Dalvik 바이트코드 분석을 위한 핵심언어 디자인

김진영 윤용호 이승중
서울대학교 프로그래밍 연구실

Dalvik 핵심 언어

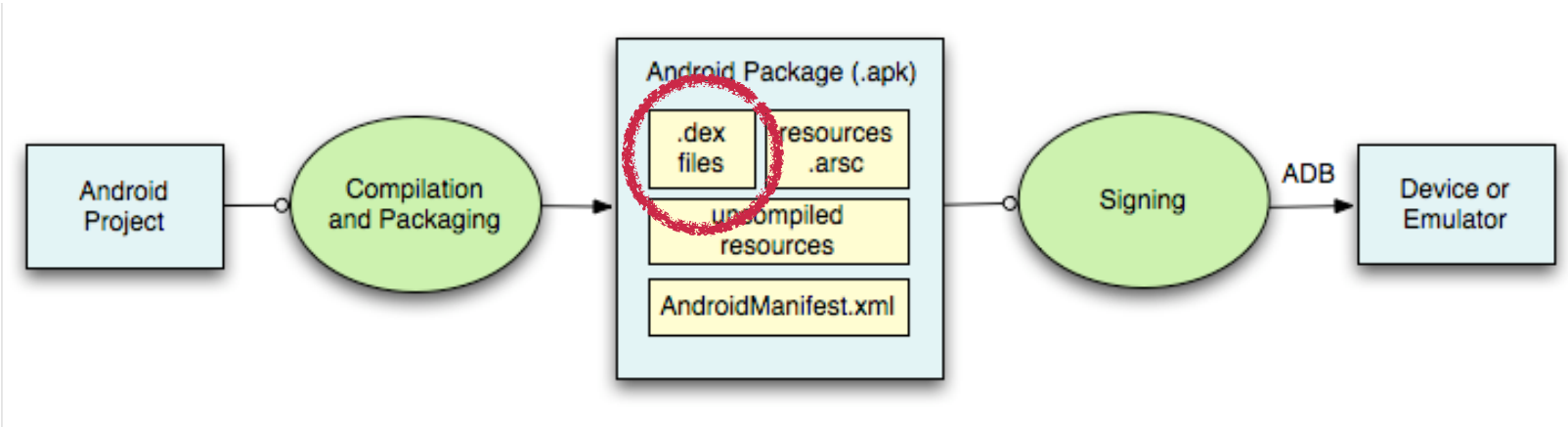
- 무엇을 하려고
- 어떻게 표현되나
- 어떻게 실행되나

무엇을 하려고

- 안드로이드 어플리케이션 분석

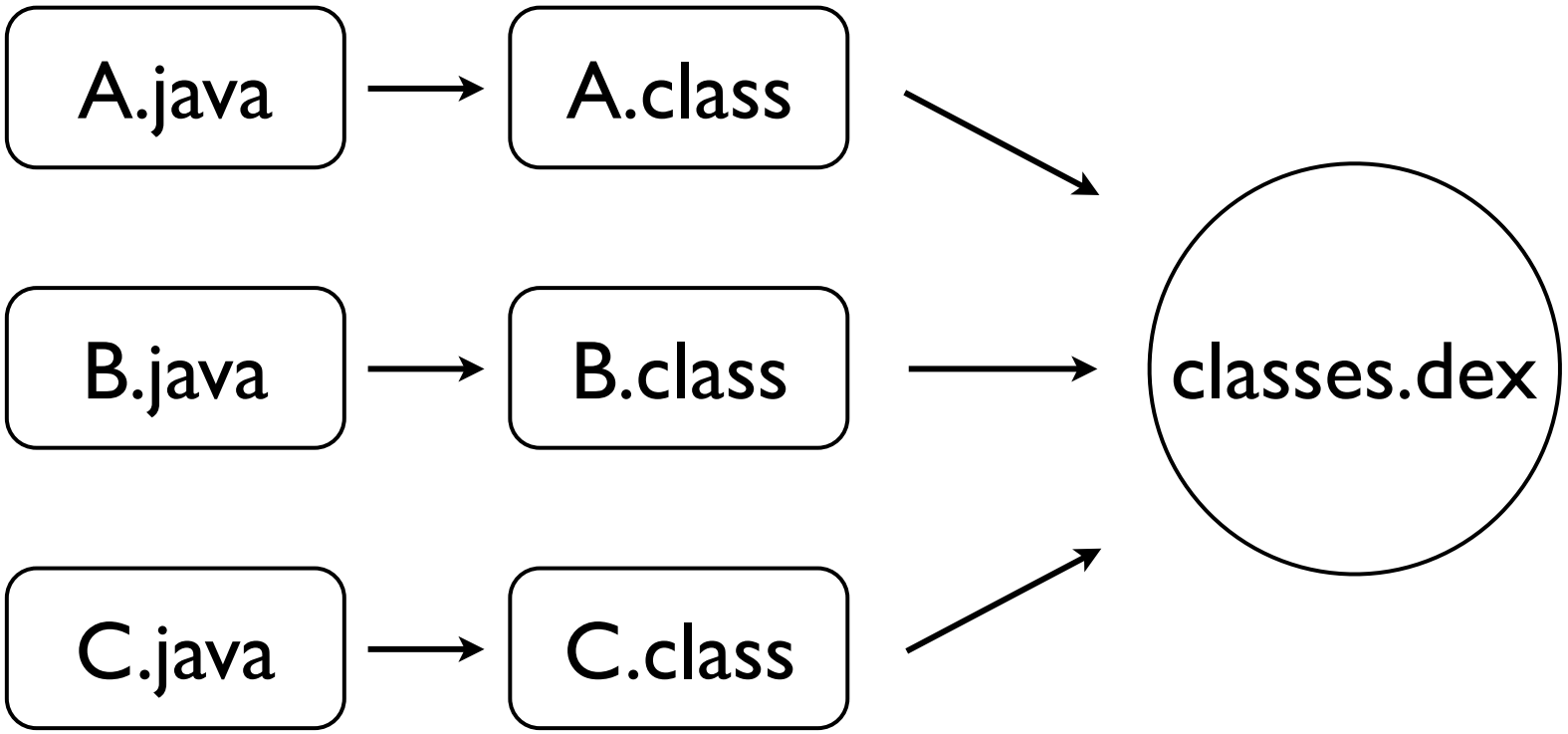
무엇을 하려고

- 안드로이드 어플리케이션 분석
  - Dalvik 가상 머신 (레지스터 기반)
  - Dalvik EXecutable



무엇을 하려고

- Dalvik EXecutable (자바로 구현한)



왜 필요한가

- 분석을 위해서는 실행 의미가 엄밀해야
- Dalvik 바이트코드
  - 200개가 넘는 명령어
  - move, move/16, move-wide, move-object, move-exception, move-result, move-result-wide, move-result-objection, ...
  - 일일히 정의/증명/분석?

왜 필요한가

move e1 e2

`move vA, vB`	move r_A r_B
`move/from16 vAA, vBBBB`	
`move/16 vAAAA, vBBBB`	
`move-object vA, vB`	
`move-object/from16 vAA, vBBBB`	
`move-object/16 vAAAA, vBBBB`	
`move-wide vA, vB`	move t_i r_B; move t_{i+1} r_{B+1}; move r_A t_i; move r_{A+1} t_{i+1}
`move-wide/from16 vAA, vBBBB`	
`move-wide/16 vAAAA, vBBBB`	
`move-result vAA`	move r_A r_{ret}
`move-result-object vAA`	
`move-exception vAA`	move r_A r_{ex}

어떻게 표현되나

Data

- move
- istype
- new
- get
- put

Control

- call-direct
- call-virtual
- return
- throw
- jmpnz
- skip

어떻게 표현되나

```
'com.kakao.talk.dex', DEX version '035'
Class #0            -
  Superclass      : 'Ljava/lang/Object;'
  Interfaces      -
  Static fields   -
  Instance fields -
  #0
    name       : '<init>'
    type       : '()V'
    registers  : 1
...
  #1
    name       : 'main'
    type       : '([Ljava/lang/String;)V'
    registers  : 7
...
```

- MethodTable

어떻게 표현되나

```
main()
  move R5 5
  move R6 10
  move R6 (R6-4)
  ...
  call-direct foo R0 R5 R6
  move R4 Rr
  ...
  call-direct bar R0
  ...

foo()
  ...
  return r3

bar()
  ...
```

- MethodTable
- BlockTable
  - 베이직 블록

어떻게 표현되나

```
main()
  move R5 5
  move R6 10
  move R6 (R6-4)
  ...
  call-direct foo R0 R5 R6
  move R4 Rr
  ...
  call-direct bar R0
  ...

foo()
  ...
  return r3

bar()
  ...
```

- MethodTable
- BlockTable
  - 베이직 블록

어떻게 표현되나

```
catches       : 3
  0x0001 - 0x001a
    <any> -> 0x001b
  0x0037 - 0x003e
    Ljava/lang/Exception; -> 0x003f
  0x004c - 0x0053
    Ljava/lang/Exception; -> 0x0054
```

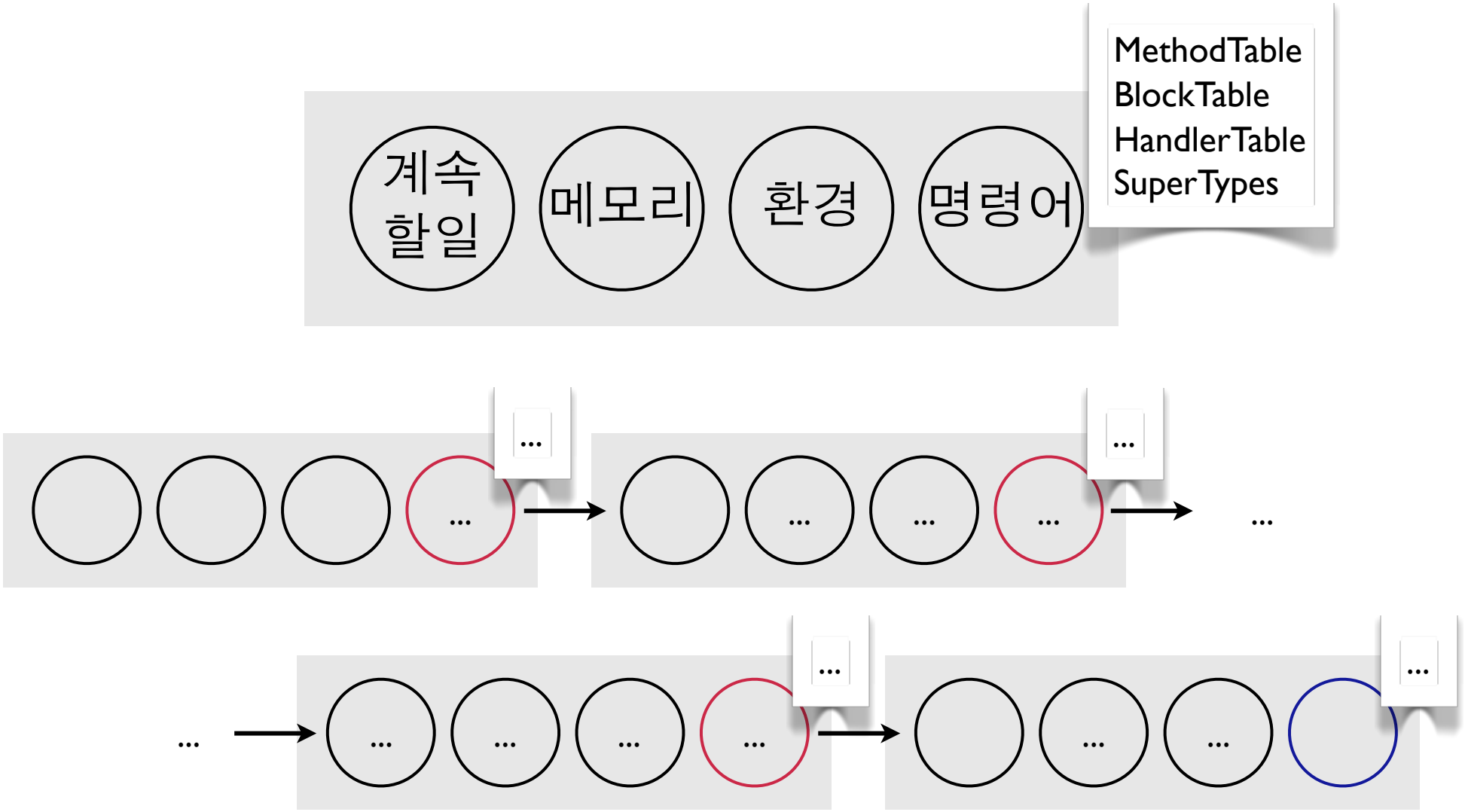
- MethodTable
- BlockTable
  - 베이직 블록
- HandlerTable

어떻게 표현되나

```
Class #19 'vcard/CustomBufferedReader;'
  Superclass : 'io/BufferedReader;'

...

Class #25 'vcard/VCardParser_V30;'
  Superclass : 'vcard/VCardParser_V21;'

...

Class #64 'activity/BaseActivity;'
  Superclass : 'Landroid/app/Activity;'

...
```

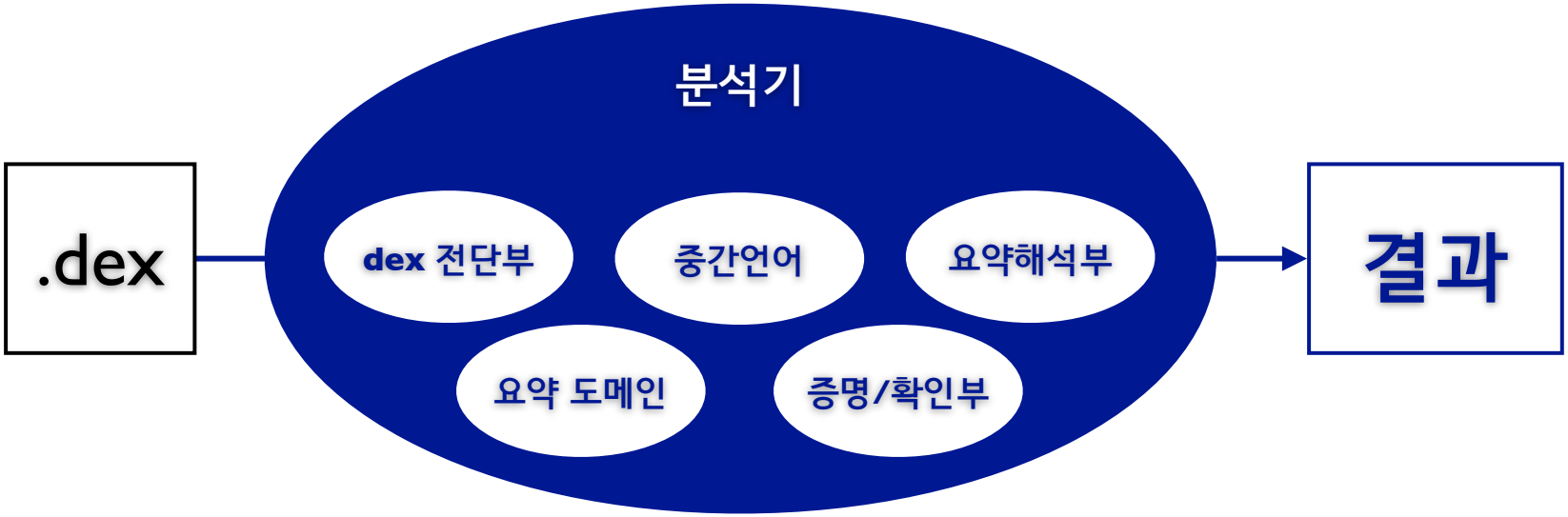
- MethodTable
- BlockTable
  - 베이직 블록
- HandlerTable
- SuperTypes

어떻게 실행되나


MethodTable
BlockTable
HandlerTable
SuperTypes
계속
할일
메모리
환경
명령어
...

Dalvik 핵심 언어

- Dalvik 바이트코드의 실행을 표현할 수 있는 핵심 언어 디자인
- 핵심 언어를 대상으로 안드로이드 앱의 개인정보 유출 여부 정적분석기 제작 중



분석기

- 개인정보들의 생성 (소스)
  - API 함수 호출
  - 휴대전화 식별번호, 위치정보, 주소록, ...
- 새어나가지는 않는가 (싱크)
  - 인터넷 등으로 전송

감사합니다.