머신러닝을 활용한 해외 기업의 악성코드 탐지 연구 소개

(보안기술연구팀, 2018.4.30.)

1 개 요

□ 보안 전문가에 의해 정의된 악성코드 탐지 규칙만으로는 꾸준히 증가하고 있는 신종·변종 악성코드[1]를 효과적으로 탐지하는 것이 어려움

○ 이에 산·학계에서는 악성코드의 규모와 다양성을 고려하여 머신러닝을 활용한 악성코드 탐지를 연구

□ 본 보고서에서는 글로벌 회사인 카스퍼스키 랩(Kasperskey Lab)과 앤비디아(Nvidia)에서 공개한 머신러닝 기반의 악성코드 탐지 연구를 간략히 소개[2]

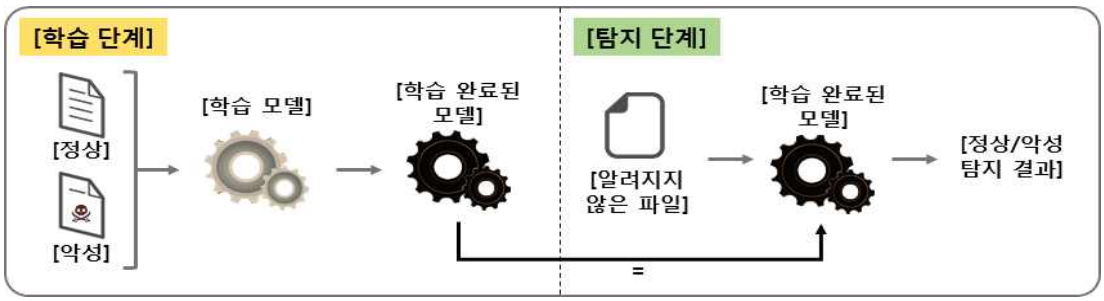
2 머신러닝 기반 악성코드 탐지

□ 머신러닝 기반의 악성코드 탐지란 **정상파일 및 악성코드가 포함된 파일**(이하 '악성코드')**로 학습 모델을 학습**시킨 후, **학습된 모델로 의심스러운 파일의 악성 여부를 탐지**하는 것(<그림 1> 참고)

○ **(학습 단계)** 학습 모델을 파일들의 특징정보(문자열, 명령어, 바이트 정보, API 호출 기록 등)와 레이블(정상/악성코드)로 학습시킴으로써 **모델이 악성코드 탐지에 최적화 되도록 함**

1) http://www.av-test.org/en/statistics/malware,
Threats Report, McAfee Labs, 2018.5.
2) Machone Learning for Malware Detection, Kasperskey Lab,
Edward Raff, Malware Detection by Eating a Whole EXE, NVIDIA, 2017.

○ **(탐지 단계)** 학습된 모델을 이용하여 **입력된 파일이 정상 파일인지 악성코드인지를 구별**



<그림 1> 머신러닝을 활용한 기본적인 악성코드 탐지 시스템 예시

□ 카스퍼스키 랩과 앤비디아는 <그림 1>과 유사한 머신러닝 기반의 악성코드 탐지 시스템을 개발

○ 시스템 설계 시 악성코드의 다양성과 규모, 분석 업무의 효율 등을 감안하여,

○ 시스템이 일반성, 강건성(Robust), 확장성, 처리율, 설명가능성 등의 특징을 갖도록 함

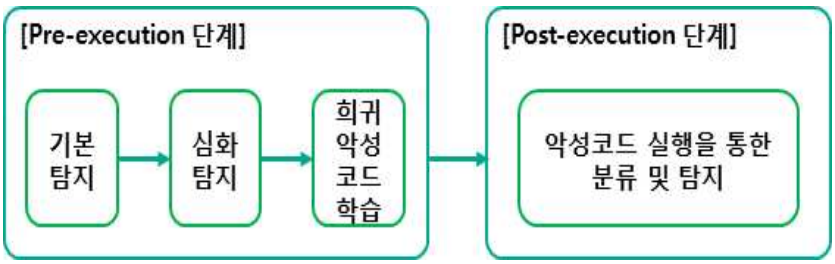
<카스퍼스키 랩과 앤비디아에서 시스템 설계 시 고려한 주요 요소>

고려 요소	설 명
일반성	- 시스템이 머신러닝에 사용된 데이터셋에 대해서만 좋은 탐지 성능을 보이는 것이 아니라 **새로운 악성코드에 대해서도 우수한 탐지 성능을 보이도록** 해야 함
강건성	- 신종 악성코드뿐만 아니라 **기존의 것을 일부 변형한 악성코드에 대해서도 탐지가 가능**하도록 해야 함
확장성 및 처리율	- 악성코드의 수가 급격히 증가함에 따라 **시스템도 확장 가능해야 하며, 처리율 역시 보장** 되어야 함
설명가능성	- 입력된 파일이 악성코드로 분류된 경우, 이에 대한 **분류 원인(예: 파일의 메타정보, 명령어 등)을 제공할 수** 있어야 함

3 카스퍼스키 랩의 악성코드 탐지 실험

□ **(실험 방향)** 카스퍼스키 랩은 악성코드 탐지 과정을 여러 단계로 구성하여 각 단계에서 머신러닝을 활용한 파일 학습 또는 악성코드 탐지를 수행

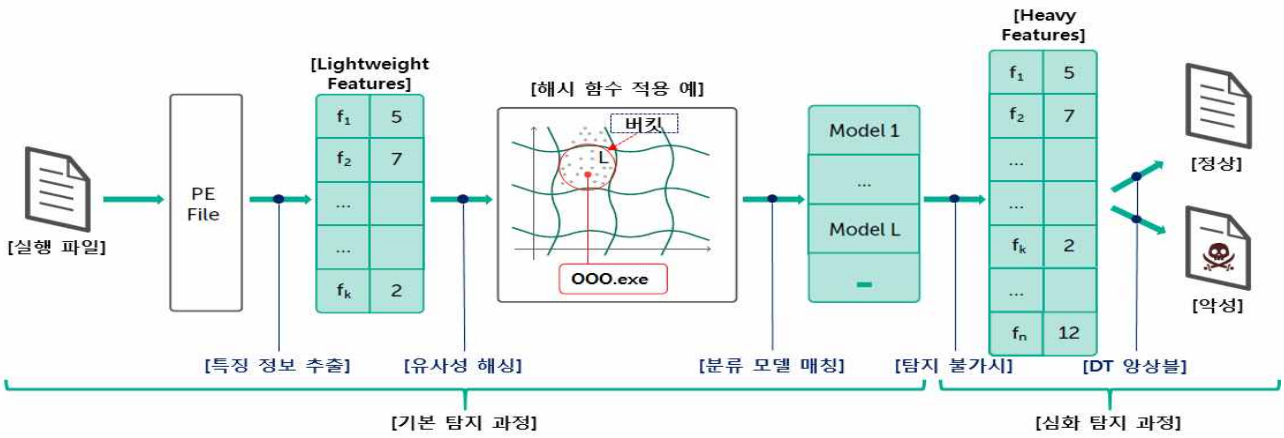
□ **(시스템 구성)** 악성코드 탐지 과정을 악성코드 Pre-execution (실행 전) 단계와 Post-execution(실행 후) 단계로 구분



<그림 2> 카스퍼스키 랩의 악성코드 탐지 시스템 구성도

○ **(Pre-execution(실행 전) 단계)** 악성코드 실행 없이 수집 가능한 정보(파일 포맷, 바이트, 추출된 텍스트 등)를 학습하여 악성코드를 탐지

- 동 단계는 사용되는 악성코드의 정보에 따라 기본 탐지, 심화 탐지, 희귀 악성코드 학습 단계로 나뉨(<그림 3> 참고)



<그림 3> Pre-execution(실행 전) 단계의 처리 과정

- **(기본 탐지)** 실행파일의 **기본적인 특징정보(Lightweight Features)로 유사성 해싱(Similarity Hashing) 함수[3]를 학습**시킨 후 정상파일과 악성코드를 분류

  • 유사성 해싱 함수는 특징정보의 해싱값에 따라 악성코드를 적절한 버킷에 분류

  • 만일, **버킷이 정상파일과 악성코드가 혼합된 경우 심화 탐지 단계를 수행**

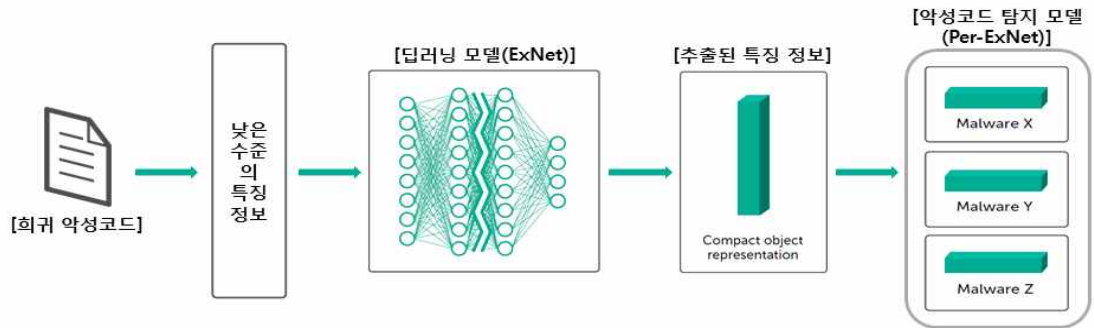
- **(심화 탐지)** 실행파일에서 **추출 가능한 모든 특징정보(Heavy Features)**로 유사성 해싱 함수를 학습시킨 후 앙상블 알고리즘[4]을 통해 정상파일과 악성코드를 분류

  • 만일, **분류 결과에 정상파일과 악성코드가 혼합된 경우 아래와 같은 희귀 악성코드 학습 단계**를 수행

< 희귀 악성코드 학습 >

○ 희귀 악성코드들을 탐지할 수 있도록 딥러닝 모델(ExNet, exemplar Network)로 이들의 특징정보를 추출하는 단계

- 추출된 특징정보는 희귀 악성코드 및 이와 유사한 악성코드를 탐지하는 딥러닝 모델(Per-ExNet, Per-exemplar classifiers)에 사용



<딥러닝 모델 기반의 희귀 악성코드 특징정보(패턴) 추출 과정>

3) 유사성 해싱(Similarity Hashing): 유사한 입력을 같은 버킷에 할당하는 지역민감해싱(LSH)에 학습 개념을 도입한 것. 입력된 파일의 특징정보와 레이블(정상/악성)로 지도학습을 수행하여 악성코드 탐지에 최적화된 해싱 함수를 만듦

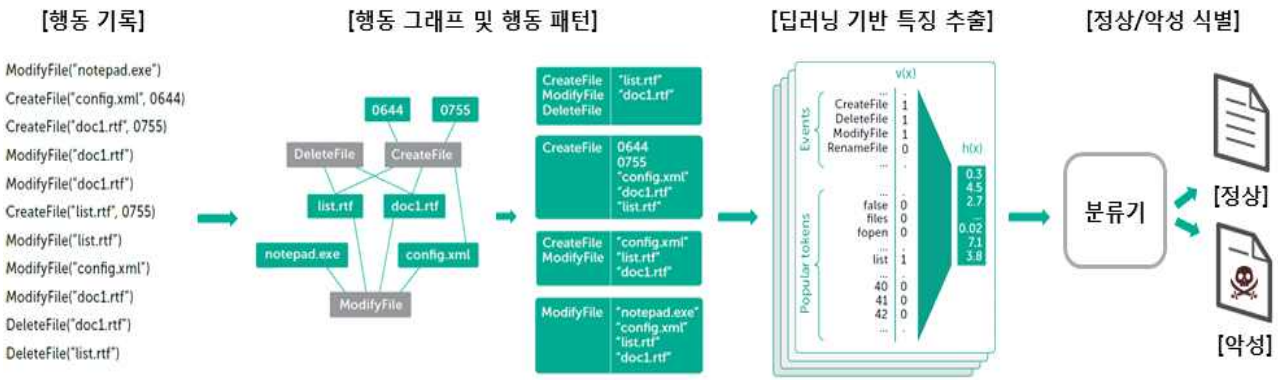
4) 의사결정나무 앙상블: 여러 개의 의사결정나무를 학습시켜 다수의 분류 결과를 도출한 후 투표 등의 방식으로 최적의 분류 결과를 찾아내는 방법

○ **(Post-execution(실행 후) 단계) 암호화, 난독화 등으로 Pre-execution 단계에서 악성코드 분류가 어려운 경우** Post-execution 단계를 수행

- 동 단계는 **악성코드를 실행하여 수집한 정보(발생한 이벤트, 프로세스 행위 기록 등)로 딥러닝 모델을 학습**시켜 악성코드를 탐지

  • 악성코드 실행 시 수집되는 프로세스의 동작 기록으로 행동 그래프 및 행동 패턴을 생성하고,

  • 딥러닝 모델로 생성된 행동 패턴의 주요 특징정보를 추출한 후 분류 모델을 거쳐 정상파일과 악성코드를 분류(<그림 5> 참고)



<그림 5> Post-execution(실행 후) 단계의 수행 과정

□ **(실험 결과)** 카스퍼스키 랩은 시스템에서 강건성, 확장성, 처리율, 설명가능성을 고려함으로써, **악성코드의 작은 변화에도 탐지 성능(오탐율 등)을 보장(강건성)하고,**

○ **대량의 악성코드 처리에 적합하며(확장성 및 처리율), 탐지 결과에 대한 담당자의 이해(설명가능성)를 도울 수 있는 시스템**을 구축

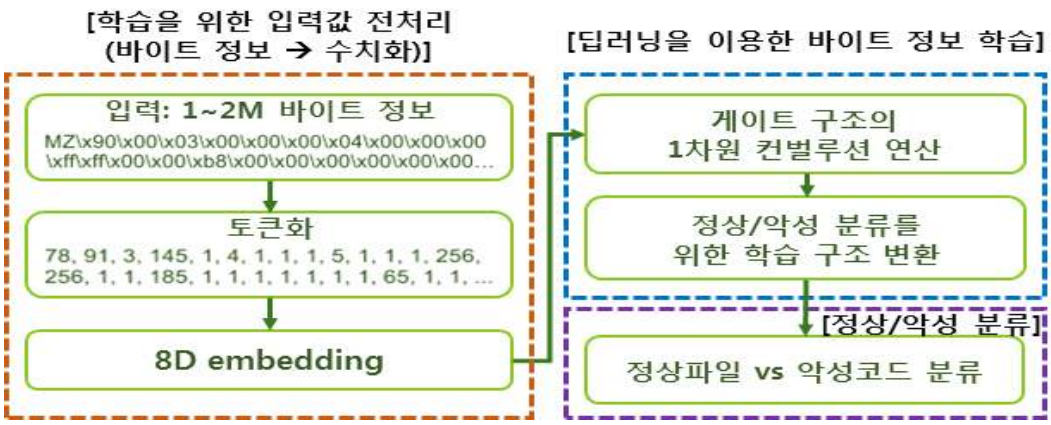
<카스퍼스키 랩 탐지시스템의 주요 설계 고려 요소>

고려 요소	탐지 시스템에 고려 요소를 적용한 방식
강건성	- 입력된 파일은 **각 단계를 거치면서 정상/악성으로 분류되는 한편, 제대로 분류되지 않은 파일은 면밀한 분석을 위해 다음 단계로 전달** - **희귀 악성코드의 경우 별도의 학습 단계를 통해 특징정보를 추출하고 추후에 사용**함으로써 탐지 가능한 악성코드의 범위를 확장
확장성 및 처리율	- 악성코드의 탐지 난이도에 따라 기본적인 단계에서 악성코드로 분류되거나, 직접 실행되는 단계를 거쳐 악성코드로 분류되기도 함 - 이는 **각 단계에 전달되는 파일의 양을 단계적으로 줄이는 것**으로써 모든 악성코드에 동일한 탐지과정이 적용되지 않게 하여 시스템의 확장 가능성과 처리 속도를 높임 - 또한 카스퍼스키 랩은 **연산 속도가 빠른 해시 함수를 사용**하여 처리 속도를 향상
설명가능성	- 악성코드 탐지에 사용되는 **특징정보를 구분하여 학습시킴으로써 탐지 결과가 설명 가능함**을 보임 - 예를 들어, 악성코드의 API 함수 사용 정보를 특징정보로 하여 학습할 경우 최종 분류 결과의 도출 근거로 API 함수를 사용

4 앤비디아의 악성코드 탐지 실험

□ **(실험 방향)** 앤비디아는 윈도우 **실행파일(.exe)의 악성코드 탐지를 위해 파일의 바이트 정보로 딥러닝 모델을 학습**시켜 악성코드 탐지를 수행

□ **(시스템 구성)** 악성코드 탐지를 위해 시스템을 실행파일의 **바이트 정보 전처리** 단계, 바이트 정보로 **딥러닝 모델을 학습**시키는 단계, 정상파일과 **악성코드를 분류**하는 단계로 구성



<그림 6> 딥러닝 기반의 악성코드 분류 모델(MalConv) 아키텍처

○ (학습을 위한 입력값 전처리) 실행파일의 바이트 정보(16진수)를 학습 알고리즘 입력값에 적합한 형태로 변환

○ (딥러닝을 이용한 바이트 정보 학습) 합성곱 신경망[5](CNN)에 2개의 컨벌루션 연산을 수행하는 게이트 구조를 적용하고, 정상/악성 분류를 위한 구조를 추가하여 학습

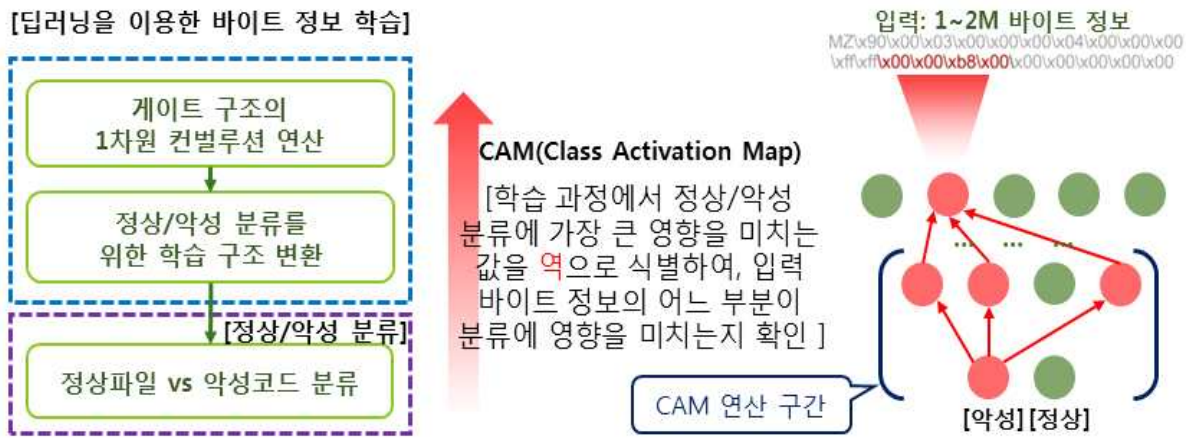
○ (정상/악성 분류) 딥러닝 학습 결과를 이용하여 정상파일과 악성코드를 분류

□ **(분류 기준 도출)** 앤비디아에서는 **CAM(Class Activation Map[6]) 방식을 활용하여 실행파일이 악성코드로 분류되게 된 근거**를 찾음

○ CAM은 정상/악성 분류 단계까지의 학습 과정에서 **연산 결과 중 가장 큰 값이 입력된 바이트 정보 어느 부분에 해당하는지를 계산**(<그림 7> 참고)

5) 합성곱 신경망(CNN): 입력 데이터의 특징을 추출하는 다수의 층(layer)과 분류를 수행하는 완전연결층(fully connected layer)으로 구성된 신경망. 특징을 추출하는 다수의 층들은 이전 층에서 전달받은 정보에서 특징 정보맵을 생성(합성곱 연산)하는 층과 이를 다음 층으로 넘기기 위해 크기를 축소(풀링 연산)시키는 층으로 구성. 특징 추출이 완료되면 완전연결층을 거쳐 악성코드 분류 등을 수행

6) Zhou, Learning Deep Features for Discriminative Localization, CVPR, 2016.



<그림 7> CAM 방식을 이용한 악성코드 분류에 가장 큰 영향을 미친 바이트 부분 탐색

□ **(시스템 테스트)** 앤비디아에서는 바이트 블록[7], 파일 메타 정보를 특징정보로 활용한 학습 모델과 제시한 모델(Malconv)의 정확도를 비교하여 제시한 모델의 타당성을 검증

□ **(실험 결과)** 앤비디아는 시스템 설계 시 일반성, 강건성, 확장성, 처리율, 설명가능성, 특징정보 의존성을 고려함으로써, **악성코드의 작은 변화에도 탐지 성능(오탐율 등)을 보장(강건성)하고,**

○ **범용적 시스템을 위해 학습만을 위한 데이터셋을 사용**(일반성)하였으며, **바이트 정보만을 사용하여 특징정보 추출 과정을 축소**(특징정보 의존성 감소)

○ 또한 **대량의 악성코드 처리에 적합하며(확장성 및 처리율), 탐지 결과에 대한 담당자의 이해(설명가능성)를 도울 수** 있는 시스템을 구축

7) 바이트 블록: 바이트 정보를 고정된 길이로 분할한 단위

<앤비디아 탐지시스템의 주요 설계 고려 요소>

고려 요소	탐지 시스템에 고려 요소를 적용한 방식
일반성	- 딥러닝 모델 학습·검증에 사용되는 데이터셋을 두 개의 전혀 다른 출처로부터 수집하여 그룹 A, B로 구분 - **두 개의 그룹 A, B 중 하나의 그룹으로만 모델을 학습**시키고 **다른 그룹으로 성능을 검증**하여 모델의 일반성을 향상
강건성	- 바이트 정보를 고정된 길이로 분할하지 않고 **파일 전체의 바이트 정보를 학습**함으로써, 일부만 **변형된 악성코드들도 탐지**할 수 있는 가능성을 높임
확장성 및 처리율	- 학습 과정의 **계산 복잡도가 바이트 정보 길이에 선형적으로 증가**하도록 설계
설명가능성	- 악성코드 탐지에 **가장 큰 영향을 미치는 요인을 추적하여 해당 요인이 바이트 정보와 매핑되는 부분을 식별**함으로써 탐지 결과가 설명 가능함을 보임
특징정보 의존성	- **바이트 정보만을 사용하여 학습을 위한 특징정보 추출 과정을 축소**시켰으며, 이는 특정 특징정보에 대한 의존성을 낮춤

5 결 론

□ 글로벌 회사*에서 머신러닝을 활용한 악성코드 탐지 실험을 공개한 것은 **사이버보안 분야에 머신러닝의 도입이 구체화되고 있는 좋은 사례**

* 사이버보안 분야의 카스퍼스키 랩, 머신러닝 하드웨어 개발 분야의 앤비디아

○ (카스퍼스키 랩) 악성코드 탐지를 단계적으로 구성하여 의심스러운 파일의 레이블(정상/악성)을 정함으로서,

- 최종적으로 보안 전문가가 분석해야하는 악성코드의 양을 줄여 분석 업무의 효율성을 향상

○ (앤비디아) 바이트 정보의 일부분을 학습에 활용하는 것이 아닌 전체 바이트 정보를 학습하는 것의 중요성을 기존의 학습 모델(바이트 블록, 파일의 메타 정보 활용)과 비교하여 보임

□ 악성코드 분석 관련 **담당자들은 머신러닝을 활용한 새로운 탐지 방식을 통해 악성코드 분석에 필요한 추가적인 유용한 정보를 얻고, 이를 활용할 수 있도록 예의 주시**할 필요가 있음