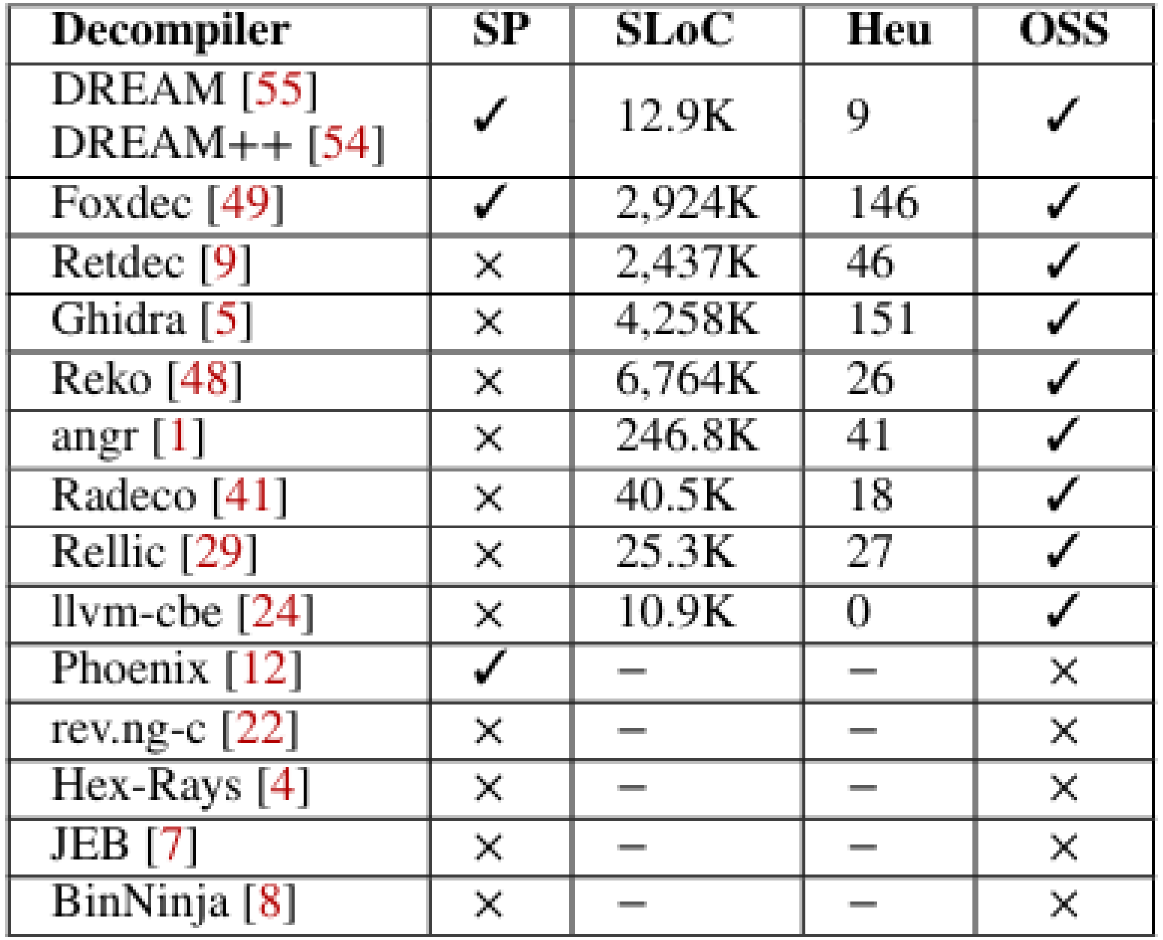
D-Helix: A Generic Decompiler Testing Framework Using Symbolic Differentiation

Motivation

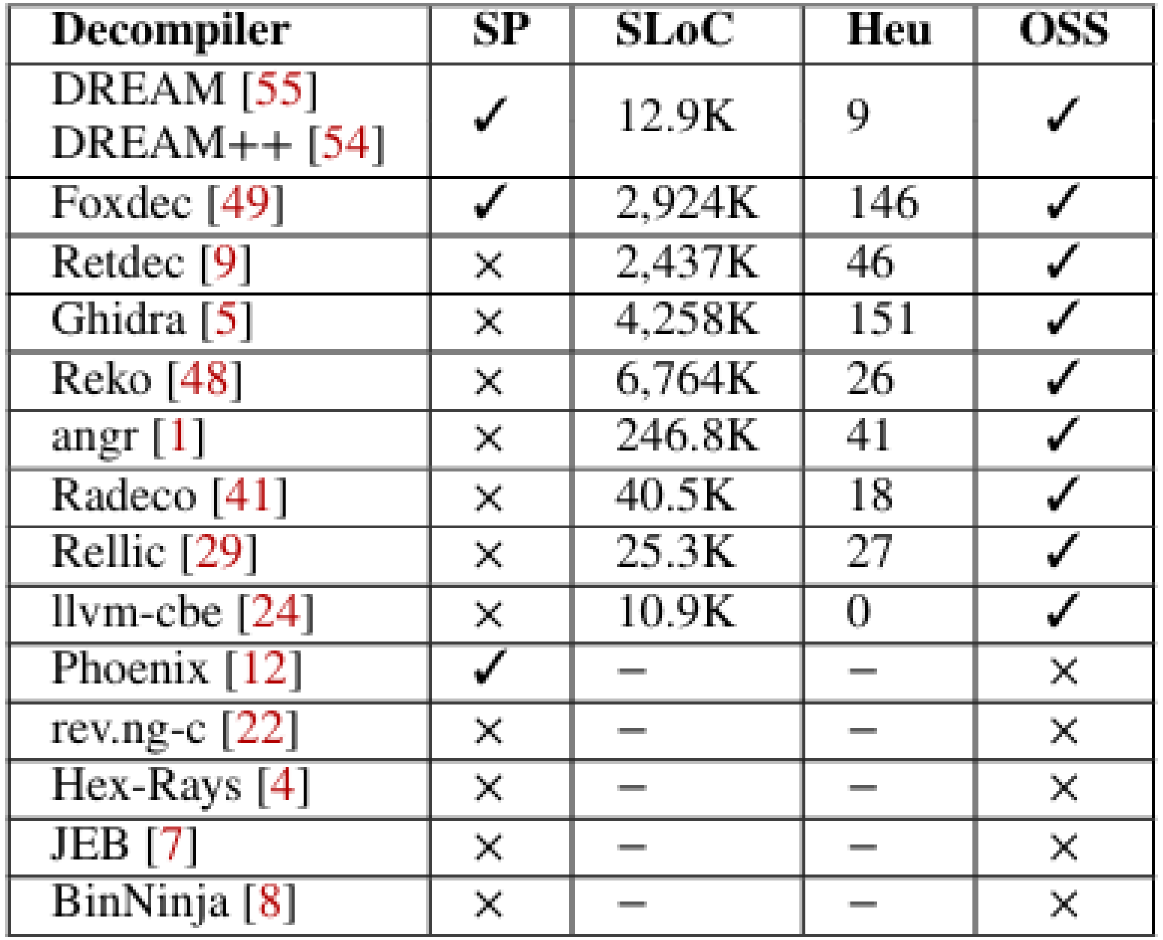


Decompiler	SP	SLoC	Heu	OSS
DREAM [55] DREAM++ [54]	✓	12.9K	9	✓
Foxdec [49]	✓	2,924K	146	✓
Retdec [9]	×	2,437K	46	✓
Ghidra [5]	×	4,258K	151	✓
Reko [48]	×	6,764K	26	✓
angr [1]	×	246.8K	41	✓
Radeco [41]	×	40.5K	18	✓
Rellic [29]	×	25.3K	27	✓
llvm-cbe [24]	×	10.9K	0	✓
Phoenix [12]	✓	–	–	×
rev.ng-c [22]	×	–	–	×
Hex-Rays [4]	×	–	–	×
JEB [7]	×	–	–	×
BinNinja [8]	×	–	–	×

Systematization of decompilers and their characteristics.

SP = Semantic-Preserving, Heu = Heuristics, and OSS =Open Source Software.

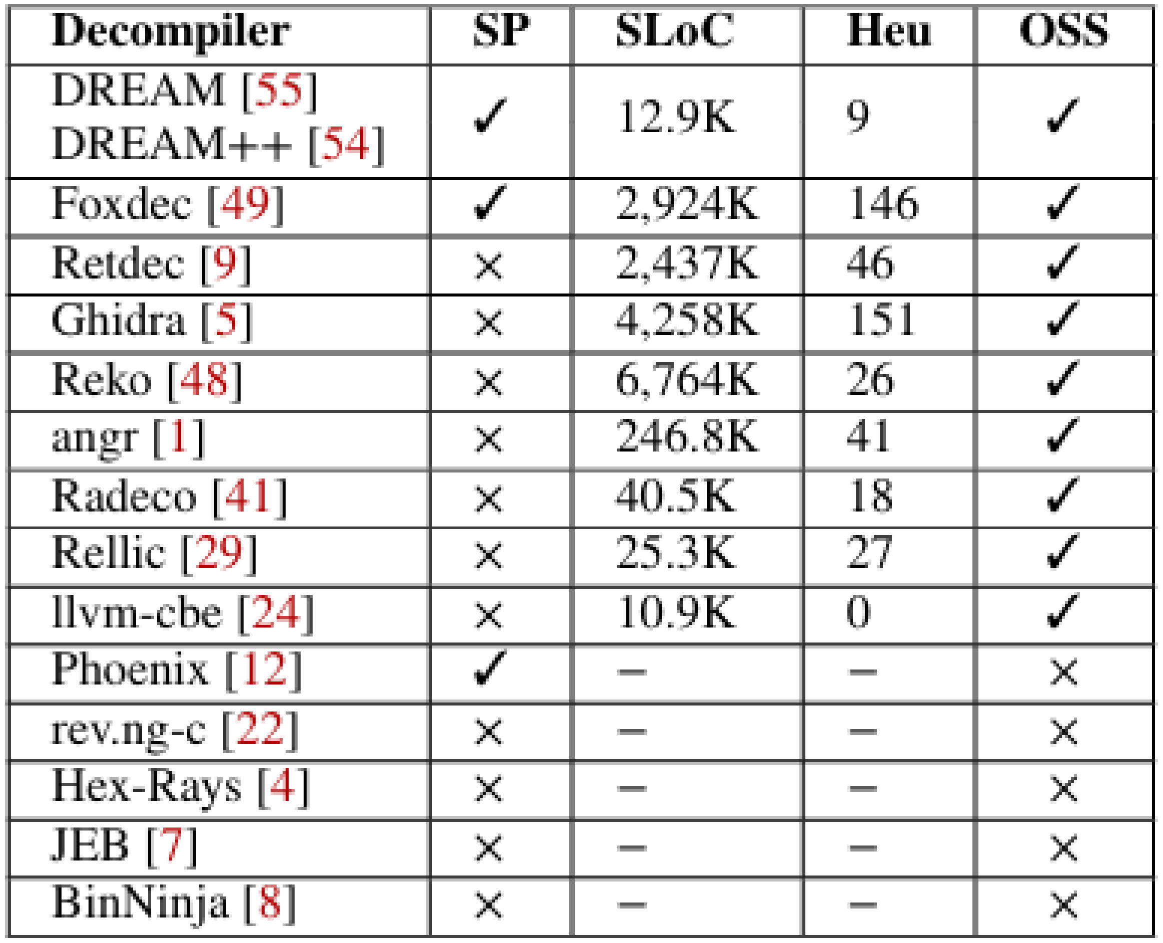
Motivation



Decompiler	SP	SLoC	Heu	OSS
DREAM [55] DREAM++ [54]	✓	12.9K	9	✓
Foxdec [49]	✓	2,924K	146	✓
Retdec [9]	×	2,437K	46	✓
Ghidra [5]	×	4,258K	151	✓
Reko [48]	×	6,764K	26	✓
angr [1]	×	246.8K	41	✓
Radeco [41]	×	40.5K	18	✓
Rellic [29]	×	25.3K	27	✓
llvm-cbe [24]	×	10.9K	0	✓
Phoenix [12]	✓	–	–	×
rev.ng-c [22]	×	–	–	×
Hex-Rays [4]	×	–	–	×
JEB [7]	×	–	–	×
BinNinja [8]	×	–	–	×

1.Decompilers tend to overlook the importance of ensuring the semantic preservation of their decompiled code

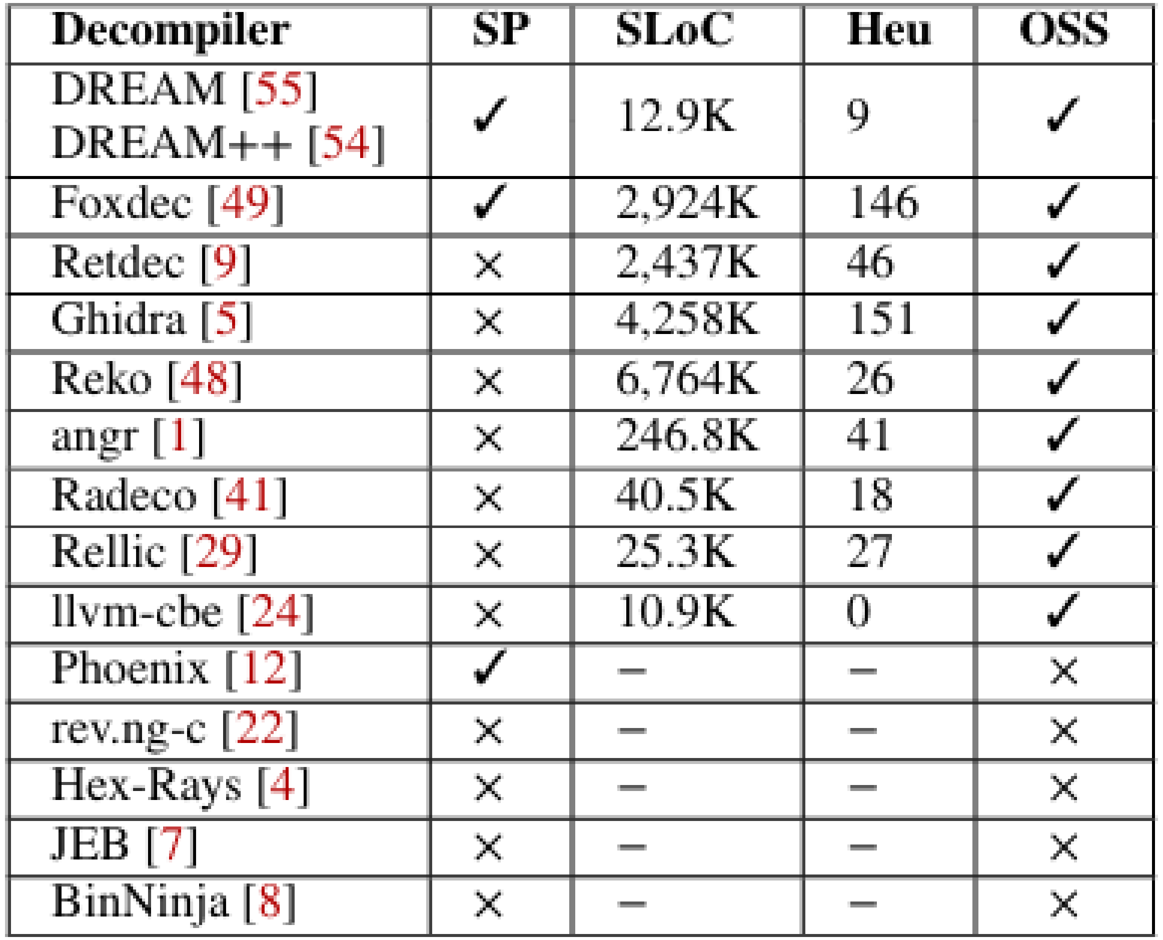
Motivation



Decompiler	SP	SLoC	Heu	OSS
DREAM [55] DREAM++ [54]	✓	12.9K	9	✓
Foxdec [49]	✓	2,924K	146	✓
Retdec [9]	×	2,437K	46	✓
Ghidra [5]	×	4,258K	151	✓
Reko [48]	×	6,764K	26	✓
angr [1]	×	246.8K	41	✓
Radeco [41]	×	40.5K	18	✓
Rellic [29]	×	25.3K	27	✓
llvm-cbe [24]	×	10.9K	0	✓
Phoenix [12]	✓	–	–	×
rev.ng-c [22]	×	–	–	×
Hex-Rays [4]	×	–	–	×
JEB [7]	×	–	–	×
BinNinja [8]	×	–	–	×

2.There lacks a generic methodology, which can soundly examine the decompilers

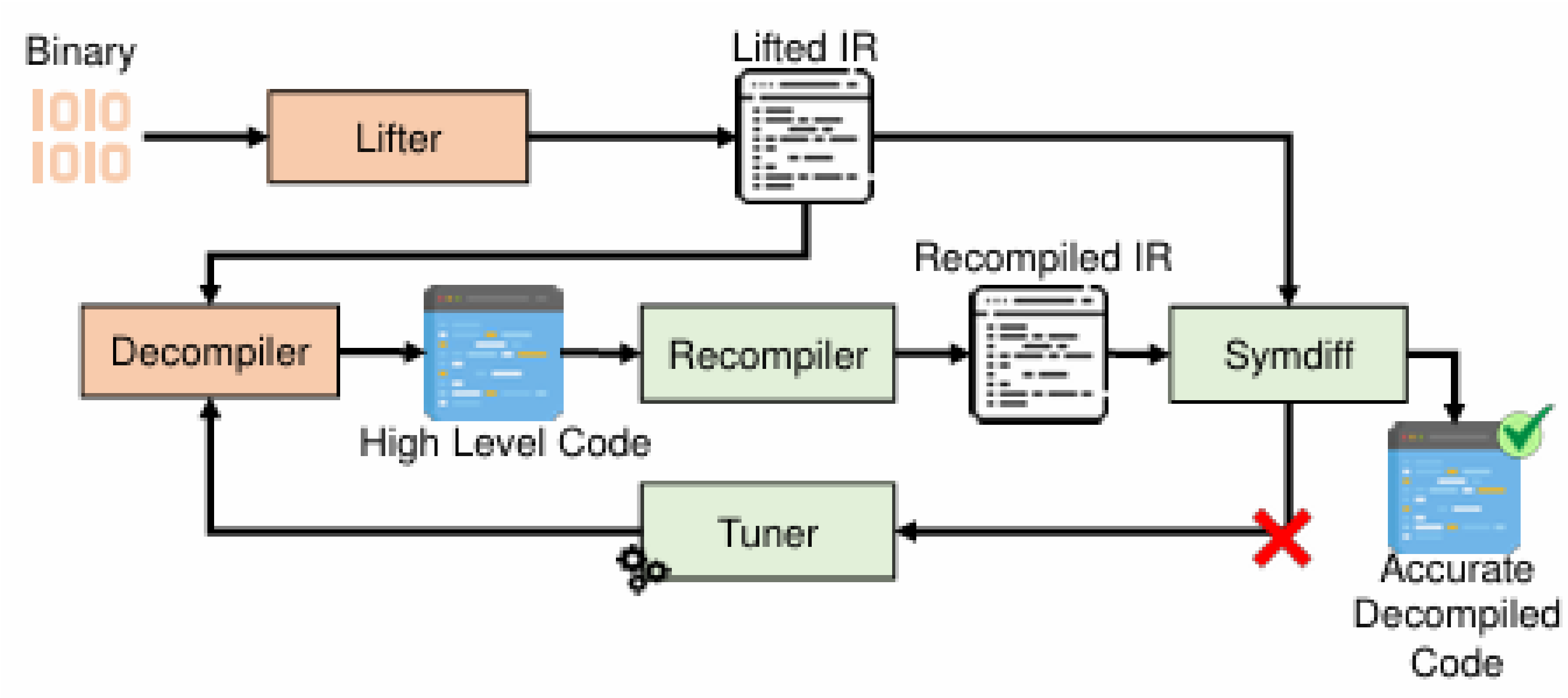
Motivation



Decompiler	SP	SLoC	Heu	OSS
DREAM [55] DREAM++ [54]	✓	12.9K	9	✓
Foxdec [49]	✓	2,924K	146	✓
Retdec [9]	×	2,437K	46	✓
Ghidra [5]	×	4,258K	151	✓
Reko [48]	×	6,764K	26	✓
angr [1]	×	246.8K	41	✓
Radeco [41]	×	40.5K	18	✓
Rellic [29]	×	25.3K	27	✓
llvm-cbe [24]	×	10.9K	0	✓
Phoenix [12]	✓	–	–	×
rev.ng-c [22]	×	–	–	×
Hex-Rays [4]	×	–	–	×
JEB [7]	×	–	–	×
BinNinja [8]	×	–	–	×

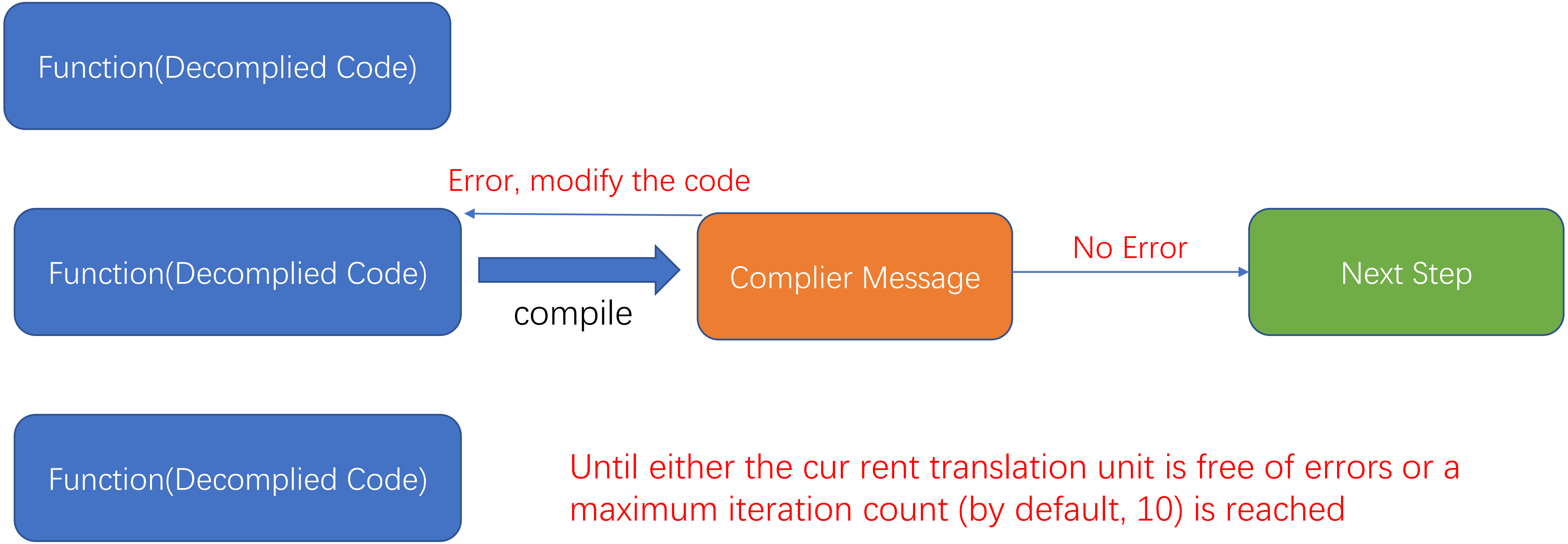
3.It is hard to debug root causes of semantic inaccuracies in decompilers

D-HELIX



D-HELIX pipeline

RECOMPLIER



Function(Decomplied Code)

Function(Decomplied Code)

Function(Decomplied Code)

Error, modify the code

compile

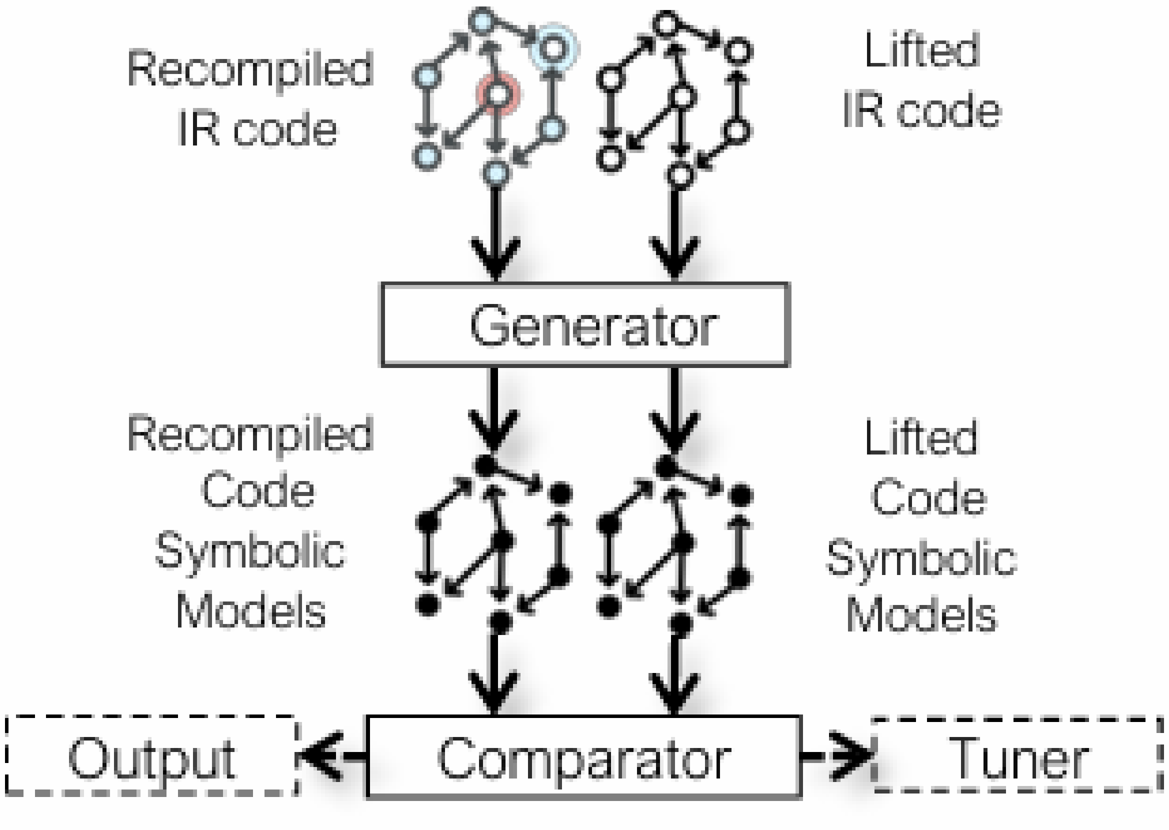
Complier Message

No Error

Next Step

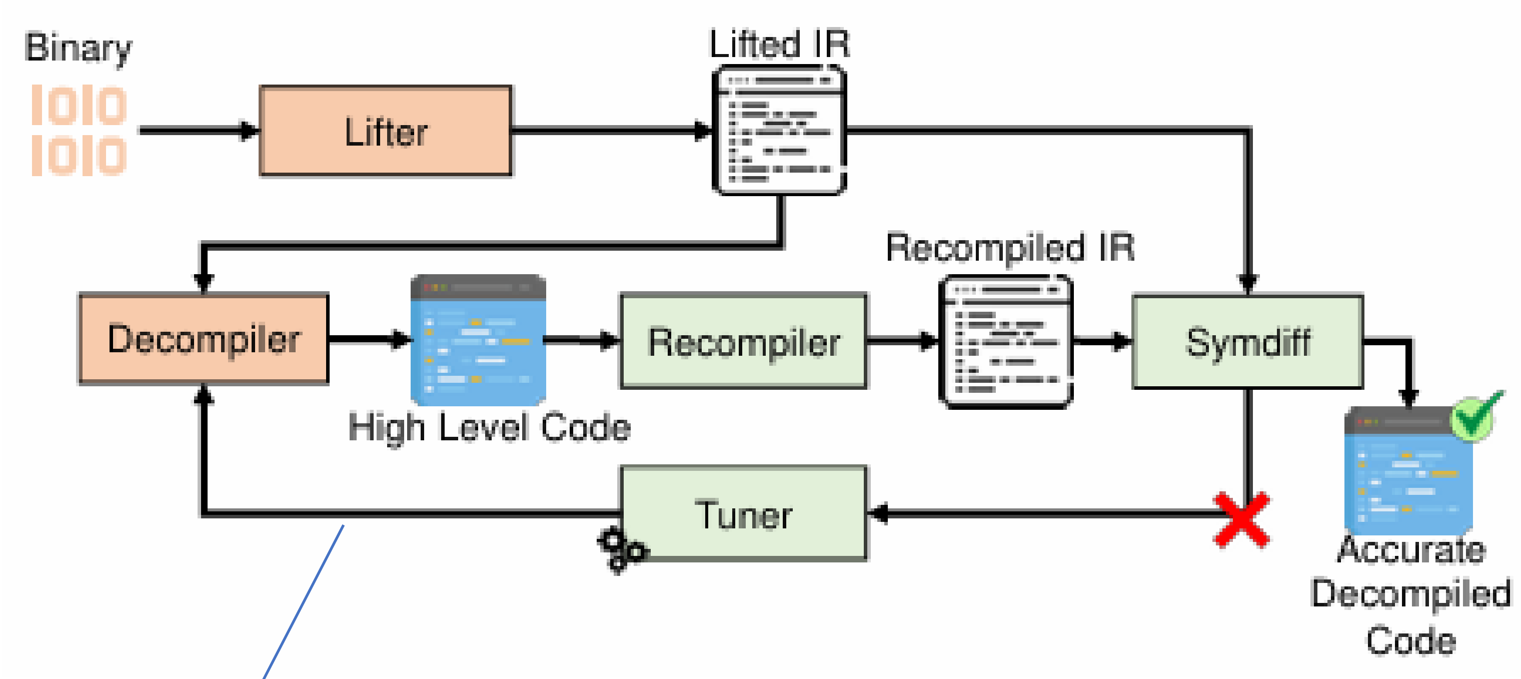
Until either the cur rent translation unit is free of errors or a maximum iteration count (by default, 10) is reached

SYMDIFF



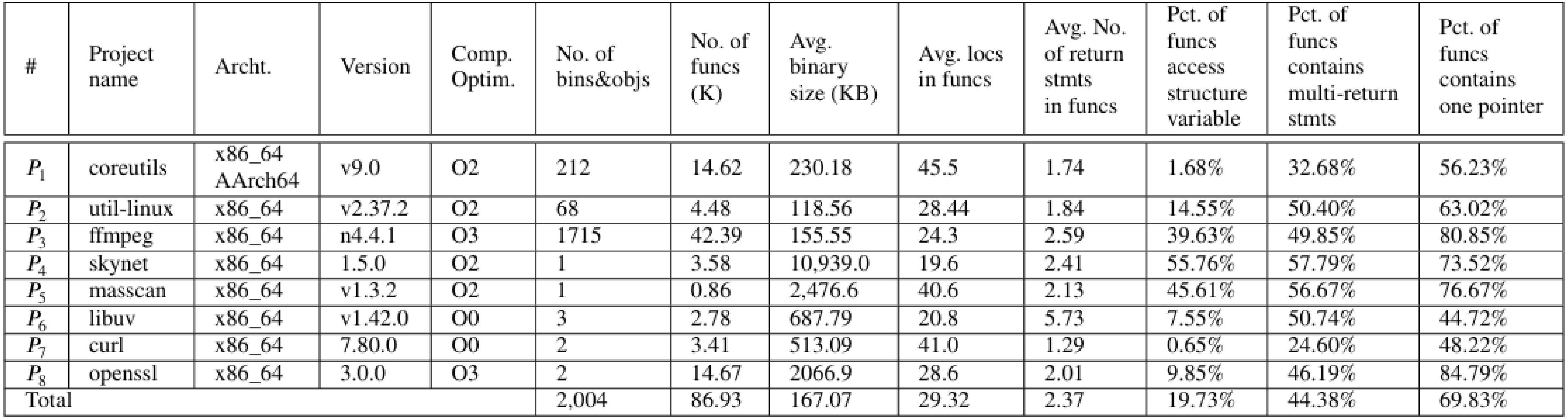
Comparator compares symbolic models, generated by Generator

TUNER



Apply Different Heuristic Rules

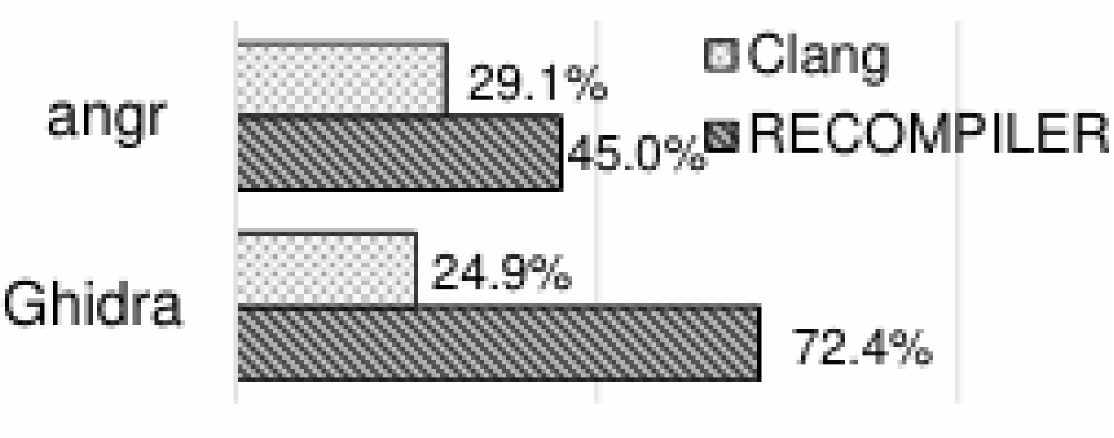
Evaluation



#	Project name	Archt.	Version	Comp. Optim.	No. of bins&objs	No. of funcs (K)	Avg. binary size (KB)	Avg. locs in funcs	Avg. No. of return stmts in funcs	Pct. of funcs access structure variable	Pct. of funcs contains multi-return stmts	Pct. of funcs contains one pointer
P_1	coreutils	x86_64 AArch64	v9.0	O2	212	14.62	230.18	45.5	1.74	1.68%	32.68%	56.23%
P_2	util-linux	x86_64	v2.37.2	O2	68	4.48	118.56	28.44	1.84	14.55%	50.40%	63.02%
P_3	ffmpeg	x86_64	n4.4.1	O3	1715	42.39	155.55	24.3	2.59	39.63%	49.85%	80.85%
P_4	skynet	x86_64	1.5.0	O2	1	3.58	10,939.0	19.6	2.41	55.76%	57.79%	73.52%
P_5	masscan	x86_64	v1.3.2	O2	1	0.86	2,476.6	40.6	2.13	45.61%	56.67%	76.67%
P_6	libuv	x86_64	v1.42.0	O0	3	2.78	687.79	20.8	5.73	7.55%	50.74%	44.72%
P_7	curl	x86_64	7.80.0	O0	2	3.41	513.09	41.0	1.29	0.65%	24.60%	48.22%
P_8	openssl	x86_64	3.0.0	O3	2	14.67	2066.9	28.6	2.01	9.85%	46.19%	84.79%
Total					2,004	86.93	167.07	29.32	2.37	19.73%	44.38%	69.83%

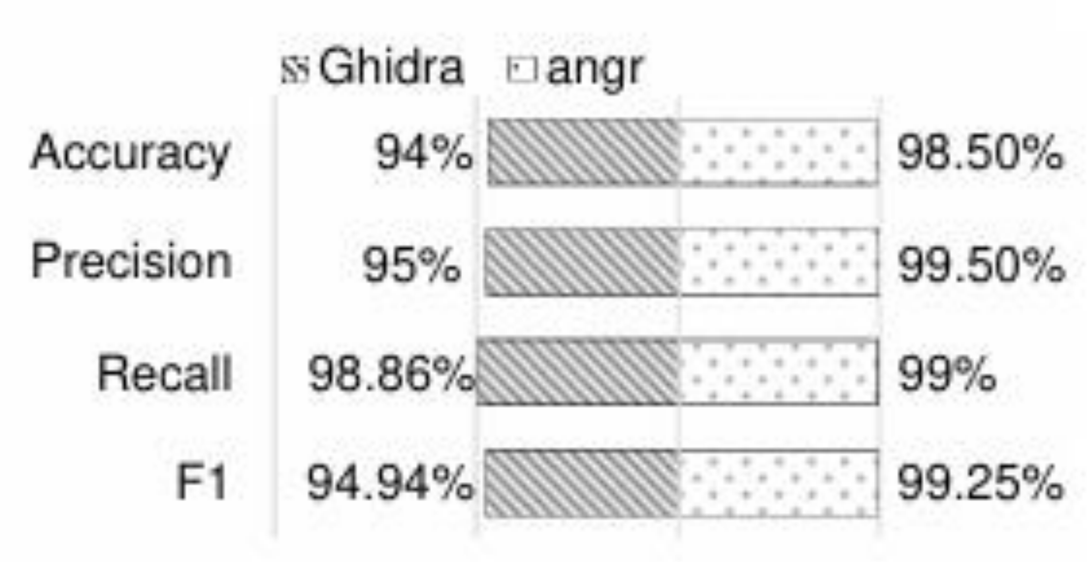
found a total of 25 (17 previously unknown) bugs in the two decompilers (Ghidra and angr)

Evaluation - RECOMPILER



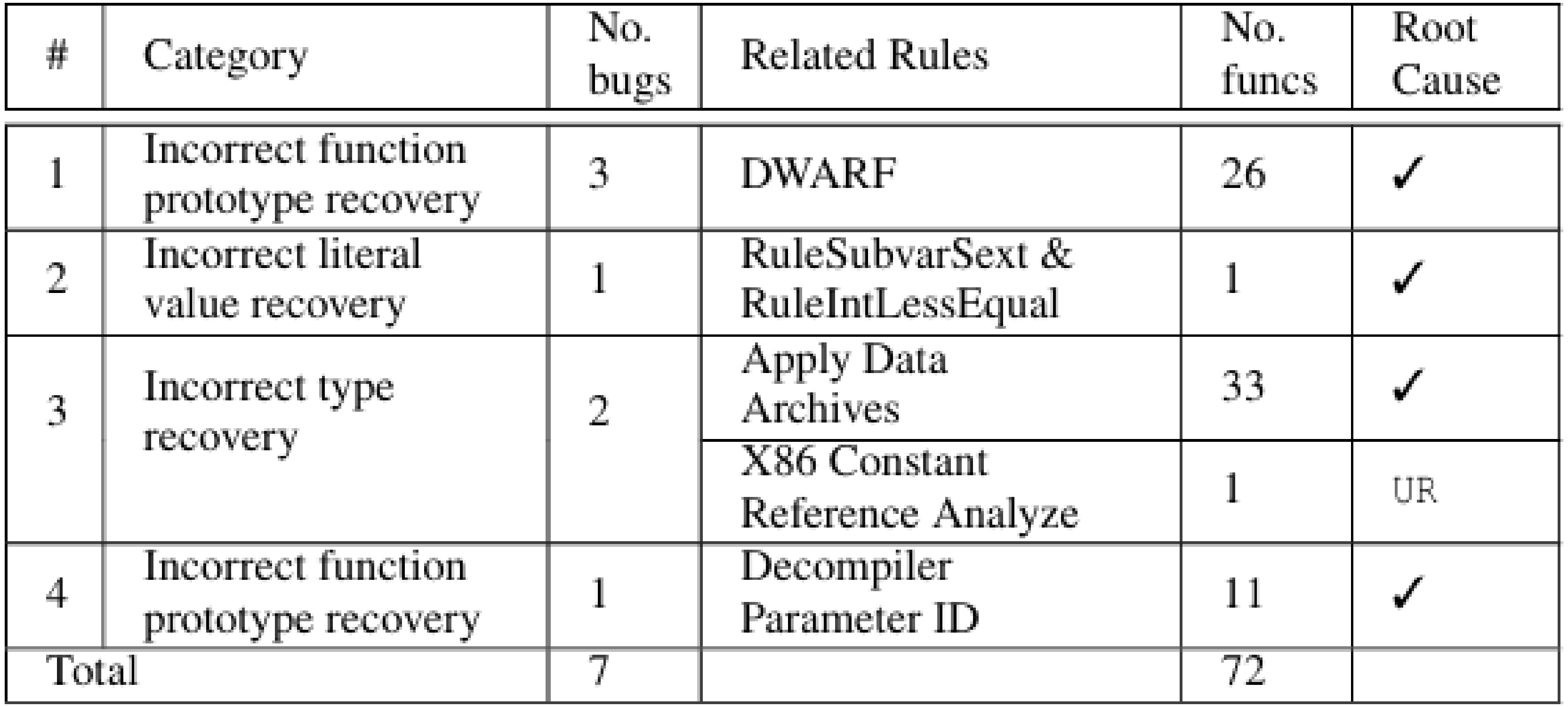
The percentage of functions that can be compiled after using RECOMPILER

Evaluation - SYMDIFF



The accuracy, precision recall and F1 score of SYMDIFF on the tested decompilers

Evaluation - TUNER



#	Category	No. bugs	Related Rules	No. funcs	Root Cause
1	Incorrect function prototype recovery	3	DWARF	26	✓
2	Incorrect literal value recovery	1	RuleSubvarSext & RuleIntLessEqual	1	✓
3	Incorrect type recovery	2	Apply Data Archives	33	✓
			X86 Constant Reference Analyze	1	UR
4	Incorrect function prototype recovery	1	Decompiler Parameter ID	11	✓
Total		7		72	

Summary of bugs in Ghidra that can be fixed by the TUNER.

End