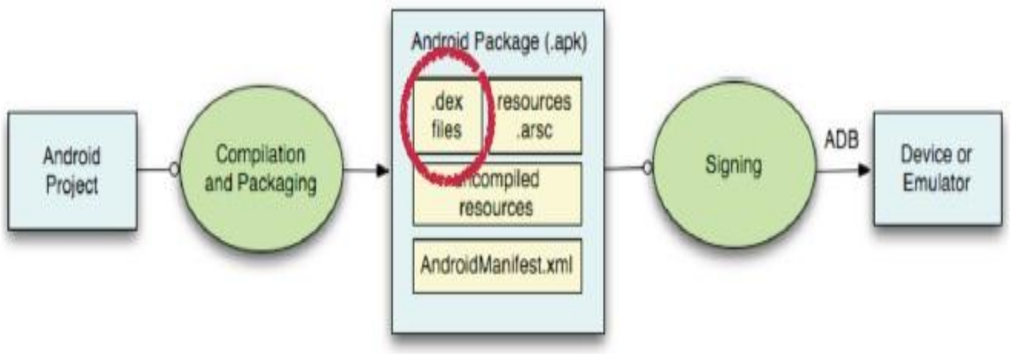
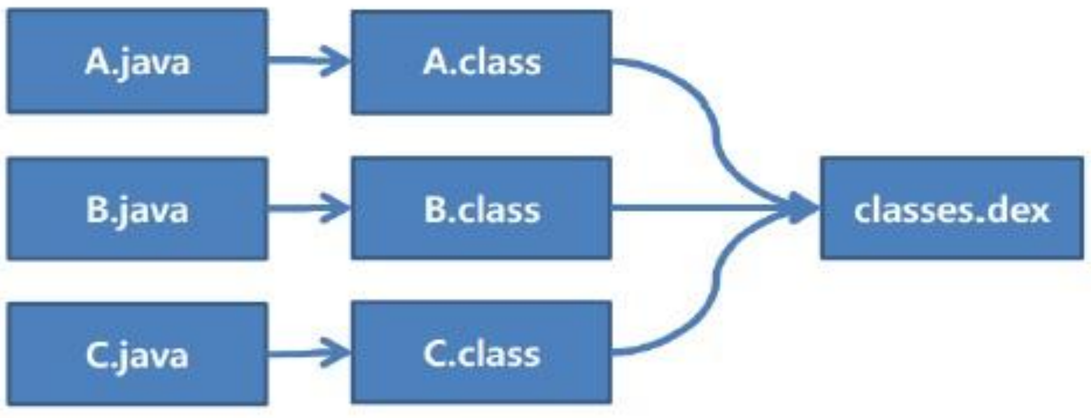
ROSAEC center
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 NRF ERC

달빅 바이트코드의 정수 범위 도메인 분석

정지수 (서울대학교 프로그래밍 연구실)

1. 달빅 핵심언어란?





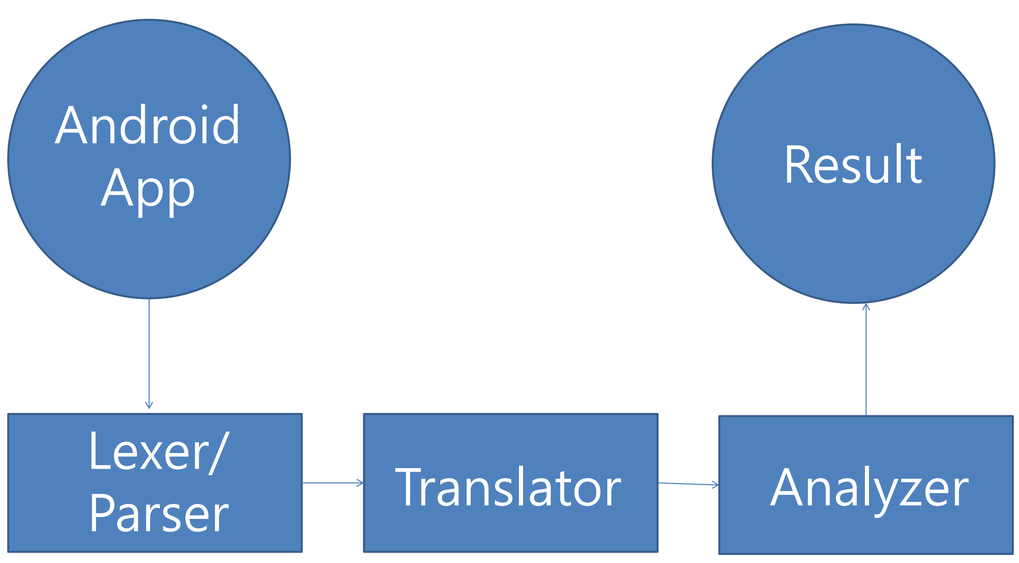
- Register 기반
- 200개 이상의 instruction -> 18개로 축약
- cmd -> data command* control command

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

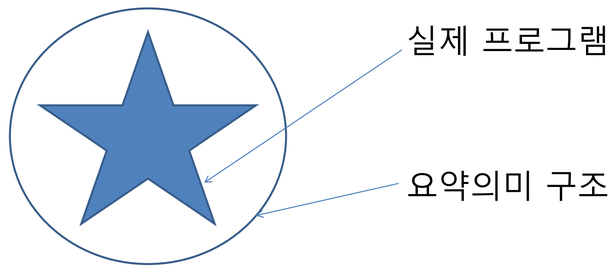
2. 동기

- 정수 범위 분석은 정적분석에서 buffer overrun과 같은 다양한 오류를 찾아내는 데 사용됨.
- Android app의 오류 분석을 위해 필요한 주춧돌 쌓기

3. 분석과정



4. 요약분석



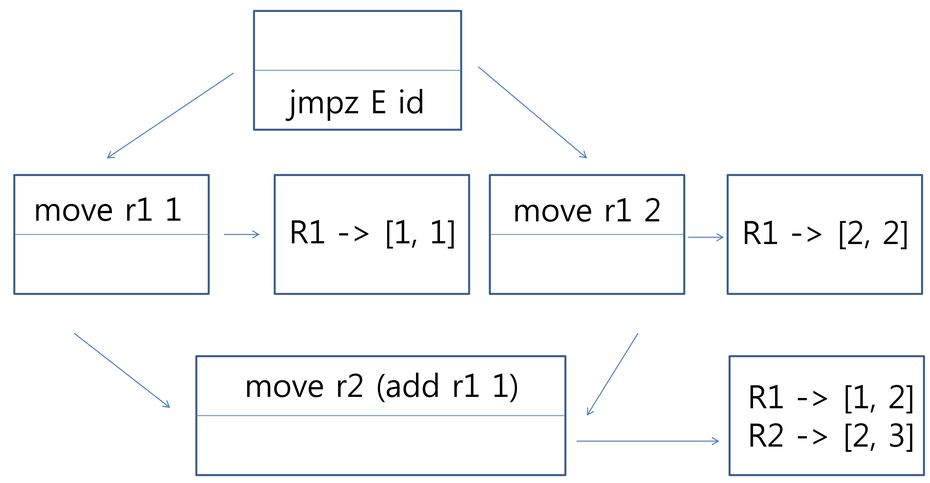
$$
\begin{aligned}
\hat{State} &= \hat{Memory} \times \hat{Environment} \times \hat{CallBacks} \times Command \times ProgramCounter \\
&\quad \times \hat{Continuation} \\
\hat{Memory} &= Location \xrightarrow{\text{fin}} \hat{Object} \\
\hat{Environment} &= Register \xrightarrow{\text{fin}} \hat{Value}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

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

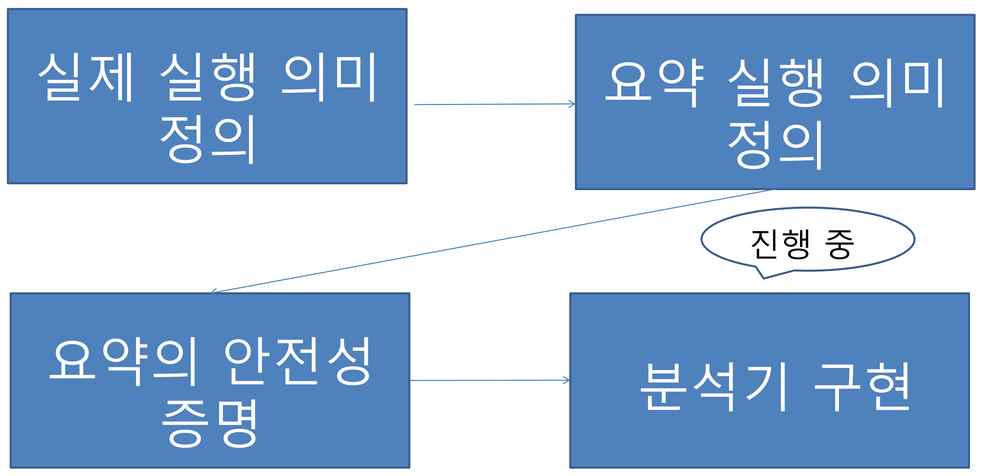
$$\hat{\pi} X = \wp(\lambda l.\{(\hat{m}, \hat{\sigma}, \hat{CB}, cmd, p, \hat{K}) \mid (\hat{m}, \hat{\sigma}, \hat{CB}, cmd, p, \hat{K}) \in X\})ProgramCounter$$

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

- 고정점 반복 알고리즘을 통한 state의 고정점 찾기
- 축지법과 좁히기의 사용



5. 진행 상황



Programming Research Laboratory

SEOUL NATIONAL UNIVERSITY