머신러닝을 이용한 퍼미션 기반 악성앱 탐지

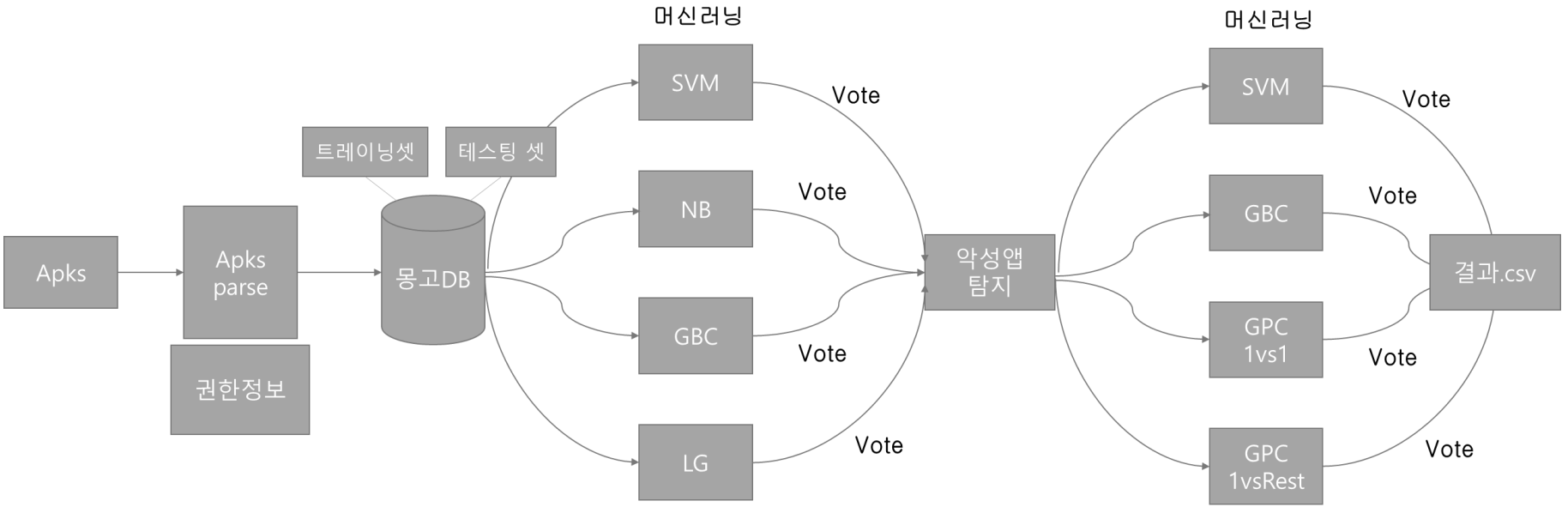
ASAP
(Android Static Analysis Platform)

하상민 강성은

목적

- AndroidManifest.xml의 권한을 Feature로 이용하여 분석에 대한 자원과 시간의 비용 절감
- 악성앱의 과도한 권한과 위험한 권한을 이용하여 멀웨어 탐지
- 탐지한 악성앱들에서 위험한 권한들의 그룹화를 통해 악성앱 패밀리 분류

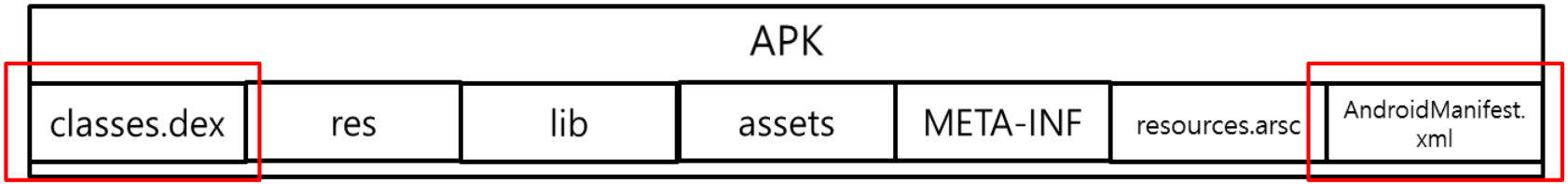
방법론


머신러닝
머신러닝
Apks
Apks parse
권한정보
트레이닝셋
테스팅 셋
몽고DB
SVM
NB
GBC
LG
Vote
Vote
Vote
Vote
악성앱 탐지
SVM
GBC
GPC 1vs1
GPC 1vsRest
Vote
Vote
Vote
Vote
결과.csv

방법론

- 트레이닝 데이터 셋에서 apk parse를 이용하여 AndroidManifest.xml의 정보 파싱
- Classes.dex를 분석하여 API 사용 여부 확인
- 몽고 DB를 통해 Json 형태로 파싱된 내용 저장
- 저장된 데이터를 4개의 머신러닝을 통해 트레이닝
- 트레이닝 된 4개의 머신러닝에 테스팅 데이터 셋 실험

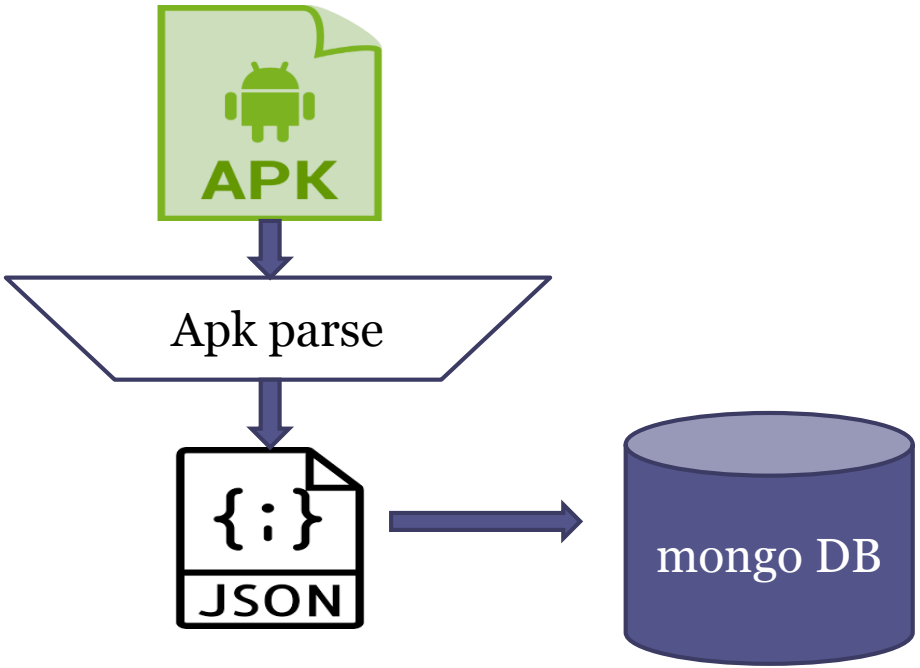
데이터 저장



Class 와 Method를 추출하여 퍼미션과 연관된 API가 실질적으로 사용 됐는지 확인함

악성앱에서 사용된 권한을 추출하여 위험한 권한이 있는지 확인함

데이터 저장



JSON 저장 리스트	
Android Manifest	Permissions Activities Services SDK …
Classes.dex	Class Method Fields Parameters …

Feature Selection

총 138개의 악성앱에서 악용될 수 있는 위험한 권한들을 Feature로 사용

퍼미션명	설명
android.permission.CALL_PHONE	어플리케이션을 통해 전화 걸기 권한
android.permission.INSTALL_PACKAGES	어플리케이션을 통한 APK 설치 권한
android.permission.INTERNET	인터넷 사용권한
android.permission.READ_CONTACTS	주소록 읽기 권한
android.permission.READ_EXTERNAL_STORAGE	SD카드 읽기 권한
android.permission.READ_SMS	SMS 읽기 권한
android.permission.RECEIVE_BOOT_COMPLETED	부팅 완료 이벤트 접근 권한
android.permission.SEND_SMS	SMS 전송 권한
android.permission.BATTERY_STATS	배터리 상태 접근 권한
android.permission.RECORD_AUDIO	오디오 녹음 권한

Feature Selection

위험한 권한과 해당 권한을 요구하는 API

퍼미션명	연관 API
android.permission.SEND_SMS	android.telephony.gsm.SmsManager.sendDataMessage android.telephony.gsm.SmsManager.sendMultipartTextMessage android.telephony.gsm.SmsManager.sendTextMessage …
android.permission.RECEIVE_SMS	com.android.internal.telephony.IccSmsInterfaceManagerProxy.copyMessageToIccEf com.android.internal.telephony.IccSmsInterfaceManagerProxy.getAllMessagesFromIccEf com.android.internal.telephony.IccSmsInterfaceManagerProxy.updateMessageOnIccEf …
android.permission.WRITE_EXTERNAL_STORAGE	com.android.providers.media.MediaProvider.openFile com.android.providers.downloads.DownloadProvider.insert …

Feature Selection

- 매칭된 API가 악성앱에서 실제로 사용됐는지 확인
- 만약 API가 사용 될 경우 API와 연관된 퍼미션을 선택함
- 선택된 퍼미션들을 Feature로 이용하여 머신러닝에 이용함

실험 결과

카테고리		실제 결과	
		Malware	Benign
실험결과	Malware	474	9
	Benign	26	1491

4개의 머신러닝의 결과의 평균을 통해 멀웨어를 탐지한 결과 98.5%의 탐지율을 얻었다.

추후연구

- so 파일 분석을 통한 Native 영역 분석
- 동적 분석을 통해 난독화된 API 분석
- 자동화된 동적 로딩 파일 추출을 통한 동적 로딩 분석
  -(ex DexClassLoader를 통한 dex, jar, apk파일 로딩)

감사합니다.