[별첨 5]

딥러닝을 이용한 악성파일 탐지 프로그램

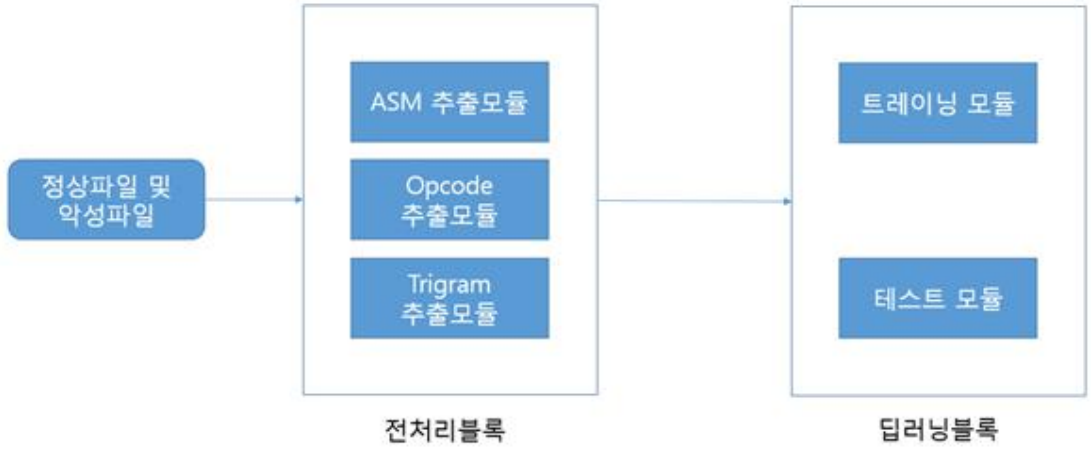
지능보안연구그룹

ETRI 한국전자통신연구원
www.etri.re.kr
초연결통신연구소/정보보호연구본부

목 차

■ 딥러닝을 이용한 악성파일 탐지 프로그램

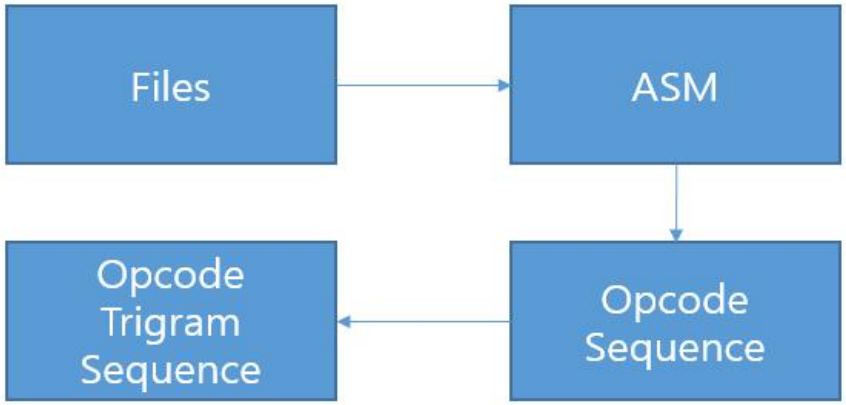
❖ 본 기술은 정상파일 및 악성파일로부터 악성파일 탐지를 위한 딥러닝 모델을 학습하고 라벨링이 되어있지 않은 정상파일 및 악성파일이 주어졌을때 악성여부를 탐지하는 프로그램에 관한 것임



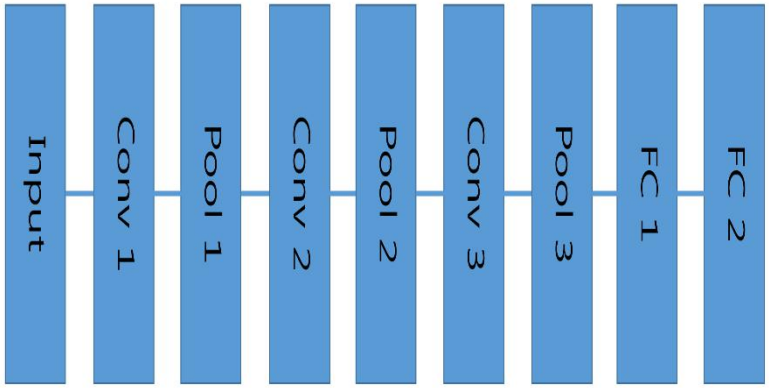
< 개발 기술 운용 개념도 >

▣ 딥러닝을 이용한 악성파일 탐지 프로그램

- ❖ 본 기술은 데이터 전처리블록과 딥러닝블록으로 구성됨
- ❖ 데이터 전처리 블록은 PE 파일로부터 opcode trigram sequence를 추출함
- ❖ 딥러닝 블록은 전처리 데이터를 사용하여 딥러닝 모델을 학습하고 라벨링 되지 않은 파일이 주어질때 악성여부를 판단함



< 딥러닝을 위한 전처리 과정 >



< 악성파일 탐지를 위한 딥러닝 모델 >

■ 기술이전 내용 및 범위

❖ 기술이전 내용

- ✓ 딥러닝을 위한 PE 파일 전처리 기술
- ✓ 딥러닝을 이용한 학습 기술
- ✓ 딥러닝을 이용한 탐지 기술

❖ 기술이전 범위

- ✓ 소스코드: 딥러닝을 이용한 악성파일 탐지 프로그램
- ✓ 문서: 시스템 설계서, 개발문서, 기술문서

■ 기술 개발 현황

❖ 기술성숙도(TRL : Technology Readiness Level) 단계 :

구 분	단계	정 의	세 부 설 명
기초 연구 단계	1	기초 이론/실험	◦ 기초이론 정립 단계
	2	실용목적의아이디어, 특허 등 개념정립	◦ 기술개발 개념 정립 및 아이디어에 대한 특허 출원 단계
실험 단계	3	실험실 규모의 기본성능 검증	◦ 실험실 환경에서 실험 또는 전산 시뮬레이션을 통해 기본성능이 검증될 수 있는 단계 ◦ 개발하려는 부품/시스템의 기본 설계도면을 확보하는 단계
	4	실험실 규모의 소재/부품/시스템 핵심성능 평가	◦ 시험샘플을 제작하여 핵심성능에 대한 평가가 완료된 단계 ◦ 3단계에서 도출된 다양한 결과 중에서 최적의 결과를 선택하려는 단계 ◦ 컴퓨터 모사가 가능한 경우 최적화를 완료하는 단계
시작품 단계	5	확정된 소재/부품/ 시스템 시작품 제작 및 성능 평가	◦ 확정된 소재/부품/시스템의 실험실 시작품 제작 및 성능 평가가 완료된 단계 ◦ 개발 대상의 생산을 고려하여 설계하나 실제 제작한 시작품 샘플은 1~수개 미만인 단계 ◦ 경제성을 고려하지 않고 기술의 핵심성능으로만 볼 때, 실제로 판매가 될 수 있는 정도로 목표 성능을 달성한 단계
	6	파일롯 규모 시작품 제작 및 성능 평가	◦ 파일롯 규모(복수 개~양산규모의 1/10정도)의 시작품 제작 및 평가가 완료된 단계 ◦ 파일롯 규모 생산품에 대해 생산량, 생산용량, 불량률 등 제시 ◦ 파일롯 생산을 위한 대규모 투자가 동반되는 단계 ◦ 생산기업이 수요기업 적용환경에 유사하게 자체 현장테스트를 실시하여 목표 성능을 만족시킨 단계 ◦ 성능 평가 결과에 대해 가능하면 공인인증 기관의 성적서 확보
실용화 단계	7	신뢰성평가 및 수요기업 평가	◦ 실제 환경에서 성능 검증이 이루어지는 단계 ◦ 부품 및 소재개발의 경우 수요업체에서 직접 파일롯 시작품을 현장 평가(성능 및 신뢰성 평가) ◦ 가능하면 인증기관의 신뢰성 평가 결과 제출
	8	시제품 인증 및 표준화	◦ 표준화 및 인허가 취득 단계
사업화	9	사업화	◦ 본격적인 양산 및 사업화 단계 ◦ 6-시그마 등 품질관리가 중요한 단계

▣ 딥러닝을 이용한 악성파일 탐지 프로그램

- ❖ 전처리 방식으로 opcode를 사용하기 때문에 동적분석형태의 API 시스템콜을 사용하는 기존의 방식에 비해 훨씬 빠르게 학습 및 테스트 데이터를 수집할 수 있음
- ❖ 딥러닝 방식을 사용하기 때문에 Decision Tree 및 Support Vector Machine (SVM)을 사용하는 머신러닝 방식에 비해 전처리가 수월하고 더 높은 탐지율을 보여줌
- ❖ 딥러닝 모델을 사용하기 때문에 대용량의 데이터에 대한 학습이 용이함

■ 예상 응용제품 및 서비스

예상 제품/서비스	예상 수요자(층)
딥러닝을 이용한 악성파일 탐지시스템	국가 CERT 및 ISP 망 사업자
딥러닝을 이용한 악성파일 탐지시스템	공공기관 및 교육기관
딥러닝을 이용한 악성파일 탐지시스템	기존 네트워크 보안 업체

■ 기술이전 업체 조건

❖ 해당사항 없음

■ 사업화시 제약조건

❖ 해당사항 없음

■ 사업성

예상 제품/서비스	예상단가 (천원)	이전기술의 비중(%)	잠재적/현재적 경쟁자와 가격, 시장 등에서 경쟁상 유리한 점	판매 가능 시기
딥 러 닝 을 이 용 한 악성파일 탐지시스템	100,000	50%	a. 가격경쟁력면: 기존 외산제품에 비해 저렴하며 기능 요구사항에 따라 변동 가능 b. 시장환경면: 제로데이 악성코드 탐지를 위한 인공지능 백신에 대한 요구가 많음 c. 기타: 국내의 특수한 악성파일에 대한 대응이 가능함	2018

- ❖가격 경쟁력: 상용제품의 기능 요구사항에 따라 변동가능
- ❖상용화를 위한 생산설비 등 추가비용: S/W 이므로 생산 설비 등 추가비용 없음
- ❖상용화를 위한 추가적인 기술개발 내용
  - ❖ 대량의 악성파일 학습을 위한 환경이 준비되어야 함

■ 국내외 시장 동향

- 국내의 악성파일 탐지 및 대응 관련 보안 시장은 `18년부터 연 평균 20%씩 성장하여, `22년 시장규모는 약 2,000억 원까지 증대될 것으로 추산함 (출처 : 2016 국내 정보보호산업 실태조사(KISA, 2016))
- 전 세계 악성파일 탐지/대응 관련 시장은 `18년 139,542억원 규모이며, 연 평균 1.4%씩 성장하여, `20년까지 시장규모는 약 149,846억원까지 증대될 것으로 추산함 (출처 : 2012 세계 지식정보보안산업 비교 분석(KISA, 2012))
  - 전 세계 사이버 공격 탐지/대응 관련 시장 자료가 없는 관계로, 국내 시장 점유율을 세계 지식정보보안 시장 현황에 반영하여 추정함
- 국내외 시장규모

(단위: 억원)

관련 제품 /서비스	시장	1차년도 (2018)	2차년도 (2019)	3차년도 (2020)	4차년도 (2021)	5차년도 (2022)	합계
악성파일 탐지 및 대응 시스템	해외 (억원)	60,000	70,000	80,000	90,000	100,000	
	국내 (억원)	1,200	1,400	1,600	1,800	2,000	

감사합니다.

www.etri.re.kr

♣ 연락처 : 초연결통신연구소/지능형보안연구그룹 김종현 책·연 (042-860-3843, jhk@etri.re.kr)