데이터 마이닝 기법을 이용한 소규모 악성코드 탐지에 관한 연구★

이 택 현*, 국 광 호**

요 약

최근 인터넷 기술을 악용하는 행위로 인하여 경제적, 정신적 피해가 증가하고 있다. 특히, 신규로 제작되거나 변형된 악성코드는 기존의 정보보호 체계를 우회하여 사이버 보안 위협의 기본 수단으로 활용되고 있다. 이를 억제하기 위한 다양한 연구가 진행되었지만, 실제 악성코드의 많은 비중을 차지하는 소규모 실행 파일에 대한 연구는 미진한 편이다. 본 연구에서는 기존에 알려진 소규모 실행 파일의 특징을 데이터마이닝 기법으로 분석하여 알려지지 않은 악성코드 탐지에 활용할 수 있는 모델을 제안한다. 데이터 마이닝 분석 기법에는 나이브베이지안, SVM, 의사결정나무, 랜덤포레스트, 인공신경망 등 다양하게 수행하였으며, 바이러스토탈의 악성코드 검출 수준에 따라서 개별적으로 정확도를 비교하였다. 결과적으로 분석 파일 34,646개에 대하여 80% 이상의 분류 정확도를 검증하였다.

A Study on Detection of Small Size Malicious Code using Data Mining Method

Taek-Hyun Lee*, Kwang-Ho Kook**

ABSTRACT

Recently, the abuse of Internet technology has caused economic and mental harm to society as a whole. Especially, malicious code that is newly created or modified is used as a basic means of various application hacking and cyber security threats by bypassing the existing information protection system. However, research on small-capacity executable files that occupy a large portion of actual malicious code is rather limited. In this paper, we propose a model that can analyze the characteristics of known small capacity executable files by using data mining techniques and to use them for detecting unknown malicious codes. Data mining analysis techniques were performed in various ways such as Naive Bayesian, SVM, decision tree, random forest, artificial neural network, and the accuracy was compared according to the detection level of virustotal. As a result, more than 80% classification accuracy was verified for 34,646 analysis files.

Keywords : Data Mining, Malware, Dynamic Analysis, Static Analysis, Anti-Virus

접수일(2018년 11월 30일), 수정일(1차: 2019년 12월 27일), 게재확정일(2019년 3월 23일)

* 주저자 : 서울과학기술대학교 IT정책전문대학 산업정보시스템
** 교신저자 : 서울과학기술대학교 기술경영융합대학 글로벌융합산업공학과

1. 서 론

국내 인터넷 이용자는 2017년에 4,428만명(90.3%)이며, 스마트폰의 이용자는 2017년에 3,800만명(77.7%)으로 세계 6위를 기록하고 있다[2]. 이는 대다수 국민이 온라인 상태를 유지하거나, 인터넷에 즉시 접속할 수 있음을 의미한다. 이러한 정보통신 환경의 변화는 초 연결사회(Hyper Connected Society)의 진입을 앞당기고 있다. 또한, 국가 기반시설을 포함한 다양한 산업 영역이 IT 인프라로 연결되면서, 새로운 사업적 기회를 창출하고 자원의 효율성을 높이는 등의 긍정적인 효과를 가져왔다[16].

하지만 세계경제포럼(World Economic Forum)에서 발간된 '글로벌 리스크 2017' 보고서에 의하면 전 지구적으로 발생 가능성이 높은 위협으로 '사이버 공격(6위)'를 언급하였으며, '글로벌 리스크 2014'는 파급력이 큰 잠재적인 위협으로 '주요 정보 인프라 붕괴(5위)'를 제시하였다. 이러한 전망은 디지털 재난이 미래형 재난으로 부각될 것을 의미한다[12][24].

또한, 미국의 보안 회사인 시만텍의 2013년 보고서에 의하면 전 세계 사이버 범죄에 의한 연간 피해액이 $1,130억으로, 사이버 피해당 평균 $298의 막대한 피해가 발생하였다고 한다[23]. 국내에서도 다수의 주요 정보를 보유하고 있는 포탈, 금융, 국방, 국가기간망 등에서 무차별적인 해킹 공격이 발생하여 사회 전반에 심각한 피해가 발생하였다[4].

이러한 사이버 범죄에서 컴퓨터 악성 코드는 주요 정보 탈취, 자원 도용 등과 같은 다양한 사이버 공격의 기본 수단으로 활용되고 있다. 특히, 신규로 제작되거나 변종된 악성코드는 기존의 정보보호체계를 우회하여 지속적으로 IT인프라 환경을 위협하며 사이버범죄에 이용되고 있다. 이를 예방하기 위하여 다양한 연구가 진행되었으나, 실제 악성 코드의 많은 비중을 차지하는 소규모 EXE파 일(1M 바이트 미만)을 식별하기 위한 연구는 미진한 편이다.

본 논문에서는 실제의 악성코드에서 많은 비중을 차지하는 소규모 EXE 파일(1Mbyte 미만)의 특징을 데이터마이닝 기법으로 분석하여 기존의 정보보호체계에서 탐지할 수 없었던 신종 및 변종 악성코드를 사전에 식별할 수 있는 최적의 모델을 제시한다.

논문의 구성은 다음과 같다. 2장은 이론적 배경과 관련 연구를 소개한다. 3장은 데이터마이닝 실행 결과에 대해 설명한다. 4장은 연구 결과를 요약하고 향후 연구 방향을 제시한다.

2. 이론적 배경과 관련 연구

2.1 최근 악성코드 동향

인터넷이 활성화되지 않았던 시기에는 단순 호기심 혹은 저작권을 보호하기 위해서 악성코드를 개발하였다. 그러나 최근에는 금전적, 정치적, 군사적 목적으로 바이러스가 제작되어 사회 인프라 전반을 위협하고 있다[4].

독일의 보안전문업체인 AV-TEST에 의하면 전체 악성코드의 개수는 2012년 99백만개 에서 2017년 719백개로 약 7.2배 증가하였다. 이는 신종 악성코드에 의한 위협이 지속해서 증가하는 것을 보여준다[19].

미국의 온라인 바이러스 분석 업체인 바이러스 토탈에 의하면 2018년 8월에 7일 동안 바이러스 유무를 검사한 파일중 상위 확장자 파일이 EXE인 파일이 3,207.303개(44.05%), Android인 파일이 657,730개(9.03%), PDF인 파일이 606,351개(8.33%)로 나타났으며, EXE 확장자가 가장 높은 바이러스 의심군인 것을 추정할 수 있다[25].

미국의 글로벌 컴퓨터 보안업체인 맥아피에 의하면 알려진 악성파일의 90.7%가 1M 바이트 미만의 크기의 가지고 있으며, 0.03%의 파일만이 25M 바이트보다 큰 크기를 가지는 것으로 보고하고 있다[20].

데스크톱 운영체제의 시장 점유율과 웹 브라우저 통계 정보를 제공하는 넷애플리케이션(Net Ap

plication)에 의하면 마이크로소프트 Windows OS의 점유율이 최근 3년간 90% 이상을 유지하며 Mac OS X와 Linux OS를 압도하고 있다[24].

결과적으로 악성코드에 의한 사이버 위협이 지속적으로 증가하고 있으며, 마이크로소프트 Windows 운영체제의 1M 바이트 이하의 EXE 실행 파일이 주요한 바이러스 대상 군이라는 것을 추정할 수 있다.

2.2 악성코드 분석에 관한 연구

악성코드에 의한 피해를 줄이기 위해서 다양한 연구가 진행되어 왔다. 기존에 알려진 악성코드는 안티바이러스 백신, 온라인 분석 시스템, 시그니처 분석 등을 통해서 식별할 수 있으며, 알려지지 않은 악성코드는 정적 분석, 동적 분석, 하이브리드 분석으로 유해성을 식별할 수 있다[15].

정적 분석은 파일을 수행하지 않고 내부 구조를 깊이 있게 분석할 수 있다. 그러나, 숙련된 기술이 필요하며, 패킹 기법으로 파일이 보호되어 있을 경우에 분석이 어려운 단점이 있다. 동적 분석은 악성코드의 실행 전 후의 상태 변화를 관찰하여 유해성을 식별한다. 하지만, 동적 분석 과정에서 실험 환경이 악성코드에 감염될 수 있으며, 분석되는 행위를 관찰하는데 일정 시간이 소요되는 단점이 있다. 이를 보완하기 위하여 온라인 분석, 하이브리드 분석 등을 수행한다.

이들과 관련된 연구 결과들은 다음과 같다. 한경수 등은 정적 분석을 이용하여 실행 파일에 내장된 API를 분석하여 악성코드를 식별하였으며[17], 강태우 등은 동적 분석을 이용하여 파일 실행 시에 발생하는 API를 분석하여 악성코드를 식별하였다[1][7][9]. G. Wagener 는 가상환경 탐지를 위해서 사용하는 함수 유무를 파악하여 악성코드를 식별하였다[21]. 정용욱은 정적 분석과 동적 분석으로 파일의 특성을 추출하여 AHP분석으로 악성코드를 분류하였다[15]. 배철민 등은 온라인분석으로 바이러스 토탈에 파일 해쉬값을 조회하여 악성 여부를 식별하였다[8][11].

선행 연구를 통하여 다양한 유형의 바이러스 분석 방법을 살펴보았다. 하지만, 악성코드의 대다수를 차지하는 소규모 파일에 대한 식별 연구는 미진하였다. 따라서, 본 연구에서는 소규모 악성코드에 대한 식별을 위해서 온라인 분석, 정적 분석, 동적 분석을 혼합하여 메타 데이터를 생성하고, 이들 데이터에 데이터 마이닝 기법을 적용하여 바이러스를 식별하는 모델을 설계한다.

2.3 데이터마이닝에 관한 연구

데이터마이닝은 방대한 데이터에서 의사 결정에 도움이 되는 일정한 규칙과 패턴을 밝혀내는 과정이다. 본 논문에서 바이러스를 식별하고자 데이터마이닝 분석 기법 중 나이브 베이즈 분류, 서포트 벡터 머신, 의사결정나무, 랜덤포레스트, 인공신경망 등을 이용하였다.

나이브 베이즈 분류(Naive Bayes Classification)는 특성값들이 서로 독립이라고 가정하고 베이즈 정리를 적용하여 가장 사후 확률이 높은 범주로 분류하는 확률 분류기의 일종으로 1950년대 이후 광범위하게 연구되고 있다. 나이브 베이즈 분류는 스팸 메일 분류와 같은 문서 분류 혹은 범주 분류에 효율적으로 적용될 수 있다[27].

서포트 벡터 머신(Support Vector Machine)은 원래 데이터를 커널을 이용하여 고차원으로 변환한 후 데이터를 최적으로 분리하는 초평면을 찾는다. 데이터를 고차원으로 변환하는 이유는 선형분리를 가능하게 하기 위해서이다. 이때 서포트 벡터 머신은 미래 데이터의 분류 오류를 최소화하기 위해 클래스들 간의 거리를 최대로 하는 최대 마진 초평면을 찾아 분류한다.

의사결정나무(Decision Tree)는 몇개의 입력변수들에 기초하여 목표변수의 값을 예측하는 모델로서 분류, 예측, 차원축소 및 변수선택 등에 다양하게 활용된다[14]. 대표적인 알고리즘에는 CART(Chi-squared Automatic Interaction Detection), CHAID(Chi-squared Automatic Interaction Detection) 등이 있다[3].

CART 알고리즘은 다음 식과 같이 정의되는 지니계수(Gini Index)로 노드의 불순도(impurity)

를 측정하며 각 노드에서는 불순도를 최대한 감소시키는 변수를 토대로 이진 분리를 수행한다. 범주의 수가 m개이고 특정노드의 총 관측수가 n, k번째 범주의 관측수가 n_k라면 그 노드의 지니 지수는 다음과 같이 계산된다.

$$G = 1 - \sum_{k=1}^{m} (\frac{n_k}{n})^2$$

CART 알고리즘은 지니 지수를 가장 크게 감소시키는 독립변수를 토대로 자식 노드를 생성한다, $G_L(G_R)$을 생성되는 왼쪽(오른쪽) 자식노드의 지니계수, $n_L(n_R)$을 왼쪽 자식노드에 속하는 관측수라 할 때($n = n_L + n_R$) 지니 지수 감소량은 다음과 같이 계산된다.

$$\triangle G = G - (\frac{n_L}{n}) G_L - (\frac{n_R}{n}) G_R$$

랜덤포레스트(Random Forest)는 다수의 의사결정 나무를 구성하고 각 의사결정 나무에 의해 예측되는 결과를 종합하여 목표변수의 값을 예측하는 앙상블 모형이다. 이때 서로 독립적인 트리들을 생성하기 위해 데이터의 부트스트랩 샘플을 생성하며, 또한 특정노드에서 자식 노드를 생성할 때 전체 특성을 대상으로 지니 지수 감소량을 비교하는 대신에 랜덤하게 선택되는 후보 특성들에 대해서만 지니 지수 감소량을 비교한다. 랜덤 포레스트를 위한 부트스트랩 샘플은 보통 반복 추출에 의해 원래 데이터의 2/3 만큼의 데이터를 추출한다[22].

인공신경망(Artificial Neural Network)은 동물의 뇌 신경계를 모방하여 분류를 위해 만들어진 모형이다. 입력층, 은닉층, 출력층으로 구성되는 다층신경망의 각 노드들은 가중치를 갖는 망으로 연결된다. 은닉층과 출력층의 각 노드들은 입력값들의 가중합을 계산하고 전이함수를 이용하여 출력값을 산출한 후 다음 노드로 전달한다. 학습 과정에서는 분류 오류를 최소화하기 위해서 오차의 역전파(Back-Progagation) 알고리즘을 통하여 가중치를 갱신한다. 은닉층의 수와 은닉 노드의 수는 사용자가 적절하게 설정한다[18].

선행 연구를 통하여 데이터 마이닝 기법이 패턴 분류와 예측 모델에 활용되는 것을 확인하였다. 본 연구에서는 기존에 알려진 바이러스의 특징 데이터베이스를 구축하고, 데이터 마이닝 기법을 활용하여 정상 파일과 악성 파일의 예측하여 분류하고자 한다.

3. 실험 및 결과

3.1 실험 환경

소규모 파일에 대한 바이러스 예측 분석을 수행하기 위해서 <표 1>과 같은 분석 환경을 구축하였다. Linux 운영체제 환경에서 Python와 PHP 프로그램으로 파일의 특징을 추출하였으며, 데이터마이닝도구에는 R-Studio을 활용하였다.

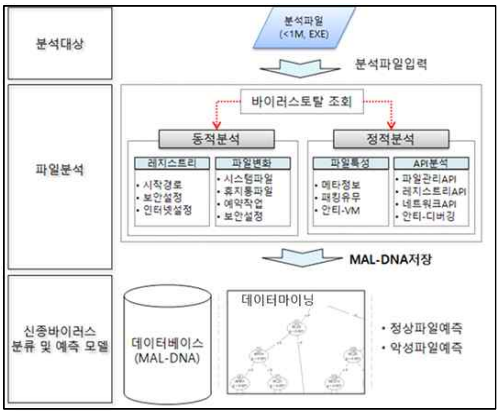
<표 1> 분석환경

분류	소분류	내용
플랫폼	운영체제	Ubuntu Linux 3.13 (x86)
	DBMS	MySQL 14.14 (x86)
분석 데이터	동적분석	자동 분석 시스템 수행 결과 전처리
	정적분석	Python/PHP 언어 기반 분석
데이터 마이닝	파일저장	마이크로소프트 엑셀 2007
	분석도구	R Studio 0.98 (의사결정나무 등)
서버 사양	서버명	HP ProLiant DL 380 G7
	CPU	인텔 Xeon X5690 (3.56 Ghz)
	메모리	8 GB

3.2 실험 과정

분석 대상은 2012년부터 2014년까지 해외에서 국내로 유입하는 1M 바이트 사이즈 미만의 EXE 확장자 파일 34,646개이다.

실험 데이터를 생성하기 위해서 기존에 알려진 정상 파일과 악성 파일을 사전에 분류하였으며, 파일의 동적/정적 특징을 바이너리 변수로 변환하여 데이터베이스화하였다. 전체적인 분석 절차는 (그림1)과 같다.



(그림 1) 전체 분석 절차

세부 분석 과정은 다음과 같다.

첫 번째, 온라인 분석 서비스인 바이러스토탈을 활용하여 총 34,646개 파일에 대하여 50여개 백신의 검출 여부를 조회하였다. 결과적으로 <표 2>와 같이 바이러스 의심 집단으로 분류되었다.

<표 2> 바이러스 토탈 조회 결과

결과(전체 34,646개 中)			
정상집단	바이러스 의심집단(백신검출개수)		
0개	1개 이상	5개 이상	20개 이상
5,292개 (18.09%)	29,354개 (84.34%)	20,379개 (58.88%)	14,116개 (40.74%)

두 번째, 분석대상파일 34,646개의 특성을 134개 항목으로 분류하여 데이터베이스화 하였다.

파일의 특성을 나타내는 구분 항목은 선행 연구에서 생성한 134개로 정의하였다[13]. 파일의 특징은 각각의 항목에 대하여 존재할 경우에는 '1', 존재하지 않을 경우에는 '0'과 같이 바이너리 데이터로 생성하였다. 세부 내용은 <표 3>와 같다.

<표 3> 파일특성필드(134개)

구분	소구분	설명	필드명	개수
정적분석	메타데이터	파일버전/설명, 버전정보 등	ME01~ME11	11개
	패커	알려진 패킹 유무(5672종)	PK01	1개
	가상머신	가상머신 탐지 여부	V01	1개
동적분석	디버깅함수	Anti디버깅 함수 등	AD01~AD05	5개
	API분석	명령함수, 네트워크 API 등	API01~API29	29개
	리소스분석	아이콘, 국가언어 등	RES01~RES30	30개
	파일관리	Windows 폴더, 어플리케이션 등 파일 생성 및 삭제	FC01~FC23	23개
	레지스트리	자동실행, 방화벽 변경 등	RC01~RC34	34개
총합계				134개

3.3 머신러닝 수행 결과

악성코드 예측 모델의 성능을 검증하기 위해서 전체 분석 파일 34,646개에 대하여 학습데이터와 검증데이터를 각각 7대 3으로 분류하였다. 또한, 학습데이터는 악성코드 유무에 대한 불균형 편차를 줄이기 위해서 R 패키지에서 제공하는 SMOTE 함수를 활용하여 1:1 샘플링을 수행하였다.

머신러닝 알고리즘은 나이브 베이즈 분류, 서포트 벡터 머신, 의사결정나무, 랜덤포레스트, 인공신경망을 활용하였다. 학습데이터를 기반으로 분석 모델을 생성하고, 검증데이터에서 모델의 성능을 검증하였다. 모델의 성능을 평가하는 주요 지표에는 정상과 악성을 판별하는 비율인 Accuracy와 Precision과 Recall의 균형을 평가하는 F1 Score를 활용하였다.

악성코드에 대한 바이러스 탐지 기준은 조직의 정보보호수준에 따라서 다르게 반영될 수 있다. 따라서 금번 연구에서는 바이러스토탈에서 판별한 바이러스 탐지 개수에 따라서 차등적으로 악성코드 분류 모델을 수행하였다. 세부적인 검증 결과는 <표 4>와 같다.

<표 4> 실험결과

악성 기준	분석 알고리즘	Recall	Precision	Specifity	**accuracy**	**F1 Score**
1개 이상	NB	0.84	0.19	0.37	0.44	0.31
	SVM	0.71	0.45	0.85	0.83	0.55
	DT	0.67	0.38	0.79	0.77	0.48
	RF	0.72	0.47	0.86	0.83	0.57
	ANN	0.52	0.48	0.89	0.83	0.50
5개 이상	NB	0.36	0.93	0.84	0.43	0.52
	SVM	0.84	0.94	0.71	**0.82**	**0.89**
	DT	0.74	0.94	0.74	0.74	0.83
	RF	0.83	0.95	0.75	**0.82**	**0.89**
	ANN	0.76	0.94	0.71	0.75	0.84
20개 이상	NB	0.63	0.67	0.79	0.73	0.65
	SVM	0.76	0.91	0.95	**0.87**	**0.83**
	DT	0.65	0.86	0.93	0.82	0.74
	RF	0.79	0.90	0.94	**0.88**	**0.84**
	ANN	0.80	0.81	0.87	0.84	0.81

결과적으로 랜덤포레스트 알고리즘의 성능이 가장 우수한 것으로 나타났다. F1 Score 평가 기준으로 악성코드에 대한 판단이 1개 이상일 경우에 56%, 5개 이상일 경우에 89%, 20개 이상일 경우에 84%의 높은 성능을 보여줬다.

4. 결론

본 연구에서는 기존에 알려진 소규모 파일의 특징을 데이터베이스화 하고, 다양한 머신러닝 기법을 활용하여 정상 파일과 악성 파일을 분류할 수 있는 효과적인 모형을 제안하였다. 결과적으로 기존에 5개의 백신이 바이러스로 판별하였을 경우에 랜덤포레스트 알고리즘으로 정확도(82%), F1 Score(89%)의 높은 성능을 검증하였다. 이를 통하여 대량으로 생산되는 신종 악성코드를 사전에 검출하는데 활용할 수 있을 것이다. 또한, 악성코드를 자동으로 빠르게 분류하여 분석 과정에 소요되는 시간과 인력을 최소화할 수 있을 것이다.

연구의 한계점으로는 소규모 EXE 파일에 한정하여 악성코드에 대한 분류 모델을 제안하였다. 향후에는 모바일 APK 파일, 문서파일 등 분석 대상을 확대하여 적합한 머신러닝 모델의 개발이 필요하다. 마지막으로 본 연구의 결과가 지속해서 증가하는 신종 악성코드 유입을 예방하여 국가와 기업의 사회적·경제적 피해를 줄일 수 있기를 기대한다.

참고문헌

[1] 강태우 외 3명 . "API call의 단계별 복합분석을 통한 악성코드 탐지", 정보보호학회논문지 제17권, 제6호, pp. 89-98, 2007.

[2] 과학기술정보통신부, "2017년 인터넷이용실태조사", 2017.

[3] 구윤회, "의사결정나무와 로지스틱 회귀분석을 이용한 태권도 수련생 이탈 예측을 위한 비교연구", 한양대학교, 2007.

[4] 국가정보보호백서, "국가정보보호백서", 국가정보원. p.149~163, 2016.

[5] 국가정보원, "국가정보보호백서", 2014.

[6] 김영진 외 3명, "의사결정트리를 이용한 날씨에 따른 화재발생 확률 예측모델", 정보과학회논문지, 제 40권, 제11호, pp. 705-715, 2013.

[7] 김태근, "악성코드 탐지 정확성 향상을 위한 행위 별 API List 비교 분석", 한양대학교, 2011.

[8] 배철민 외 3명, "Hybrid 악성코드 수집 기술 기반 Unknown 악성코드 선별 방안 연구", 한국인터넷진흥원, 가을학술발표논문집 제 39권, 제 2호, pp. 135-137, 2012.

[9] 서희석 외 2명. "윈도우 악성코드 분류 방법론의 설계", 정보보호학회논문지, 제 19권, 제 2호, pp. 88~92, 2009.

[10] 송주영, 한영선, "한국 남자 청소년의 범죄지속 위험예측 요인분석", 한국형사정책연구원 , 제98권, pp. 239-260, 2014.

[11] 유병길, "Domain 증적을 통한 효율적인 악성코드 예방 체계에 관한 연구", 고려대학교, 2013.

[12] 유영성, 이명수, "미래형 재난대응과 통합플랫폼 구축-경기도 빅데이터 프로젝트 실현에 활

용", 미래비전연구실, 2014.
[13] 이택현, "소형 악성 실행 파일의 식별 방법에 관한 연구", 서울과학기술대, pp. 54~57, 2015.
[14] 장한두, "의사결정나무분석을 통한 중소형 아파트 거주세대의 이주와 리모델링 결정요인", 대한건축학회지, 제 30권, 제 9호, pp. 45-56, 2014.
[15] 정용욱, "속성기반 악성코드 유사도 분류 문제점 개선을 위한 가중치 분석 연구", 정보보호학회지, 제23권, 제 3호, pp. 501-504, 2013.
[16] 주대영, 김종기, "초연결시대 사물인터넷(IoT)의 창조적 융합 활성화 방안", 산업연구원, pp. 17, 2014.
[17] 한경수, 김인경, 임을규, "API 순차적 특징을 이용한 악성코드 변종 분류기법", 보안공학연구논문지, 제8권, 제2호, pp. 319-335, 2011.
[18] Artificial_neural_network, https://en.wikipedia.org/wiki/Artificial_neural_network, 2018.
[19] AV-TEST, http://www.av-test.org, 2018.
[20] Darkmegi, http://blogs.mcafee.com/mcafee-labs/darkmegi-not-the-rootkit-youre-looking-for, 2017.
[21] G . Wagener, R. State, and A. Dulaunoy, "Malware behaviour analysis," Journal in Computer Virology, 2007.
[22] LEO BREIMAN, "Random Forests", Volume 45, pp. 5-32, 2001.
[23] Marian Merritt, Kevin Haley, "Norton cybercrime report 2013", Norton by Symantec, 2013.
[24] NetMarketShare, http://www.netmarketshare.com/, 2018.
[25] Virustotal Statistics, https://www.virustotal.com/ko/statistics/, 2018.
[26] World Economic Forum, http://www.weforum.org/docs/GRR17_Report_web.pdf, 2017.
[27] Nir Friedman, Dan Geiger, and Moises Goldszmidt.Bayesian network classifiers. Machine Learning, 29:131 - 163, 1997.

〔저자소개〕

이 택 현 (Taek-Hyun Lee)
2013년 2월 산업정보시스템 석사
2018년 ~ 현재 서울과학기술대 산업정보시스템 박사과정
email : futp@naver.com

국 광 호 (Gwang-Ho Kook)
1981년 2월 서울대학교 산업공학과 석사
1989년 2월 조지아공대 산업공학과 박사
2018년 ~ 현재 서울과학기술대학교 글로벌융합산업공학과 교수
email : khkook@seoultech.ac.kr