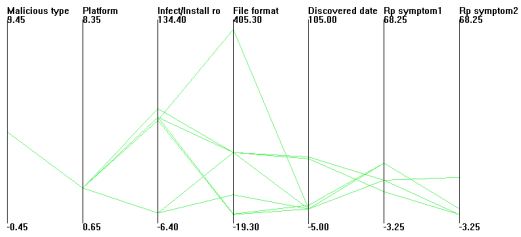
(그림 10) 다운로더 시각화 그래프

다운로더 또한 증상이 다양한 편에 속했으며, 트로이목마와 겹쳐져서 나타나는 편이 눈에 띄였다.



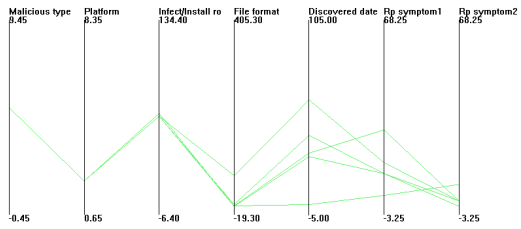
(그림 11) 바이러스 시각화 그래프

바이러스는 증상에 있어서 비슷한 증상이 중복되어 나타나는 것을 알 수 있었다.



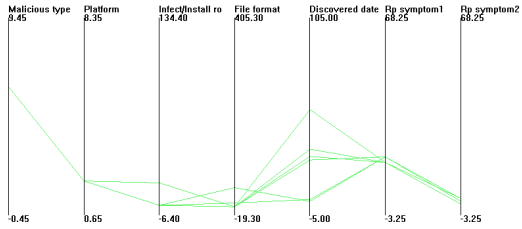
(그림 12) 웜 시각화 그래프

웜은 가장 다양한 경로로 유포되었으며 파일 포맷 형태도 다양하게 나타났다. 또한 시각적으로 가장 많은 영역을 차지하여 보였으며, 증상도 여러 가지가 중복되어 나타나는 점을 보였다.



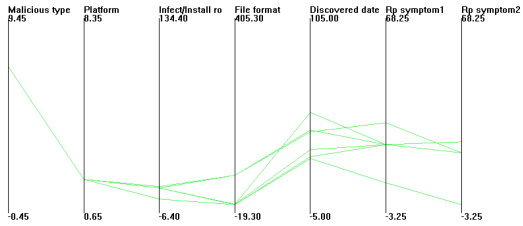
(그림 13) 드로퍼 시각화 그래프

드로퍼는 다른 악성코드에 비해 적은 양이 발견되었고 또한 감염 및 설치 경로와 파일 포맷 형태는 비교적 단순하게 나온 것을 알 수 있었다.



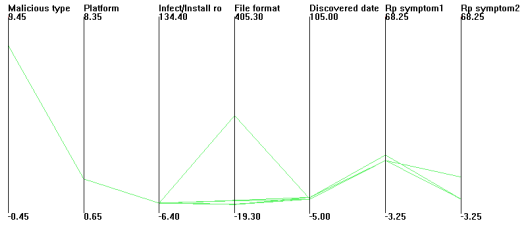
(그림 14) 키로거 시각화 그래프

키로거는 가장 뚜렷한 특징을 나타냈다. 증상에 있어서도 수집 시에 개인정보유출과 사생활 침해에 대한 내용으로 많이 좁혀졌듯이 그래프 상에서도 일관된 형태로 나타났다.



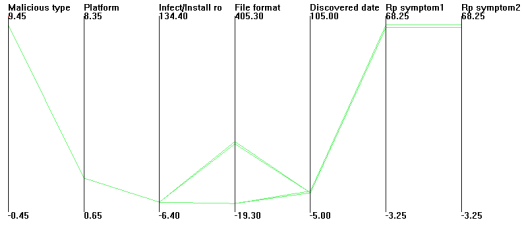
(그림 15) 봇 시각화 그래프

봇은 발견일이 어느 한때에 몰려있는 특징을 보였으며 증상에 있어서 비교적 단순한 형태를 띄고 있었다.



(그림 16) 에드웨어/스파이웨어 시각화 그래프

에드웨어 및 스파이웨어는 트로이목마를 비롯하여 많은 발견이 된 악성코드지만 비교적 단순한 증상을 보이고 있었다.



(그림 17) 미정된 악성코드 시각화 그래프

종류가 정해지지 않은 악성코드에 대한 속성 또

한 비슷한 형태를 띄고 있었다. 미정된 증상들은 정해진 번호로 전화를 걸게 한다거나 하는 증상으로 기존의 악성코드 종류에 속하지 않은 악성코드였다. 미정된 악성코드는 증상이 매우 단순한 형태로 구분됨을 알 수 있었다.

5. 결 론

본 논문은 악성코드 속성을 수집하여 수치화한 데이터를 시각화 기법 중 Parallel Coordinates를 통하여 악성코드 속성을 분석하고 분류하는 기법에 대한 연구를 하였다. 시각화 기법을 통한 다량의 데이터 분석은 많은 분야에서 활용되고 있지만 악성코드에 대한 분석은 연구되지 않았다. 악성코드를 나타낸 그래프를 통해 악성코드의 분석의 하나의 방법으로 제시하였다. 각 악성코드만을 나타낸 그래프를 통한 분석은 해당 악성코드에 대한 특징점을 볼 수 있었으며, 또한 악성코드마다 증상이 대부분 나뉘어져 나타났지만 중복되어 나타나는 악성코드가 있다는 것을 알 수 있었다. 알고리즘에 의해 분석되는 정적인 분석에 비해 시각화 기법을 통한 분석은 속성을 가지고 있는 악성코드는 모두 표현이 이론적으로 알 수 없었던 악성코드의 유사성을 그래프 상으로 볼 수 있었다. 알고리즘에 의한 분석과 분류와 혼용한다면 동향과 대책을 좀 더 효율적으로 대응할 수 있다.

본 논문의 제안 방법은 모든 속성을 표현할 수 있으며 그동안 눈에 보이지 않았던 미정된 악성코드에 대한 특징도 볼 수 있는 장점과 악성코드의 분석을 글이 아닌 시각화를 통해 보다 쉽게 파악할 수 있는 점이 장점이다. 하지만 상관관계의 정확한 임계치 제시와 연관성에 대한 정확한 증명이 어렵다. 향후 이에 대한 연구가 필요하며, 모든 안티 바이러스 업체에서 제공하는 속성을 통한 분석 연구가 필요하다.

참 고 문 헌

[1] E Skoudis and L Zeltser, “Malware : Fighting malicious code”, books.google.com, 2004.
[2] 안철수 연구소, “ASEC_Annual_Report”, 2009~2010.
[3] 장영준, 차민석, 정진성, 조시행, “악성코드 동향과 그 미래 전망”, 정보보호학회논문지, 제18권, 제3호, pp. 1-16.
[4] 서희석, 최중섭, 주필환, “윈도우 악성코드 분류 방법론의 설계”, 정보보호학회논문지, 제19권 제2호, pp. 88-92, 2009.
[5] 서희석, 최중섭, 주필환, “윈도우 악성코드 분류 시스템에 관한 연구”, 한국시뮬레이션학회논문지, 제18권, 제1호, pp. 63-70, 2009.
[6] 배성재, 권오철, 문종섭, 조재익, “Native API 빈도 기반의 퍼지 군집화를 이용한 악성코드 재그룹화 기법연구”, 정보보호학회논문지, 제18권, 제6호, pp. 115-127, 2008.
[7] Daniel A. Keim, “Information Visualization and Visual Data Mining”, IEEE transactions on visualization and computer graphics, Vol. 7, No. 1, 2002.
[8] MO Ward, “XmdvTool : Integrating multiple methods for visualizing multivariate data”, portal.acm.org, 1994.
[9] Stephen Few, “Multivariate Analysis Using Parallel Coordinates”, Perceptual edge, 2006.
[10] H Choi, H. Lee, “PCAV : Internet attack visualization on parallel coordinates”, Springer, Information and Communications Security, 2005.
[11] Rawiroj Robert Kasemsri, and Ying Zhu, “A Survey, Taxonomy, and Analysis of Network Security Visualization Techniques”, Citeseer, 2005.
[12] Nwokedi, Aditya P. Mathur, “A Survey of Malware Detection Techniques”, Citeseer, Purdue University, 2007.
[13] http://www.cve.mitre.org/.

송 인 수

2009년 남서울대학교 컴퓨터학과 (공학사)
2009년~현재 경기대학교 산업보안학과 석사과정

이 동 휘

2000년 경기대학교 전자계산학과 (이학사)
2003년 경기대학교 정보보호기술 공학과(공학석사)
2007년 경기대학교 정보보호학과 (정보보호학 박사)
현재 경기대학교 산업기술보호 특화센터 연구교수

김 귀 남

미국 캔자스대학(학사)
미국 콜로라도주립대학(석사)
미국 콜로라도주립대학(박사)
현재 경기대학교 산업보안학과 교수