안드로이드 앱의 개인정보 유출 여부 분석

김진영, 윤용호, **이승중**
서울대학교 프로그래밍 연구실
2011.06.27

개요

- **동기**
  - 무엇 때문에 분석하는가?
  - 정적 분석하면 좋은 점
- 분석하기 위해 필요한 것들
  - 개인정보가 새는 곳, 개인정보를 보내는 곳
  - Android Platform의 행동 묘사
  - Dalvik 핵심 언어(Dalco), 요약 실행

동기

- 스마트폰 사용자 증가
  - 국내 스마트폰 보급대수 1000만대 ('11년 03월)

IT&과학

스마트폰 이용자 1,000만명 시대

2009년 본격 보급 1년 4개월만에 휴대폰 이용자 5명중 1명꼴 해당
안드로이드 OS가 전체 60% 차지

임석훈기자 shim@sed.co.kr

스마트폰 이용자가 1,000만명을 넘어섰다. 24일 방송통신위원회와 통신업계에 따르면 지난 23일 기준으로 국내 스마트폰 가입자는 1,002만명으로 집계됐다. 통신사별로는 SK텔레콤 510만명, KT 375만명, LG유플러스 117만명 등이다.

동기

- 더불어 악성코드도 증가
  - 개인정보의 접근이 비교적 쉬움
  - PC에 비해 보안 프로그램이 상대적으로 적음

스마트폰 악성코드 2천개...좀비폰 출현 우려

[스마트 시대 暗]출처모호 앱 사용과 개인정보 남발 자제해야

머니투데이 조성훈 기자 › 기자의

스크랩 :

스마트폰은 우리 삶을 18
간 뇌의 확장판 격인 PC
스마트폰 속 각종 애플리
의 결합 이상의 활용성을

스마트폰 악성코드 급증, 올들어 100개 넘어

지면일자 2011.06.13 장윤정기자 linda@etnews.co.kr ▶ 기자의 다른 기사 보기

기사구매하기 PDF보기 번역의뢰 한마디쓰기 (0) -작게 | 기본 | +크게

[AD] 임베디드 S/W에서 H/W 까지 토탈 솔루션 세미나 7/12

최근 안드로이드 스마트폰 악성코드가 폭발적인 증가세를 보이고 있다. 스마트폰 악성코드는 개인정보유출은 물론이고 과금 발생 등의 소비자 피해를 유발할 수 있어 대책 마련이 시급하다.

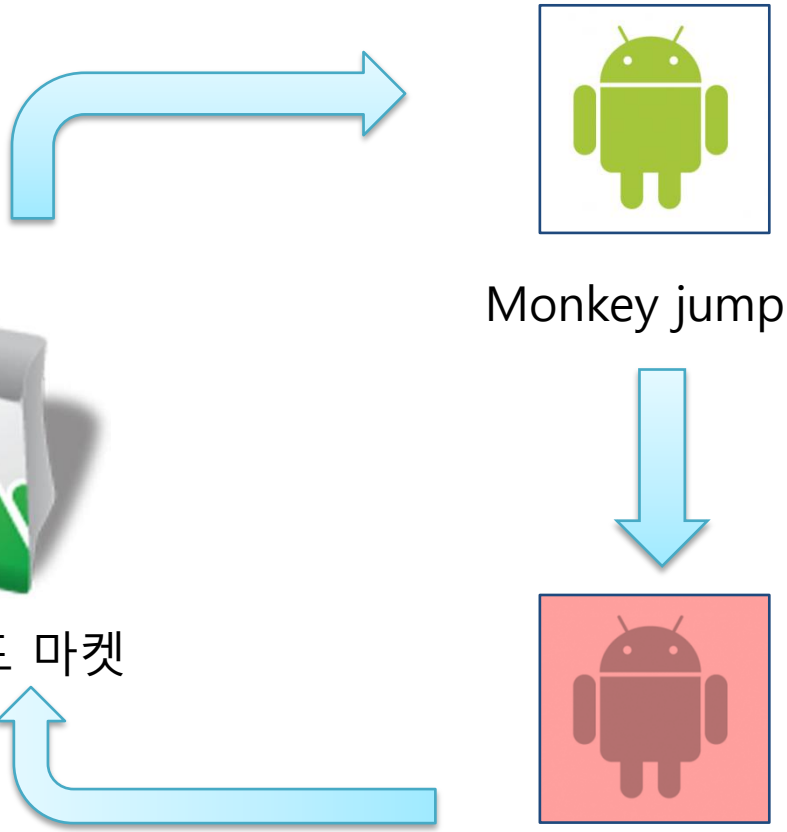
개인정보를 유출하는 앱의 예(1)

- 인터넷 라디오 앱
- 생일, 위치(GPS), 성별 정보
- 앱에서 광고 라이브러리 **5개** 발견(Veracode)

AdMarvel, AdMob, comScore, Google.Ads, Medialets

개인정보를 유출하는 앱의 예(2)

- 변조된 Monkey jump
- 중국 market에서 발견
- 개인정보 송신기능 추가
  - 연락처
  - 사용자 GPS
  - 폰 번호 등


Monkey jump 게임
안드로이드 마켓
변조된 Monkey jump 게임

권한 지정 방식

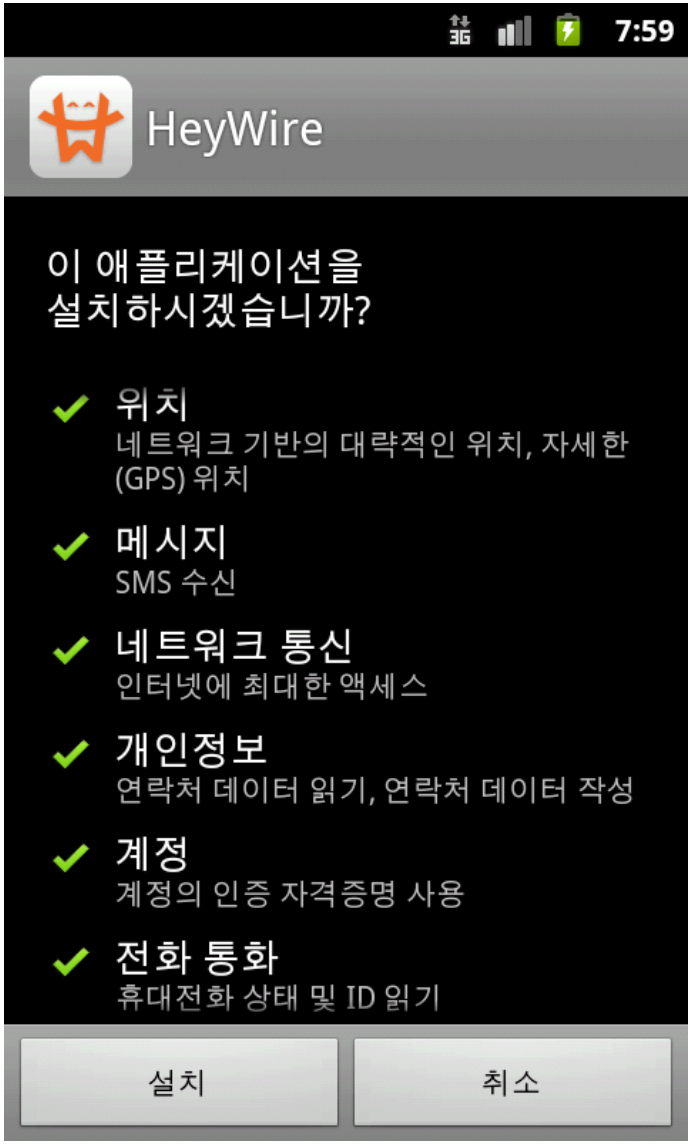
- 앱은 기본적으로 아무런 권한 없음
- 개발자가 직접 앱의 설정파일에 추가해서 배포
- 앱 설치 때 사용자에게 알려줌

AndroidManifest.xml

```
<uses-permission android:name="android.permission.READ_SMS">
</uses-permission>
```

READ_SMS, ACCESS_FINE_LOCATION (GPS)
CALL_PHONE, CAMERA

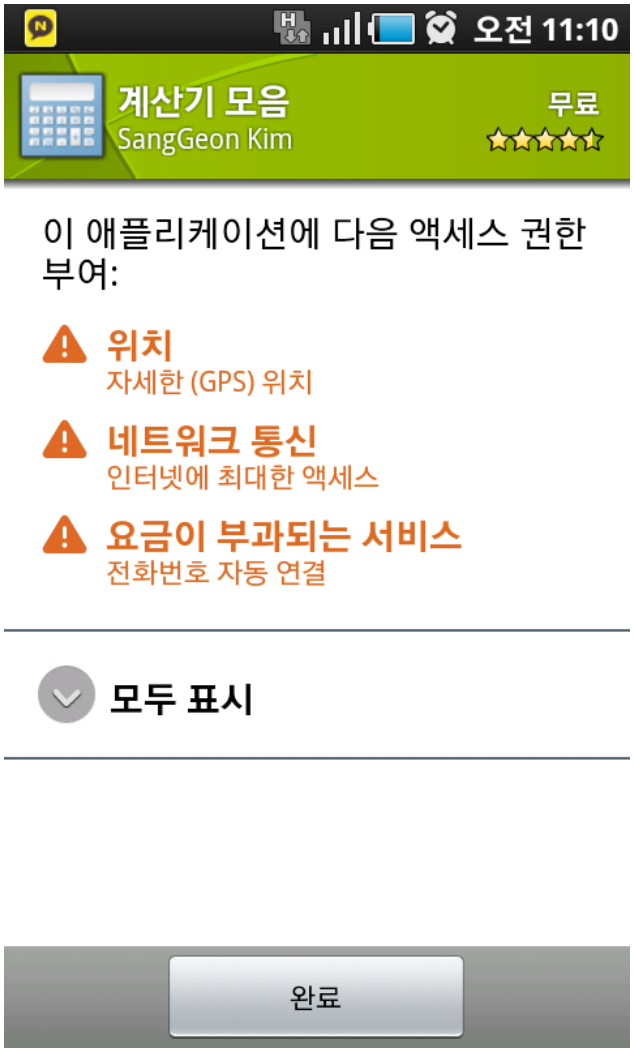
자세한 정보 얻기 힘듦



- 어느 사이트에 접속하는 지
- 연락처를 읽어서
  어디에 사용하는 지
- 전화 통화를 하는 지,
  ID만 읽는 지

잦고 뻔한 권한 허용 여부 질문

- 자세히 살피지 않게 되고
- 그냥 동의하고 넘어감



실행중에 유출을 알아내는 방법

- Taintdroid
  - 실행 중 얻어지는 개인정보에 표시
  - 표시된 정보가 외부로 나가게 될 때 경고

  - 휴대폰에 부담 14% 증가, 메모리, IPC 부담
  - 계산량 증가 -> 배터리 지속시간 감소

정적으로 분석하면 좋은 점

- 프로그램에 추가적인 성능 부담 없음
  - 좀더 적극적인 테스트 가능
- 실행 가능한 모든 경우를 잡아낼 수 있음

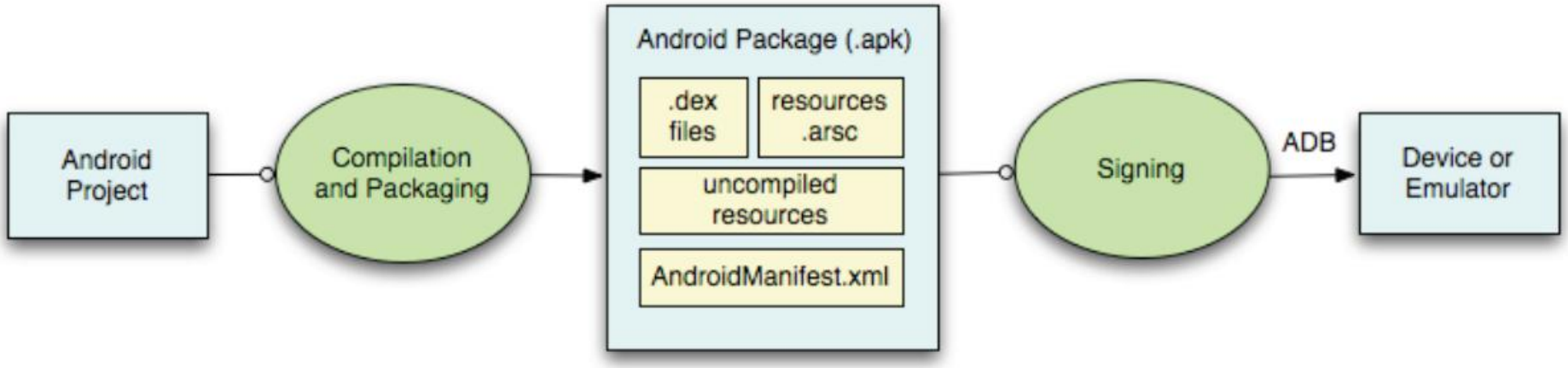
개요

- **동기**
  - **무엇 때문에 분석하는가?**
  - **정적 분석하면 좋은 점**

- **분석하기 위해 필요한 것들**
  - 안드로이드 앱의 기본적인 사항
  - 개인정보가 새는 곳, 개인정보를 보내는 곳
  - Android Platform의 행동 묘사
  - Dalvik 핵심 언어(Dalco), 요약 실행

안드로이드 앱의 기본적인 사항

- 소스코드가 공개되어 있지 않음(
- JAVA, C로 작성
  - 일단 JAVA에 집중
- 앱 제작 프로세스



Source / Sink 정리

- 개인정보를 얻는 부분(Source)
- 정보가 외부로 나가는 부분(Sink)
- 방법
  - 안드로이드 소스코드(공개)를 검색
  - API document 참고
  - Taintdroid에 있는 정보 참고

기기 정보(Source)

- READ_PHONE_STATE 권한

- android.telephony.TelephonyManager
  - getLineNumber() – 폰번호
  - getDeviceID() – 국제 핸드폰 식별번호
  - getSubscriberID() – 가입자 ID
  - getSimSerialNumber() – SIM 번호

위치 (Source)

- 콜백함수 등록

- Android.location.LocationListener()
  - 위치가 바뀔때마다 함수를 호출
  - LocationListener에 등록해야함

  - getLatitude(), getLongitude()

Contacts, … (Source)

- 문자열 인자로 얻어오는 데이터 구분

- getContentResolver().query(…)
  - “content://sms” – 문자
  - “content://mms” – MMS
  - “content://calendar” – 달력
  - “content://subscribedfeed” - 피드
  - “browser/bookmarks” – 브라우저 즐겨찾기

네트워크(Sink)

- 다양한 네트워크 접속 방법

- java.net.URLConnection
  - set() – URL 입력 부분
  - getOutputStream() – 네트워크로 보내는 부분

- org.apache.http.client.methods.HttpClient
  - setEntity() – POST로 데이터 전송

이진 코드 분석

- Dalvik 바이트코드
  - 안드로이드 플랫폼에서 주가 되는 명령어
  - 고수준 명령어
    - 가상 레지스터
    - 가상함수, 예외처리
  - 200여개의 명령어
    - 기능을 대표하는 20여개의 명령어로 추림

핵심언어 문법구조

$$
\begin{array}{rcrcl}
pgm & \in & Program & = & MethodTable \times BlockTable \times SuperTypes \times BlockId \\
r & \in & Register & & \\
ty & \in & Type & & \\
id & \in & Id & & \\
bid & \in & BlockId & & \\
s & \in & String & &
\end{array}
$$

$$
\begin{array}{rcl}
MethodTable & = & (Type \times Id) \xrightarrow{\text{fin}} (BlockId \times Register^*) \\
BlockTable & = & BlockId \xrightarrow{\text{fin}} Command \times HandlerTable \times BlockId \\
HandlerTable & = & Type \xrightarrow{\text{fin}} BlockId \\
SuperTypes & = & Type \xrightarrow{\text{fin}} 2^{Type}
\end{array}
$$

핵심언어 문법구조

Data commands

$$
\begin{array}{lcll}
d & \in & D & \\
d & ::= & \texttt{move}\ e\ e & \\
 & | & \texttt{istype}\ e\ e\ ty & \\
 & | & \texttt{new}\ e\ ty & \\
 & | & \texttt{get}\ e\ e\ id & \\
 & | & \texttt{put}\ e\ e\ id &
\end{array}
$$

Control commands

$$
\begin{array}{lcll}
c & \in & C & \\
c & ::= & \texttt{call-direct}\ ty\ id\ e^* & \\
 & | & \texttt{call-virtual}\ id\ e^+ & \\
 & | & \texttt{return}\ e & \\
 & | & \texttt{throw}\ e & \\
 & | & \texttt{jmpnz}\ e\ bid & \\
 & | & \texttt{skip} &
\end{array}
$$

$$
\begin{array}{lcll}
e & \in & E & \\
e & ::= & r & \text{// register} \\
 & | & n & \text{// integer} \\
 & | & s & \text{// string} \\
 & | & \texttt{unit} & \\
 & | & \texttt{X} & \text{// static member storage} \\
 & | & \star E & \text{// unary operator} \\
 & | & E \diamond E & \text{// binary operator}
\end{array}
$$

$$
\begin{array}{rcl}
r & \in & Reg \\
ty & \in & Type \\
id & \in & Identifier \\
bid & \in & BlockId
\end{array}
$$

$$Command = D^*C$$

어떻게 분석하나

- 요약 실행으로 정보 흐름 분석

- Instrumented Value
  - 핵심언어의 실제 실행 의미구조에 실행 시 알아내고 싶은 정보를 삽입

$$InstrValue = Value \times BirthPlace$$

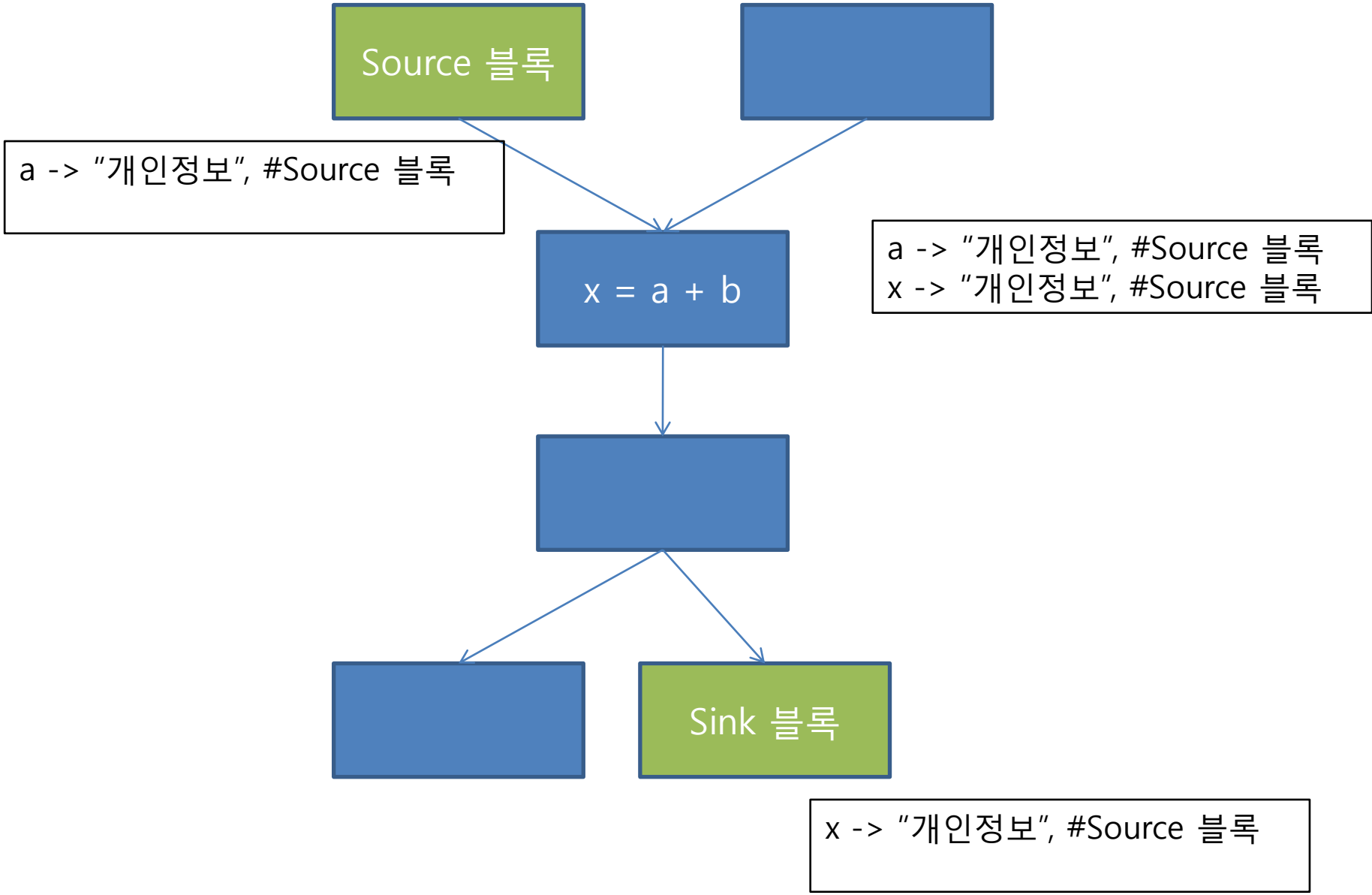
  - 값들마다 어느 위치에서 생성 되었는지를 기록

어떻게 분석하나

- 요약된 Instrumented Value

$$\hat{InstrValue} = \hat{Value} \times 2^{BlockId}$$

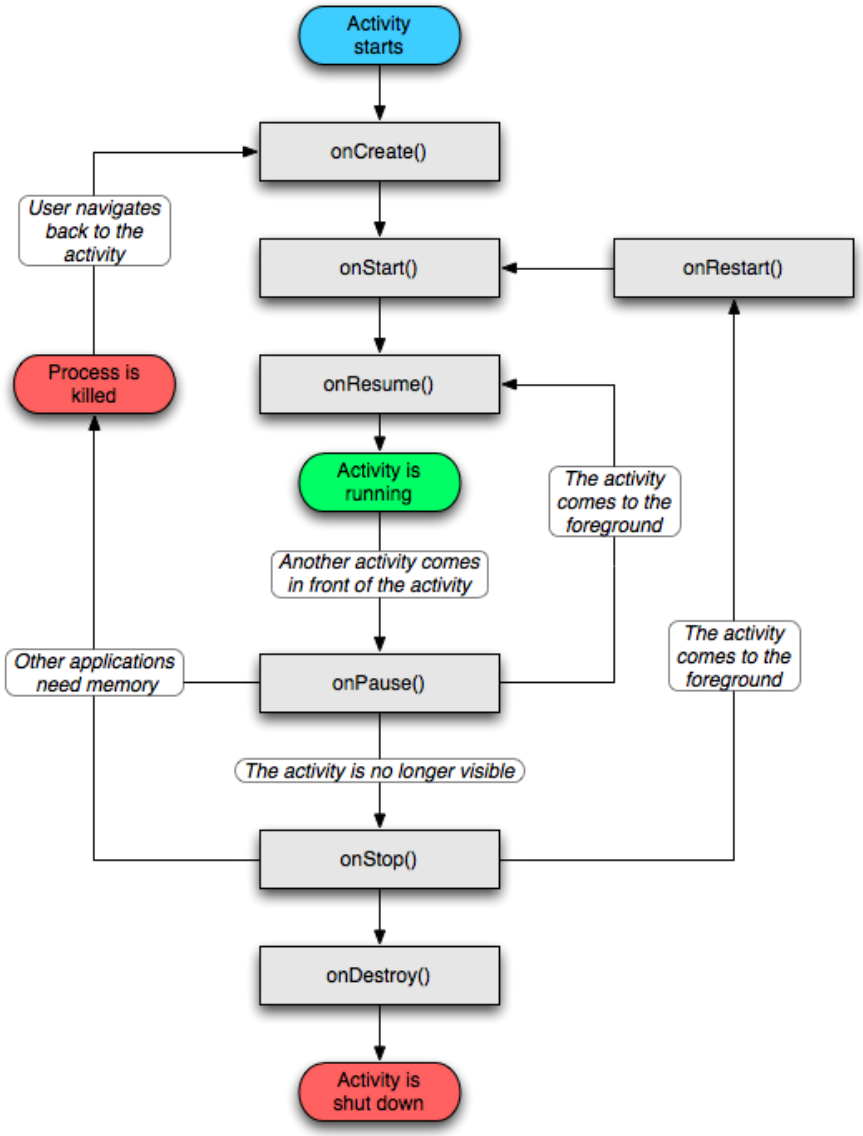
  - 요약실행 후 얻어진 상태에서
    - Sink에 해당하는 블록이
    - Source에 해당하는 BlockId를 가진 요약된 Value를 갖고 있으면 개인 정보유출


Source 블록
a -> "개인정보", #Source 블록
x = a + b
a -> "개인정보", #Source 블록
x -> "개인정보", #Source 블록
Sink 블록
x -> "개인정보", #Source 블록

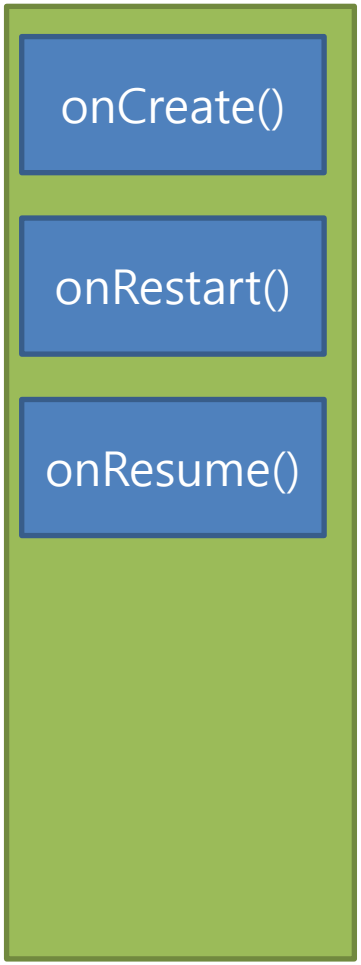
안드로이드 플랫폼이 해주는 일

- 앱 안에 제작된 Activity를 직접 생성하고 이벤트가 발생시 Activity의 함수 호출
- 앱 안에 들어있는 실행 흐름이 아님

안드로이드 플랫폼이 해주는 일

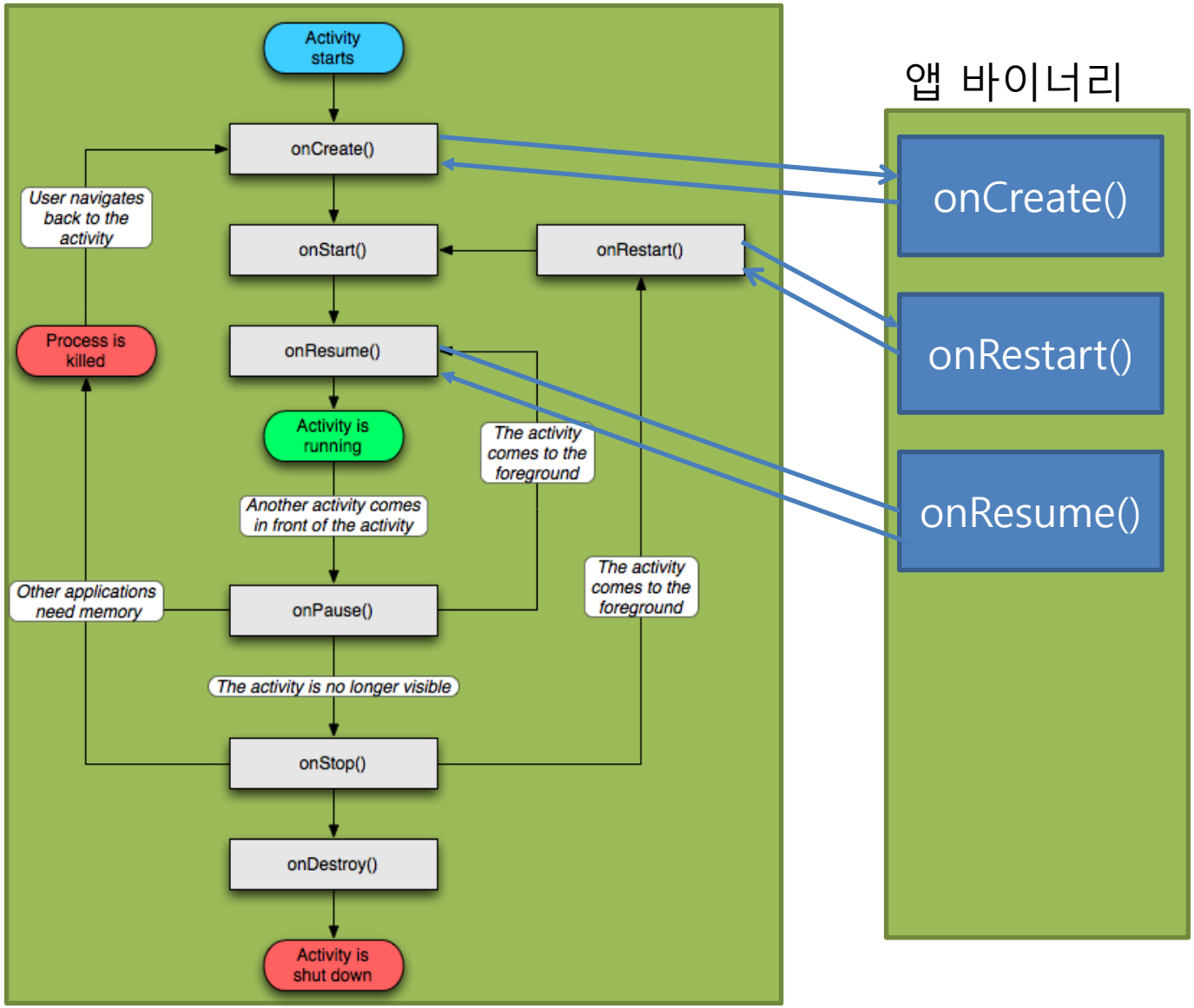


앱 바이너리



안드로이드 플랫폼이 해주는 일

분석기 제공부분



앱 바이너리

기타, 앞단 작성

Dalvik 바이트코드

```
002eea: 7020 1001 ec00
002ef0: 28be
002ef2: 220c 8100
```

Dalvik 핵심언어

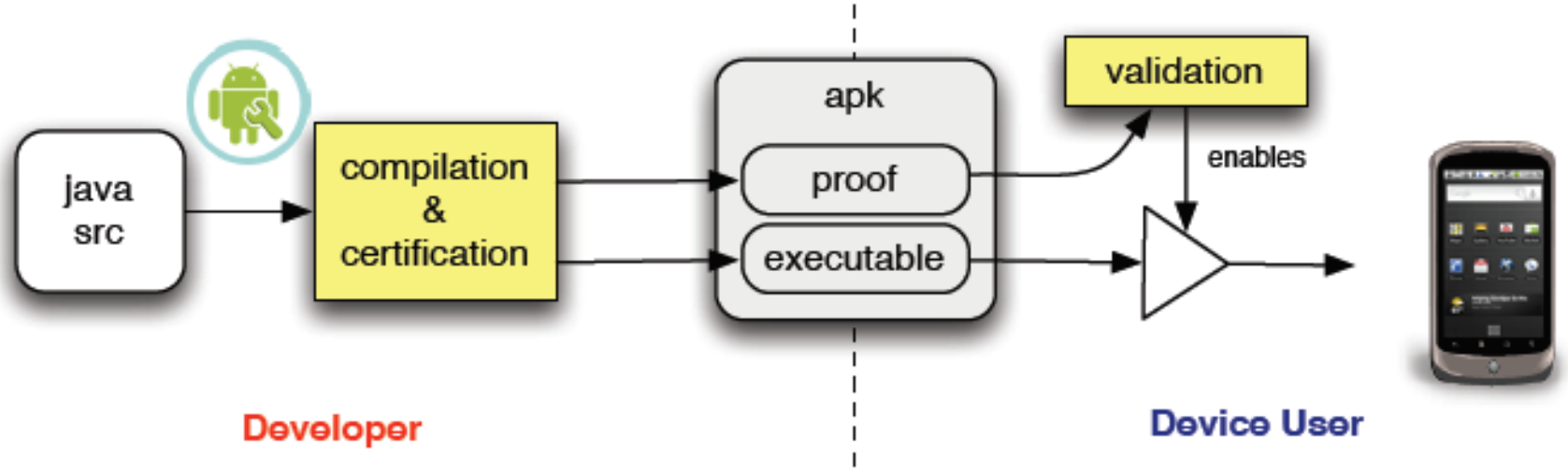
```
call-direct v12, v14, method@0110
jmpnz 000e
new v12, class@0081
```

사용 가능한 시나리오

- 앱 스토어 운영하는 개발사가 앱을 인증
  - 앱 스토어 신뢰성 상승
- 개인이 앱을 분석
- PCC(Proof Carrying Code)

Android PCC

- 증명 확인 시간 << 증명하는 시간
- 개발자가 코드에 증명추가, 받은 사용자가 검증



앱 제작 방식에 대한 연구

- 정교한 분석을 위해
- 요약 실행에서 어떤 정보들을 보존하는 것이 도움이 되는가
  - 정수범위, 문자열, 배열, 동기화 등
- 연구하여 분석기에 반영

정리

- 스마트폰 앱의 개인 정보 유출이 문제
  - 정적으로 분석
- 정적으로 분석하기 위해 필요한 것들
  - Source / Sink 정보
  - Dalvik 핵심언어
  - 안드로이드 플랫폼이 해주는 일 채워 넣기
- 쓰일 수 있는 곳
  - 앱스토어의 앱 인증
  - PCC