RESAEC center
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 NRF ERC

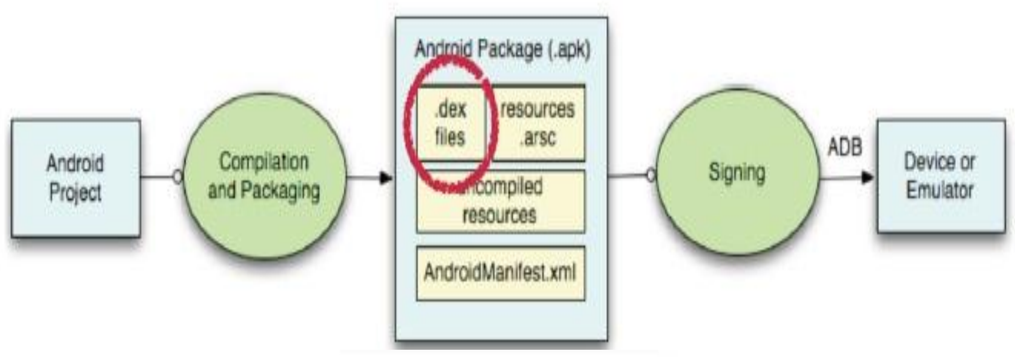
# 정수 범위 도메인의 안전한 달빅 바이트 코드 정적 분석기

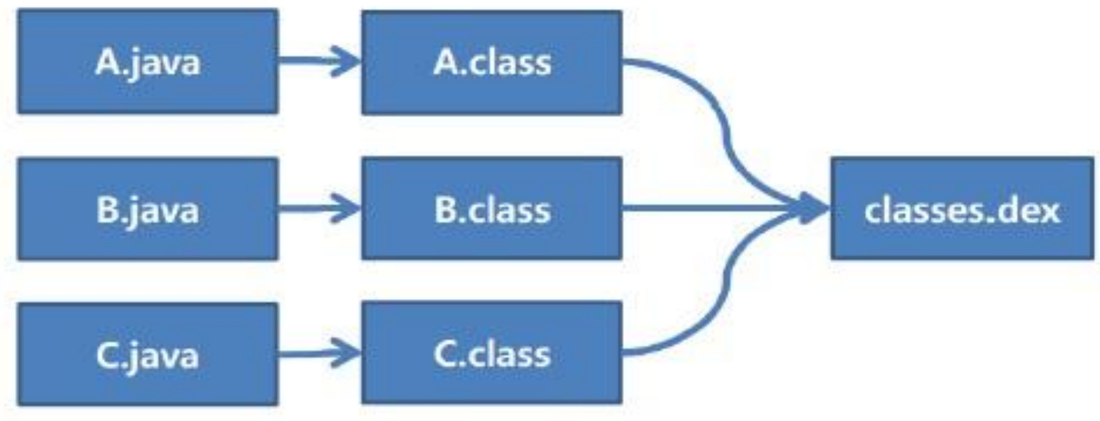
정지수 (서울대학교 프로그래밍 연구실)

## 1. 동기

- 정수 범위 분석을 통해 divide by zero, buffer overrun과 같은 다양한 오류를 정적으로 찾아내기 위함

## 2. 달빅 핵심언어란?

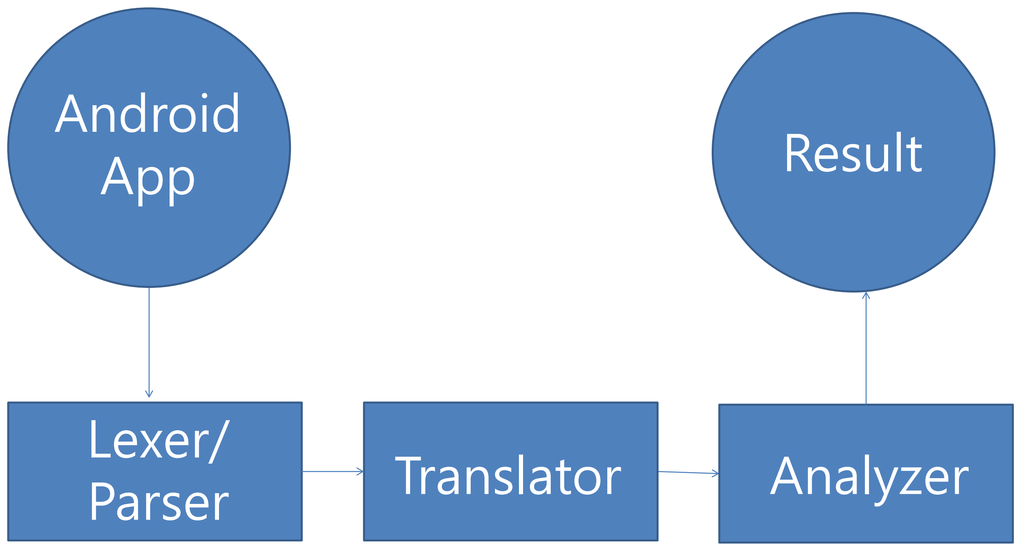



- Register 기반
- 200개 이상의 instruction -> 18개로 축약
- cmd -> data command* control command

| Data Command | Control Command |
|---|---|
| move | call-direct |
| istype | call-indirect |
| new | return |
| get/gets/geta | throw |
| put/puts/puta | jmpnz |
| addcallback | switch |
| | wait |
| | skip |

## 3. 분석과정




## 4. 요약분석

$$
\begin{aligned}
\hat{State} &= \hat{Memory} \times \hat{Environment} \times \hat{CallBacks} \times Command \times ProgramCounter \\
&\quad \times \hat{Continuation} \\
\hat{Memory} &= Location \xrightarrow{\text{fin}} \hat{Object} \\
\hat{Environment} &= Register \xrightarrow{\text{fin}} \hat{Value} \\
\hat{CallBakcs} &= 2^{Id \times \hat{Value}} \\
\hat{Continuation} &= (\hat{Environment} \times BlockId)^* \\
\hat{Location} &= 2^{Location} \\
\hat{Object} &= Type \times (\hat{Record} + \hat{Array}) \\
\hat{Value} &= \hat{\mathbb{Z}} + \hat{Location} + String + Type + \{\bot, \top\} \\
\hat{Record} &= Id \xrightarrow{\text{fin}} \hat{Value} \\
\hat{Array} &= \mathbb{Z} \xrightarrow{\text{fin}} \hat{Value} \\
\hat{\mathbb{Z}} &= (\mathbb{Z} + \{-\infty\}) \times (\mathbb{Z} + \{+\infty\}) + \{\bot\}
\end{aligned}
$$

$$\pi X = \wp(\lambda l.\{(m, \sigma, CB, cmd, p, K) \mid (m, \sigma, CB, cmd, p, K) \in X\}) ProgramCounter$$
$$\hat{\pi} X = \wp(\lambda l.\{(\hat{m}, \hat{\sigma}, \hat{CB}, cmd, p, \hat{K}) \mid (\hat{m}, \hat{\sigma}, \hat{CB}, cmd, p, \hat{K}) \in X\}) ProgramCounter$$

$$\hat{Next} = \wp_{\sqcup} \circ \hat{\pi} \circ \wp_{\cup} \hat{next}$$

- Abstract Interpretation 프레임워크을 통한 안전한 요약
- 고정점 반복 알고리즘을 통한 state의 고정점 찾기
- 축지법과 좁히기를 통해 무한을 유한으로

## 5. 분석기의 성능

| Application | 사이즈(kb) | 시간(second) | |
|---|---|---|---|
| | | Pure | Worklist |
| Kids Preschool Puzzle | 87 | 16.6 | 13.4 |
| Kids Preshool Puzzle | 101 | 74.2 | 33.1 |
| Backgrounds HD Wallpapers | 109 | 10.7 | 5.9 |
| Kids ABC Phonics | 134 | 142.9 | 57.8 |
| Bible Quotes | 138 | 263.7 | 187.9 |
| ES Task Manager | 158 | 258.8 | 198.5 |
| Job Search | 167 | 3370.2 | 1060.3(17m) |
| Multi Touch Paint | 198 | 11.3 | 4.8 |
| Kid Shapes | 225 | 232.7 | 84.6 |
| Adao File Manager | 255 | 17.2 | 10.8 |
| The Day Before | 293 | 51.5 | 19.9 |

- Worklist 알고리즘을 통해 안전한 분석을 적당한 시간 안에 실행
- 사이즈와 시간은 완전히 비례하지는 않음(프로그램 구조 등의 변수)

## 6. 결론

- Abstract Interpretation 프레임워크와 정수 범위 도메인을 통한 안전한 분석기의 디자인, 증명 및 구현 완료
- Worklist 알고리즘을 통한 성능향상
- 사이즈가 큰 특정 앱의 경우 시간이 과다하게 걸림
- Selective join과 같은 정적분석 기술의  추가도입으로 성능향상 예정

Programming Research Laboratory

SEOUL NATIONAL UNIVERSITY