컴퓨터공학부 캡스톤디자인 계획서 발표회 답변서

팀명: 5 조 assist(a security safety important special team)
조원: 손현기 (조장), 김주환, 김호준, 오예린, 이동윤, Ruslan

질문
● 이전 캡스톤 프로젝트와의 차별점은 무엇인가? ● 악성코드 파일을 두 개의 인공지능으로 분류하겠다는 것인가? 그렇게 했을 때 장점은 무엇인가?
답변
기존 프로젝트는 정상 파일과 악성 파일을 분류하는 문제를 인공지능으로 해결하는 프로젝트였습니다. 저희의 주제는 파일을 분류하는 것이 목표가 아니라, 인공지능이 분류하지 못하는 파일을 보안 전문가가 직접 분석할 때 소요되는 시간을 줄이기 위해 전문가가 검사해야 하는 코드의 범위를 줄여주는 인공지능 기반 소프트웨어를 개발하는 것입니다.

질문
● asi 가 제시한 의심내용이 적절한지를 평가할 수 있는 기준이 필요함 ● 문서의 악성과 정상으로 분류 후 다음 단계의 필터링을 진행하는 데에 있어서 최종 결과물에 대한 작업의 내용이 매우 모호합니다. 정확하게 목표를 세우고, 취급할 악성 코드의 종류들을 분석하고 나열하여 그룹화하는 작업이 필요합니다.
답변
본 프로젝트와 유사한 기능을 수행하는 상용 프로그램이 없기 때문에 기존 소프트웨어와의 비교가 어려운 상황입니다. 저희는 이를 완화하기 위해 다음과 같은 방법론을 도입했습니다. 1. asi 가 예측한 악성 행위 수행 구문을 저장한다. 2. Elasticsearch 를 이용해 정상/악성 파일 데이터셋에서 해당 구문을 포함하는 파일을 찾는다. 3. 해당 구문을 포함하는 파일 중 악성 파일의 비중이 높다면, asi 가 적절히 예측했다고 판정한다. 한편, 본 프로젝트의 목표는 악성 코드 패밀리 분류 (악성 코드의 종류를 분류하는 문제)가 아니므로 데이터셋의 악성 코드의 종류는 큰 의미가 없습니다. 다만, 일반성을 잃지 않기 위해서 KISA, Microsoft 와 같은 공신력 있는 기관에서 제공하는 정상/악성 데이터셋을 활용하여 다양한 패밀리의 악성코드를 포함하도록 실험할 예정입니다.

심사의견
● 어떤 인공신경망 학습모델을 사용할 것인지에 대한 구체적인 설명 추가하면 좋을듯 합니다.
답변
제안서에 작성한 것과 같이 다양한 RNN(recurrent neural network) 신경망을 이용해 실험을 수행하고, 이 중 최적의 신경망을 이용해 최종 프로젝트 결과를 만들 예정입니다. 사용할 신경망의 구조는 가장 기본적인 RNN 구조인 Vanilla RNN 과 이를 변형한 LSTM(long short-term memory), GRU(gated recurrent units) 각각에 대해 실험을 수행할 예정입니다.