시스템 콜을 이용한 안드로이드 앱의 특성 변화 분석

단국대학교 소프트웨어학과 임찬수, 조성제, 이승민, 정성원

1.서론

- 안드로이드 악성 앱의 현황 : 2019년 기준 최소 10% 이상의 악성 앱이 분포
- 악성 앱 탐지 기법 : API, Permission 등의 데이터를 특징 정보로 머신러닝에 활용
- 본 연구의 목표
  - 시스템 콜(System call)이란 동적 특징 정보로 사용해 악성 앱을 분류함으로써 암호화/난독화된 앱의 시스템 서비스의 요청 패턴을 분석
  - API나 Permission 등과 다르게 가변성이 적다는 특징을 통해 개념 드리프트(Concept Drift)와 앱의 패턴 변화를 파악

2. 실험 방법 및 데이터셋

- 사용 데이터
  - AndroZoo에서 수집한 APK(Android Package)
  - 연도별 양성 앱 1,000개, 악성 앱 500개, 총 4개년도 6,000개.
- 데이터 추출 절차

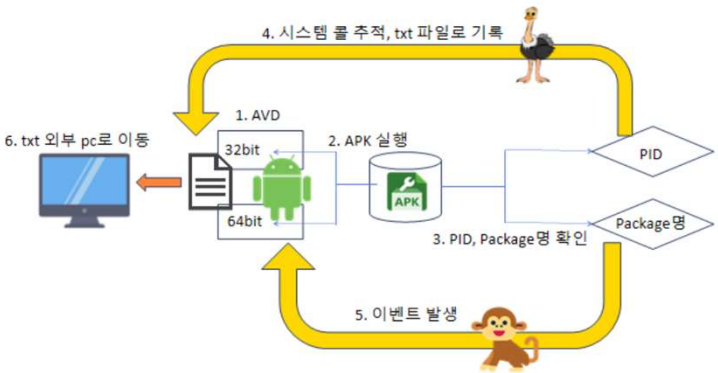

그림 1. 안드로이드 앱들의 시스템 콜 추출 과정

1. AVD 생성
- 32 비트, 64 비트 안드로이드 에뮬레이터를 설치
- 32 비트 안드로이드 API 버전 : 25(Nougat)
- 64 비트 안드로이드 API 버전 : 33(Tiramisu)

2. APK 설치 및 실행

3. PID, Package 명 확인
- PID : pidof 명령어 사용
- package 명 : aapt 도구 사용

4. PID를 이용해 시스템 콜 정보를 txt 파일로 기록
- 시스템 콜을 기록하는 strace 도구 사용

5. Package 명을 이용해 이벤트 발생
- 안드로이드에 내장된 테스트 도구인 Monkey로 총 1만 개의 이벤트를 발생
- 사용자 이벤트(클릭, 터치 등)은 75%,
  시스템 이벤트(홈 이동, 전화, 메시지 등)는 25%로 설정

6. txt 파일 외부 pc로 이동
- 기록된 시스템 콜 정보를 에뮬레이터 외부로 이동

- 데이터 추출 결과
  - 최종 추출된 APK 데이터 : 6,000개 중 실행이 되지 않는 앱 제외 총 5,200개
  - 최종 추출된 시스템 콜 종류 : 2017~2018년 111개, 2019~2020년 112개, 통합 117개 추출
- 데이터 전처리
  - 특징 정보는 32 비트/64비트의 환경 차이를 극복하기 위해 32비트 운영체제의 시스템 콜 10개를 64비트 운영체제의 시스템 콜에 맞추어 매핑
- 데이터 학습
  - 분류 모델 : Random Forest, LightGBM, XGBoost, Gradient Boosting
  - 2017~2018년도/2019~2020년도 데이터 셋의 특징 중요도 파악
  - 2017~2018년도 데이터셋의 80%로 학습한 모델을 구성하고, 이 모델을 사용하여 2017~2018년도 데이터셋의 20%로 테스트, 2019~2020년도 데이터셋에 대해서 테스트를 수행

3. 실험 결과

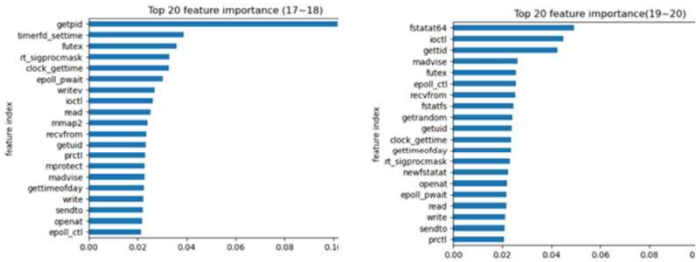

그림 2. 2017~2018년도(좌), 2019~2020년도(우) 데이터셋에서 높은 중요도의 시스템 콜

모델 / 성능	RF	LightGBM	XGBoost	GradientBoosting
17~18년도 테스트 데이터	93.07	93.22	92.47	92.44
19~20년도 테스트 데이터	71.03	59.28	66.63	66.05

표 1. 데이터셋에 따른 모델별 정확도 (단위: %)

4. 결론

- 시스템 콜은 API나 Permission과 같은 정적 특징 정보에 비해 업데이트 빈도가 낮기 때문에 시간에 영향을 덜 받음.
- 그럼에도 시스템 콜 기반의 분류 모델이 시간이 흐름에 따라 성능이 저하됐다는 것은, 작동의 패턴 자체가 바뀌었다는 것을 시사함.

- 한계점
  - API와 Permission 기반의 기존 악성 앱 분류 모델과 비교했을 때, 제안한 악성 앱 분류 모델의 평균 정확도가 낮음.
  - 시스템 콜이 개념 드리프트 문제를 해결하는데 있어 비효과적
- 향후 연구
  - 특징 정보 측면 : 동적 분석 기반의 악성 앱 탐지를 위해 시퀀스(sequence) 정보를 반영
  - 모델 측면 : 클러스터링을 적용하는 연구를 진행할 계획

Dankook univ.
Computer Security & OS LAB