http://dx.doi.org/10.5392/JKCA.2014.14.10.032

악성코드 유사도 측정 기법의 성능 평가 모델 개발

Development of a Performance Evaluation Model on Similarity Measurement Method of Malware

천성택*, 김희석**, 임광혁***, 김규일**, 서창호*
공주대학교 융합과학과*, 한국과학기술정보연구원 과학기술사이버안전센터**, 배재대학교 전자상거래학과***

Sung-Taek Chun(stchun@kisti.re.kr)*, HeeSeok Kim(hs@kisti.re.kr)**,
Kwang-Hyuk Im(khim@pcu.ac.kr)***, Kyu-il Kim(kisados@kisti.re.kr)**,
Chang-Ho Seo(chseo@Kongju.ac.kr)*

요약

날로 급증하는 대량의 악성코드들을 분류하여 악성코드에 대한 분석시간을 단축하고 신종의 악성코드를 발견하기 위한 악성코드 분류의 필요성이 대두됨에 따라 대량의 악성코드들을 분류하기 위한 다양한 악성코드 유사도 측정 기법이 제안되고 있다. 하지만 제안된 기존 연구들은 대부분 유사도 측정 기법을 소개하고 해당 기법에 의한 악성코드 분류 결과만을 제시하고 있으며, 다른 유사도 측정 기법과의 성능 비교 결과는 제시하지 않는다. 이는 유사도 측정 기법의 성능을 비교할 수 있는 평가 모델이 존재하지 않기 때문이다. 본 논문에서는 다양한 악성코드 유사도 측정 기법들의 성능을 비교 및 평가할 수 있는 악성코드 유사도 측정기법의 성능평가 모델로 성공확률과 신뢰도의 두 지표를 제안한다. 또한 본 논문에서는 두 지표를 이용해 기존 유사도 측정 기법들의 성능을 비교 및 평가한다.

■ 중심어 : | 악성코드 분류 | 유사도 측정 기법 | 정적 분석 | 동적 분석 | 허니팟 |

Abstract

While there is a great demand for malware classification to reduce the time required in malware analysis and find a new type of malware, various similarity measurement methods of malware to classify a lot of malwares have been proposed. But, the existing methods to measure similarity just represented the classification results by them and have not carried out performance comparison with other methods. This is because an evaluation model to compare the performance of similarity measurement methods is non-existent. In this paper, we propose a new performance evaluation model on similarity measurement methods of malware by using two indicators: success rate and degree of confidence. In addition, we compare and evaluate the performance of existing similarity measurement methods by using these two indicators.

■ keyword : | Malware Classification | Similarity Measurement Method | Static Analysis | Dynamic Analysis | Honeypot |

* 본 연구는 2014년도 한국과학기술정보연구원 창의연구과제인「대용량 보안 이벤트 자동검증 고도화 기술연구」의 지원을 받아 수행된 연구임(K-14-L06-C15-S01)

접수일자 : 2014년 08월 27일
수정일자 : 2014년 09월 11일
심사완료일 : 2014년 09월 11일
교신저자 : 김규일, e-mail : kisados@kisti.re.kr

I. 서 론

인터넷의 발전에 따라 웜·바이러스, 자료훼손 및 유출, 홈페이지 위·변조와 같은 사이버 해킹 공격의 수가 급속도로 증가하고 있으며 그 목적과 방법 또한 지능화·다양화되어지고 있다. 악성코드를 이용한 사이버 해킹 공격이 다양화됨에 따라 안티바이러스 제품 개발 업체와 연구자들은 해당 악성코드를 분석하고 특징을 파악하는데 많은 투자를 하고 있으며, 기존의 악성코드들과 유사한 악성코드들을 분류해 신종의 악성코드를 찾아내는 악성코드 분류기법에 대한 연구도 활발히 진행되고 있다.

악성코드 분류를 위해서는 악성코드와 기존 분류되어 있는 악성코드 그룹간의 유사도 측정을 필수적으로 수행해야 한다. 이 유사도 측정을 효율적으로 수행하기 위한 다양한 연구가 진행되어 왔으며, 유사도 측정에 사용할 악성코드 특징정보 추출 방법, 데이터 정규화 방법, 유사도 계산 방법의 세 단계에 초점을 두고 다양한 연구 결과들이 발표되어 왔다[1-5]. 하지만 유사도 측정 기법과 관련된 기존 연구들은 유사도 측정 기법 제안 후, 해당 기법에 의한 악성코드 유사도 측정 결과만을 실험결과로 제시하였고 다른 악성코드 분류 기법과의 성능 비교를 통한 공정한 평가를 수행하지 않았다.

본 논문에서는 다양한 악성코드 유사도 측정 기법들의 성능을 공정하게 평가하기 위한 성능평가 모델을 제안한다. 제안하는 성능평가 모델에서는 유사도 측정기법의 성능을 나타내는 지표로 성공확률(Success Rate)과 신뢰도(Degree of Confidence)를 이용한다. 성공확률이란 해당 유사도 측정 기법에 의해 악성코드들이 자신의 그룹으로 옳게 분류될 확률을 의미한다. x개의 악성코드에 대해 유사도 측정 기법에 의해 옳은 그룹으로 분류되는 악성코드의 개수가 y개일 때, 성공확률은 y/x가 된다. 유사도 측정 기법을 성능 평가하기 위한 두 번째 지표는 신뢰도이다. 신뢰도는 유사도 측정기법에 의한 분류 결과를 신뢰할 수 있는 정도를 의미하는 지표로서 특정 악성코드가 옳은 그룹으로 분류될 확률이 옳지 않은 그룹으로 분류될 확률과 차이가 나는 정도를 나타내는 지표이다.

본 논문에서는 제안하는 두 지표를 이용해 기존 악성코드 유사도 측정 기법들의 성능을 측정·비교하였다. 이 평가를 위해 본 논문에서는 가장 널리 사용되고 있는 세 개의 안티바이러스 제품 Microsoft, Avast, AVG의 탐지명으로 악성코드 그룹 다섯 개를 선택한 후, 기존 유사도 측정 기법들의 성능을 비교하였다[6]. 해당 실험을 통해 특징정보 추출, 데이터 정규화, 유사도 비교의 세 단계로 수행되는 유사도 측정 기법의 성능은 데이터 정규화 기법보다는 추출하는 특징 정보 및 유사도 비교 방법에 의해 결정됨을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2절은 악성코드 분석 기법 및 유사도 측정기법을 소개하고 3절에서는 악성코드 유사도 측정기법의 성능평가 모델을 제안한다. 해당 성능 평가 모델에 의한 기존 유사도 측정기법들의 성능 평가 결과는 4절에서 제시하며, 5절에서 본 논문을 결론짓는다.

II. 악성코드 유사도 측정 기법

1. 악성코드 분석

악성코드 분석은 정적 분석과 동적 분석의 두 가지 방법으로 구분되어진다. 정적 분석이 악성코드를 실행하지 않고 악성코드의 바이너리 값 혹은 기계어로 역변환 후 바이트 정보, 스트링 정보, 명령어 정보 등과 같은 악성코드의 특징 정보를 얻는 반면, 동적분석은 해당 코드를 직접 실행함에 의해 보다 유용한 특징 정보들을 얻어낸다. 동적 분석을 통해 얻어낼 수 있는 정보는 API 후킹 정보와 DNS, HTTP 요청과 같은 네트워크 분석 결과, 프로세스 정보 등이 있다.

2. 악성코드 유사도 측정 기법

악성코드 유사도 측정 기법이란 유사한 특성을 지닌 악성코드들로 구성된 악성코드 그룹과 특정 악성코드 사이의 유사도 비교를 통해 이 악성코드가 해당 그룹에 포함되는지 여부를 판단하는 방법을 의미한다.

유사한 특성을 지닌 악성코드들 Ψ_1, Ψ_2, ..., Ψ_t로 이

루어진 악성코드 그룹 G와 악성코드 M의 유사도 측정은 다음과 같은 세 단계의 과정을 통해 수행되어진다.

1. 악성코드 특징 추출(ψ): 악성코드 유사도 측정을 위한 첫 번째 과정은 악성코드로부터 유사도 비교에 사용될 특징을 추출하는 과정이다. 악성코드의 정적·동적 분석 결과로부터 얻은 다양한 정보(바이트, 스트링, 명령어, API 후킹 정보, 네트워크 분석 결과 등)로부터 유사도 측정에 사용할 특징을 추출하는 단계이다.
2. 데이터 정규화(η): 악성코드 유사도 측정을 위한 두 번째 단계는 추출된 특징들의 정규화 과정이다. 추출된 특징들은 문자열 혹은 API 명, IP, URL, Hostname 등의 악성코드 분석 결과인 원천 데이터이다. 데이터 정규화 과정에서는 이러한 추출 특징들을 유사도 비교 과정에 입력하여 사용할 수 있도록 데이터를 변형 및 정규화한다..
3. 유사도 비교(ζ): 유사도 비교 과정은 정규화된 악성코드 $\eta(\psi(M))$과 악성코드 그룹 G의 원소들을 정규화한 악성코드 그룹 $\eta(\psi(G)) = \{\eta(\psi(\Psi_i))|1 \le i \le t\}$간의 유사도 $\rho = \zeta(\eta(\psi(M)),\eta(\psi(G)))$를 측정하는 단계이다. 이러한 유사도 비교 단계는 악성코드 유사도 측정뿐 아니라 다양한 분야에서 연구되고 있다[7][8].

이러한 악성코드 유사도 측정 기법과 관련된 기존 연구에서 Bailey 등과 Rieck 등은 NCD (Normalized Compression Distance)를 이용한 악성코드 간 거리 계산 방법을 통한 분류 방법[1]과 Window API명에서 추출한 스트링을 통한 분류방법[2]을 각각 제안하였다. 최근에는 J. Nakazato 등이 2-gram 기법을 활용하여 Window API에 대한 호출 시퀀스의 빈도수를 계산하고 악성코드의 특징을 이 빈도수로부터 TF-IDF 기법을 이용해 추출한 후 유사도 비교를 수행하는 유사도 측정 기법을 제안하였다[3]. 이 방법들 이외에도 시퀀스의 존재여부를 0과 1로 표현하는 정규화하는 기법[4] 및 PCA를 통해 얻은 고유벡터에서 큰 값들의 위치로 확률 벡터의 원소를 추출 후 정규화하는 기법[5] 등이 위에서 명시한 3단계의 과정을 통해 악성코드 유사도를 측정·비교하였다.

III. 악성코드 유사도 측정 기법의 성능 평가 모델

본 절에서는 인터넷으로부터 수집된 대량의 악성코드들의 특징들을 이용해 이들을 그룹화하거나 새로운 신종의 악성코드를 찾아내기 위한 악성코드 유사도 측정 기법의 성능 평가 모델을 제안한다.

제안하는 유사도 측정 기법의 성능 평가 방법은 악성코드 특성에 따라 분류된 악성코드 그룹들 G_1, G_2, ..., G_n과 함께 수행되어진다. 각 그룹 G_i는 각각 k_i개의 악성코드로 다음과 같이 구성되어진다.

$$G_i = \{M_{iu}|1 \le u \le k_i\}$$

이 n개의 악성코드 그룹에 포함된 $\sum_{r=1}^{n} k_r$개의 악성코드와 n개의 그룹간의 유사도 측정 결과로부터 다음과 같은 n개의 유사도 측정 결과 행렬을 생성한다.

$$\Omega_r = (\rho_r(i,j))_{1 \le i \le k_r, 1 \le j \le n} \quad (1 \le r \le n)$$

$$\rho_r(i,j) = \begin{cases} \zeta(\eta(\psi(M_{ri})),\eta(\psi(G_j - \{M_{ri}\}))) & (j = r) \\ \zeta(\eta(\psi(M_{ri})),\eta(\psi(G_j))) & (j \ne r) \end{cases}$$

각 행렬 Ω_r의 i번째 행의 의미는 G_r의 i번째 원소 M_{ri}와 각 그룹 G_1, G_2, ..., G_n 간의 유사도를 의미한다. 단, M_{ri}와 G_r의 유사도 $\rho_r(i,r)$을 계산할 경우에는 그룹 G_r에서 M_{ri}를 뺀 $G_r - \{M_{ri}\}$와 M_{ri}의 유사도를 측정한다.

두 번째 단계에서는 유사도 측정 결과인 각 행렬 Ω_r을 토대로 순위 행렬 R_r을 생성한다. R_r의 각 행은 Ω_r의 각 행에 해당하는 유사도 결과에 대해 순위를 정해 그 순위를 원소로 구성한다. 예를 들어, 상관계수(Pearson Correlation Coefficient)와 같이 유사도 비교 결과 값이 큰 경우 유사하다고 판정되는 유사도 측정 방법을 사용했을 경우, Ω_r에 대해 R_r은 다음과 같이 구성되어진다.

$$\Omega_r = \begin{pmatrix} 0.3 & 0.1 & 0.2 & 0.5 \\ 0.2 & 0.5 & 0.4 & 0.7 \\ 0.1 & 0 & 0.3 & 0.2 \end{pmatrix} \rightarrow R_r = \begin{pmatrix} 2 & 4 & 3 & 1 \\ 4 & 2 & 3 & 1 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$

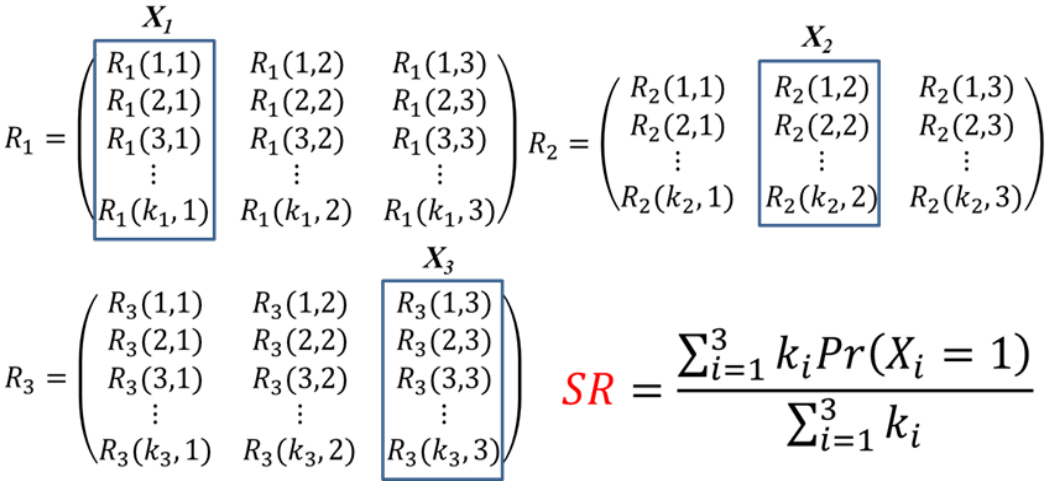


그림 1. 성공확률 (SR) 연산 방법

본 논문에서는 결과 행렬 Ω_r과 순위 행렬 R_r을 이용해 유사도 측정 기법의 성능을 평가한다. 각 악성코드를 n개의 그룹과 유사도 측정 했을 때, 악성코드가 실제 포함되는 그룹과의 유사도가 가장 높은 경우, 유사도 측정 기법의 성능이 좋은 것으로 직관적인 판단을 할 수 있다. 따라서 이러한 판단을 근거로 유사도 측정 기법의 성능을 수치화 할 수 있는 성공확률을 계산한다면 다음과 같다.

$$Success\ Rate\,(SR) = \frac{\sum_{r=1}^{n}\sum_{i=1}^{k_r}\Theta_r(i)}{\sum_{r=1}^{n}k_r}$$

위의 식에서 $\Theta_r(i)$는 다음과 같이 정의된다.

$$\Theta_r(i) = \begin{cases} 1 & if\ R_r(i,r) = 1 \\ 0 & otherwise \end{cases}$$

악성코드 그룹이 총 3개일 경우, 유사도 측정 기법의 성공확률을 연산하는 과정을 도식화하면 [그림 1]과 같다.

본 논문에서는 성공확률과 함께 또 다른 평가 지표로 유사도 측정 기법의 신뢰도를 제안한다. 신뢰도는 같은 그룹에 포함된 악성코드들 간의 유사도 결과 값이 다른 그룹에 포함된 악성코드들 간의 유사도 결과 값과 구분되는 정도를 나타낸다.

신뢰도에 대한 정확한 연산 방법은 다음과 같다.

1. 우선, Ω_r의 각 열의 원소로 구성되는 n개의 그룹 X_{r1}, X_{r2}, ..., X_{rn}이 각각 정규분포 $N(m_{r1},\sigma_{r1}^2)$, $N(m_{r2},\sigma_{r2}^2)$, ..., $N(m_{rn},\sigma_{rn}^2)$를 따른다고 가정한다. 이 가정은 각 그룹의 원소의 개수가 충분히 클 경우 사실이다. 이 n개의 그룹에서 X_{rr}을 제외한 나머지 $n-1$개의 그룹은 다른 그룹에 포함된 악성코드들 간의 유사도를 원소로 하는 그룹이다.
2. 두 번째 단계에서는 같은 그룹에 포함된 악성코드들 간의 유사도 값이 다른 그룹에 포함된 악성코드들 간의 유사도들에 해당하는 정규분포보다 클 확률을 모두 계산한다. 즉, 각 Ω_r에서 같은 그룹에 포함된 악성코드들 간의 유사도를 의미하는 r번째 열에 있는 k_r개의 원소들이 나머지 $n-1$개의 정규분포 각각보다 클 확률을 계산한다. 각 Ω_r에 대해 총 $k_r(n-1)$개의 확률이 계산되어질 수 있다.
3. 두 번째 단계를 반복하여 모든 Ω_r에 대해 이 확률을 계산한 후, 이 값들의 평균 값을 계산해 유사도 측정 기법의 신뢰도로 정의한다.

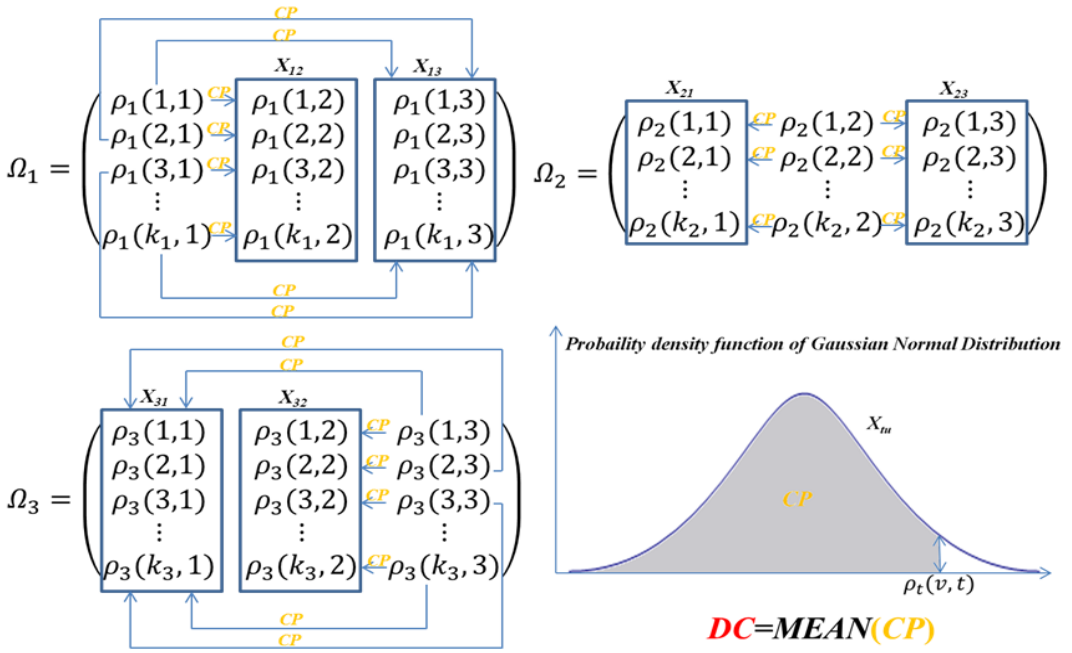


그림 2. 신뢰도 (DC) 연산 방법

표 1. 신뢰도 (DC) 계산 알고리즘

Input: $\Omega_r (1 \le r \le n)$

Output: DC

1. For r=1 to n and for j=1 to n

1.1. $X_{rj}=\{\rho_r(i,j)|1 \le i \le k_r\} \sim N(m_{rj}, \sigma_{rj}^2)$

1.2. $m_{rj}=E(X_{rj})$, $\sigma_{rj}^2=V(X_{rj})$

2. CP=0

3. For r=1 to n

3.1. For j=1 to n

3.1.1. If j==r then j=j+1

3.1.2. $CP=CP+\sum_{i=1}^{k_r} f(X_{rj} < \rho_r(i,r))$

(f는 정규분포의 확률밀도함수)

4. Return $DC=\dfrac{CP}{\sum_{r=1}^{n} k_r(n-1)}$

[그림 2]는 악성코드 그룹이 총 3개일 때 신뢰도 DC를 연산하는 과정을 도식화 한 것이며 [표 1]은 신뢰도를 연산하는 과정을 일반화한 것이다.

IV. 기존 유사도 측정기법들의 성능평가

본 절에서는 3절에서 제안한 유사도 측정기법의 성능평가 모델을 이용해 기존 제안된 유사도 측정기법들의 성능을 평가·비교한다.

우선 본 논문에서 성능평가에 활용된 악성코드들은 자체 구축한 허니팟을 통해 수집하였다. 또한 수집된 악성코드들을 가장 널리 사용되고 있는 3개의 안티바이러스 제품(Microsoft, Avast, AVG)의 진단명에 의해 분류한 후, [표 2]와 같은 5개의 악성코드 그룹을 선택한다[6].

다섯 그룹에 포함된 악성코드들의 특징을 추출하기 위해 악성코드들에 대한 정적 분석 및 동적 분석을 수행하였다. 해당 분석은 허니팟 프로젝트에 의해 개발된

Cuckoo 시스템[9]을 통해 수행하였으며 이로부터 API 호출 정보, Network 분석 정보, 포함 문자열 등 다양한 악성코드 정보를 수집할 수 있었다.

표 2. 악성코드 그룹

Group	Microsoft	Avast	AVG
G_1	PWS:Win32/OnLineGames.xx	Win32:Crypt-IGU[Trj]	Win32/Patched.FZ
G_2	Virus:Win32/Sality.G	Win32:Sality-U	Win32/Sality
G_3	Trojan:Win32/Yoddos.A	Win32:Trojan-gen	Dropper.Agent.RCT
G_4	TrojanDropper:Win32/Oficla.x	Win32:Malware-gen	Genericxx.xxxx
G_5	Trojan:Win32/Startpage.xx	NSIS:StartPage-N[Trj]	Dropper.NSIS.D

이 다섯 그룹에 포함된 악성코드들과 함께 평가하고자 하는 유사도 측정기법들의 목록은 [표 3]과 같다. 본 논문에서 평가하고자 하는 7가지의 유사도 측정기법은 모두 특징추출함수로 API 시퀀스 정보를 활용하였다. 즉, 악성코드의 API 호출 리스트에서 API의 호출 전후 관계를 특징으로 추출하였다. 본 논문에서 이용하는 악성코드의 정규화 기법은 총 4가지 방법으로 API 시퀀스에 대한 빈도수를 확률 벡터로 정규화하는 기법, 이 확률 벡터를 PCA로 압축하는 기법, [5]의 방법과 같이 PCA를 통해 얻은 고유벡터에서 큰 값들의 위치로 확률 벡터의 원소를 추출 후 정규화하는 기법, 시퀀스의 존재여부를 0과 1로 표현하는 정규화하는 기법[4]을 활용하였다.

표 3. 유사도 측정 기법

구분	특징추출(ψ)	정규화(η)	유사도 비교(ζ)
S_1	API 시퀀스	빈도수에 대한 확률 벡터	상관계수
S_2			유클리디안 거리
S_3		확률 벡터를 PCA로 압축	상관계수
S_4			유클리디안 거리
S_5		빈도수에 대한 [5]의 정규화 기법	상관계수
S_6			유클리디안 거리
S_7[4]		시퀀스의 존재여부를 0과 1로 표현한 벡터	$\frac{2\lvert X\cap Y\rvert}{\lvert X\rvert+\lvert Y\rvert}$

[그림 3]은 악성코드 M으로부터 동적분석에 의해 API 호출리스트를 얻어내고 이로부터 API 시퀀스

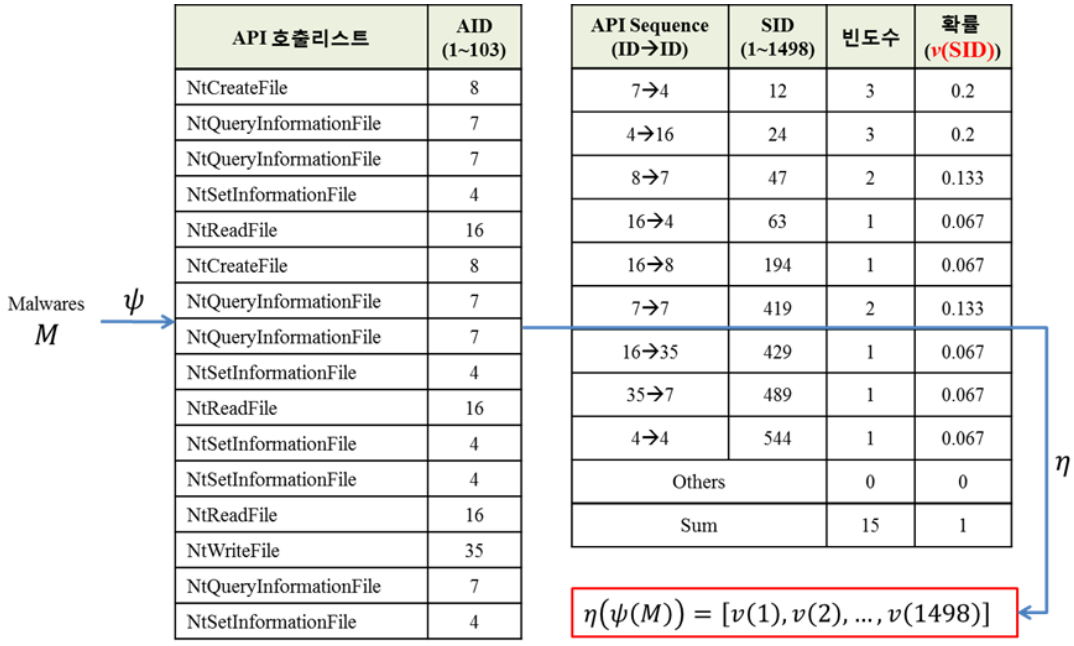


그림 3. 수집 악성코드의 특징추출 및 정규화 방법 (S_1, S_2)

(API 전후 호출관계) 및 빈도수를 추출 후, 각 시퀀스에 대한 확률을 계산해 정규화된 벡터 $[v(1), v(2),..., v(1498)]$를 얻어내는 일련의 과정을 나타낸다. [그림 3]에서 AID는 수집된 전체 악성코드들에 대해 호출된 이력이 있는 API의 ID를 의미하며 SID는 전체 악성코드들의 API 시퀀스에서 존재하는 이력이 있는 시퀀스들에 대한 ID를 의미한다. 본 논문에서 실험데이터로 선정된 악성코드들에 대해 호출된 이력이 있는 API는 총 103개였으며, 전체 API 호출리스트에 존재하는 API 시퀀스는 1498개가 존재하였다. 이 외에 PCA를 이용한 데이터 압축은 상위 고유값 100개에 대한 고유벡터를 활용하여 수행하였다. 특정 악성코드 M과 특정 그룹 G간의 유사도 비교값 $\zeta(\eta(\psi(M)), \eta(\psi(G)))$ 은 M과 그룹 G의 원소 각각에 대한 유사도를 계산 후, 이에 대한 평균값으로 유사도 비교 값을 결정했다.

제안하는 평가 모델을 활용하여 유사도 측정 기법의 성능을 평가했을 때, 안티바이러스 제품에 의해 분류된 악성코드 그룹들은 비교적 잘 분류가 되어 있음을 확인할 수 있었다. 7가지의 유사도 측정 기법 중 5가지의 기법이 성공확률(SR)로서 100%의 값을 나타냈으며 유사도 비교 방법으로 유클리디안 거리(Euclidean Distance)를 사용하는 두 개의 유사도 측정기법 S_4, S_6만이 98.61%의 성공확률을 보였다. 이는 G_4에 포함된 일부 원소들이 자기 자신이 포함된 그룹이 아닌 다섯 번째 그룹 G_5와의 유사도가 더 높은 것으로 나타났기 때문이다.

이미 분류가 잘 되어있는 악성코드 그룹들에 대한 유사도 측정기법들의 성능은 위의 결과에서 보이는 바와 같이 성공확률 값만으로 비교가 어려움을 확인할 수 있었다. 본 논문에서는 유사도 측정 기법의 성능을 측정하는 두 번째 지표인 신뢰도(DC) 연산을 통해 좀 더 정밀하게 성능을 평가하였다. 각 유사도 측정 기법에 대한 성공확률(SR)과 신뢰도(DC)는 다음과 같다.

[표 4]와 [그림 4]에서 보는 바와 같이 신뢰도가 가장 큰 유사도 측정 기법 S_1이 성능이 가장 좋은 것으로 평가되었으며 S_1, S_3, S_5의 성능이 서로 유사하고 S_2, S_4, S_6의 성능이 서로 유사함을 확인할 수 있었다. 이는 악성코드 유사도 측정 기법의 성능은 악성코드 정규화 기법보다는 특징 추출과 유사도 비교 방법에 의존함을 의미한다.

표 4. 유사도 측정 기법의 성능 비교

유사도 측정기법	SR	DC
S_1	100%	1-4.09e-07
S_2	100%	1-4.89e-02
S_3	100%	1-2.34e-06
S_4	98.61%	1-4.89e-02
S_5	100%	1-4.26e-05
S_6	98.61%	1-5.00e-02
S_7	100%	1-1.87e-02

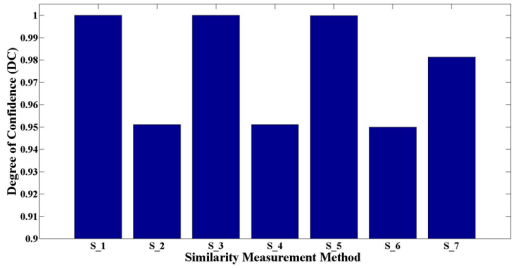


그림 4. 신뢰도 (DC)

V. 결 론

본 논문에서는 다양한 유사도 측정 기법들을 대상으로 공정한 성능 평가를 수행할 수 있는 악성코드 유사도 측정 기법의 성능 평가 모델을 제안하였다. 제안하는 평가모델에서는 각 유사도 측정 기법들의 성능을 나타내는 지표로 성공확률과 신뢰도를 이용하였다. 또한 사례 연구로 허니팟으로부터 수집된 악성코드들을 기존 유사도 측정 기법들에 의해 분류하고, 제안하는 모델의 두 지표를 활용해 유사도 측정 기법들의 성능을 평가 및 비교하였다. 해당 평가의 결과로 악성코드 유사도 측정 기법의 성능이 악성코드 정규화 기법보다는 특징 추출 및 유사도 비교 방법에 의존함을 알 수 있었으며, 향후, 본 연구의 결과인 성능 평가 방법을 활용해 기 제안되었거나 향후 제안될 악성코드 유사도 측정기

법들의 성능이 공정하게 비교·평가될 수 있을 것으로 기대되어진다.

참 고 문 헌

[1] M. Bailey, J. Oberheide, J. Andersen, and Z. M. Mao, "Automated classification and analysis of Internet malware," RAID 2007, LNCS 4637, Springer-Verlag, pp.178-197, 2007.

[2] K. Rieck, T. Holz, C.Willems, P. Dussel, and P. Laskov, "Learning and classification of malware behavior," DIMVA 2008, LNCS 5137, Springer-Verlag, pp.108-125, 2008.

[3] J. Nakazato, J. Song, M. Eto, D. Inoue, and K. Nakao, "A novel malware clustering method using frequency of function call traces in parallel threads," IEICE Trans. on Inf. And Syst., Vol.E94-D, No.11, pp.2150-2158, 2011.

[4] K. Iwamoto and K. Wasaki, "Malware Classification based on Extracted API Sequences using Static Analysis," AINTEC 2012, ACM, pp.31-38, 2012.

[5] V. P., H. Jain, Y. K. Golecha, M. S. Gaur, and V. Laxmi, "Medusa: MEtamorphic Malware Dynamic Analysis Using Signature from API," SIN 2010, ACM, pp.263-269, 2010.

[6] http://www.opswat.com/about/media/reports/antivirus-january-2014

[7] 김성환, 조환규, "PAM 행렬 모델을 이용한 음소간 유사도 자동 계산 기법", 한국콘텐츠학회논문지, 제12권, 제3호, pp.34-43, 2012.

[8] 유주원, 김종원, 최종욱, 배경율, "개선된 비디오 장면 유사도 검출 알고리즘", 한국콘텐츠학회논문지, 제9권, 제2호, pp.43-50, 2009.

[9] http://www.cuckoosandbox.org/

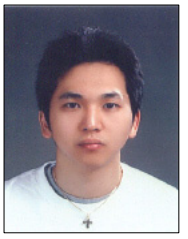
저 자 소 개

천 성 택(Sung-Taek Chun) 정회원

- 2006년 2월 : 경동대학교 전자상거래과 졸업
- 2012년 2월 : 공주대학교 바이오정보공학과 석사 졸업
- 2012년 3월 ~ 현재 : 공주대학교 융합과학과 박사과정 수료

<관심분야> : 내용기반 영상검색, 멀티미디어

김 희 석(HeeSeok Kim) 정회원



- 2006년 : 연세대학교 수학과(학사)
- 2008년 : 고려대학교 정보보호대학원(공학석사)
- 2011년 : 고려대학교 정보보호대학원(공학박사)
- 2011년 ~ 2012년 : Bristol University 박사후 연구원
- 2013년~현재 : 한국과학기술정보연구원 (KISTI) 과학기술정보보호실 선임연구원

<관심분야>: 부채널 공격, 암호시스템 안전성 분석 및 고속구현, 암호칩 설계 기술, 보안관제, 네트워크 보안

임 광 혁(Kwang-Hyuk Im) 정회원

- 1995년 2월 : 한국과학기술원 전산학과(공학사)
- 2000년 2월 : 한국과학기술원 산업공학과(공학석사)
- 2006년 2월 : 한국과학기술원 산업공학과(공학박사)
- 2006년 3월 ~ 2008년 2월 : 삼성전자 반도체연구소 책임연구원
- 2008년 3월 ~ 현재 : 배재대학교 전자상거래학과 부교수

<관심분야> : 지식서비스, 경영정보시스템, 데이터마이닝, 지능정보시스템, 전자상거래, 고객관계관리

김 규 일(Kyu-il Kim) 정회원

- 2005년 2월 : 성균관대학교 컴퓨터공학과 석사(공학석사)
- 2010년 2월 : 성균관대학교 컴퓨터공학과 박사(공학박사)
- 2010년 6월 ~ 현재 : 한국과학기술정보연구원 과학기술정보보호실 선임연구원

<관심분야> : 보안관제, 침해사고대응, 악성코드 분석

서 창 호(Chang-Ho Seo) 정회원

- 1992년 2월 : 고려대학교 수학과(이학석사)
- 1996년 2월 : 고려대학교 수학과(이학박사)
- 1996년 3월 ~ 1999년 : 한국전자통신연구원 선임연구원
- 2000년 ~ 현재 : 공주대학교 응용수학과 교수

<관심분야> : 암호 알고리즘, PKI, 무선 보안 등