Incremental CSA

Philipp Dominik Schubert, Balázs Benics

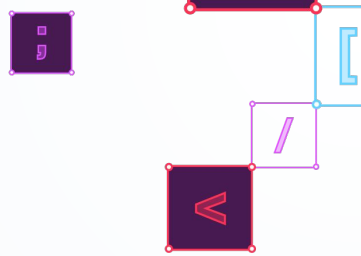
sonar

Agenda

Motivation

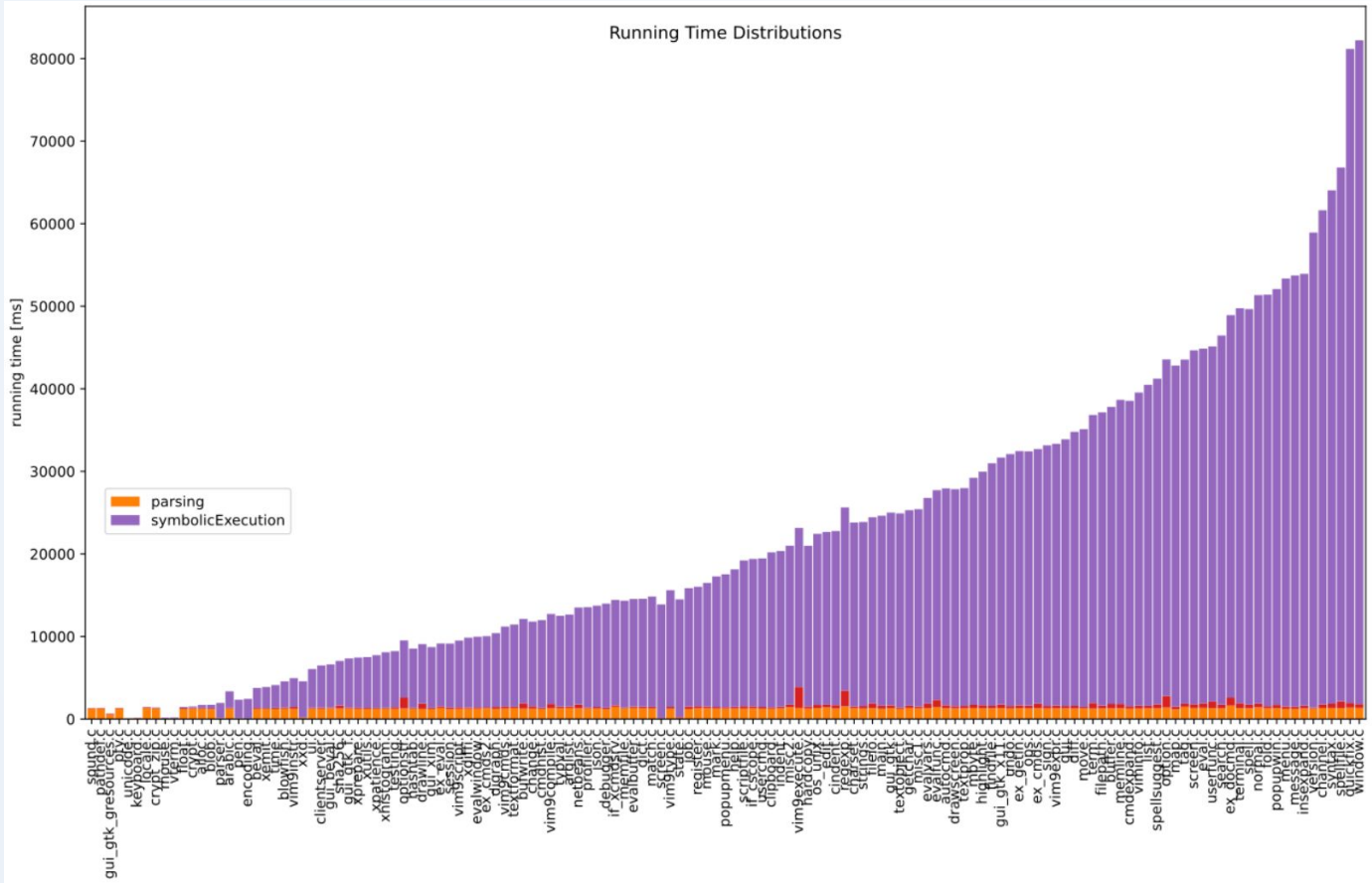
The prototype

Questions



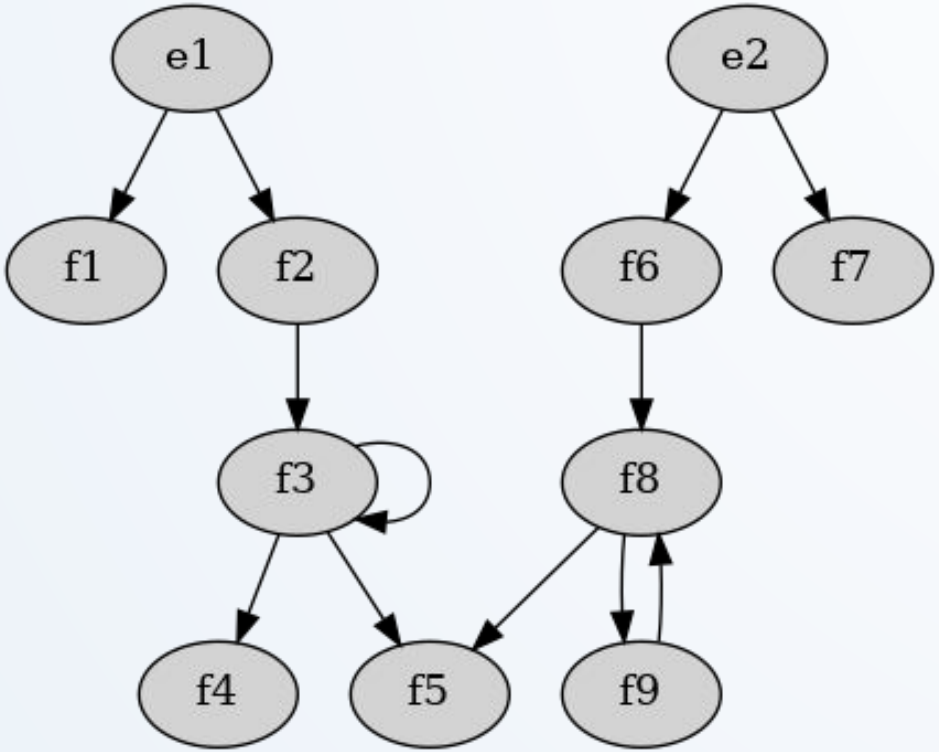
sonar

Motivation

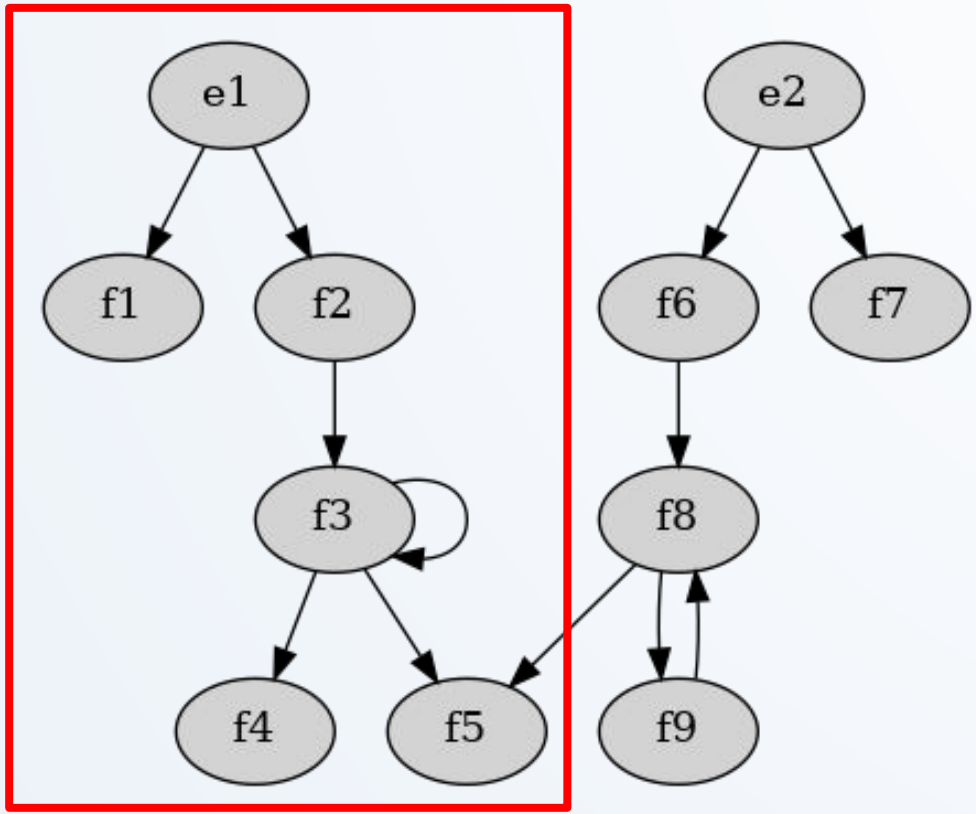


Vim

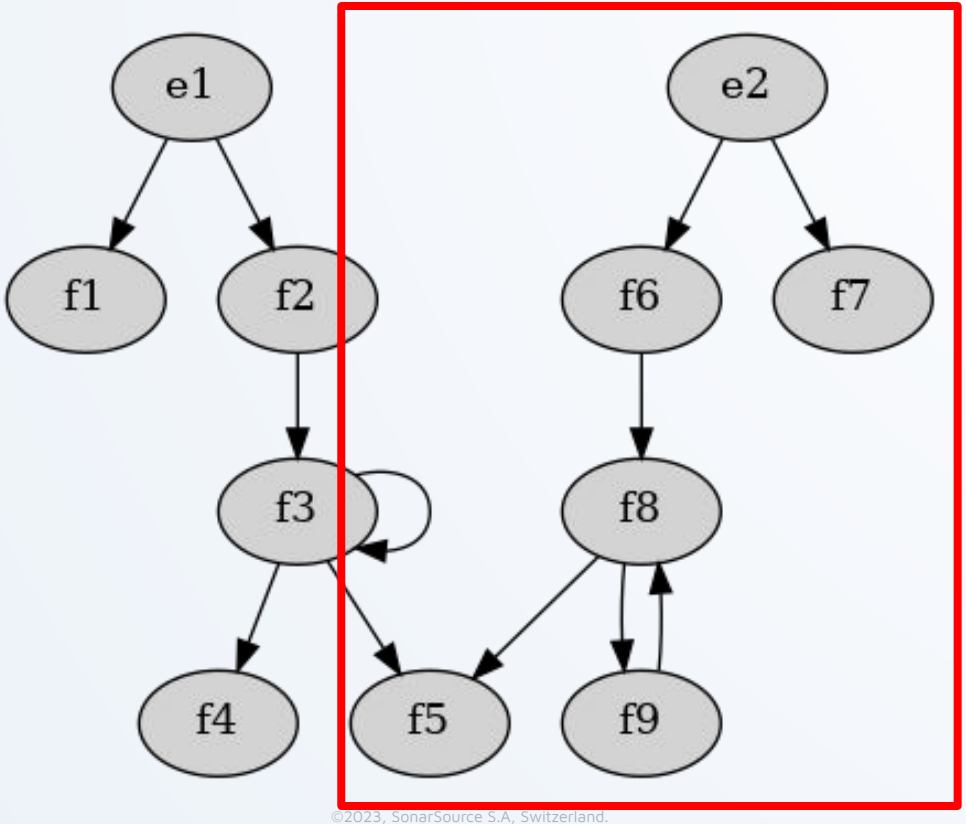
Call graph

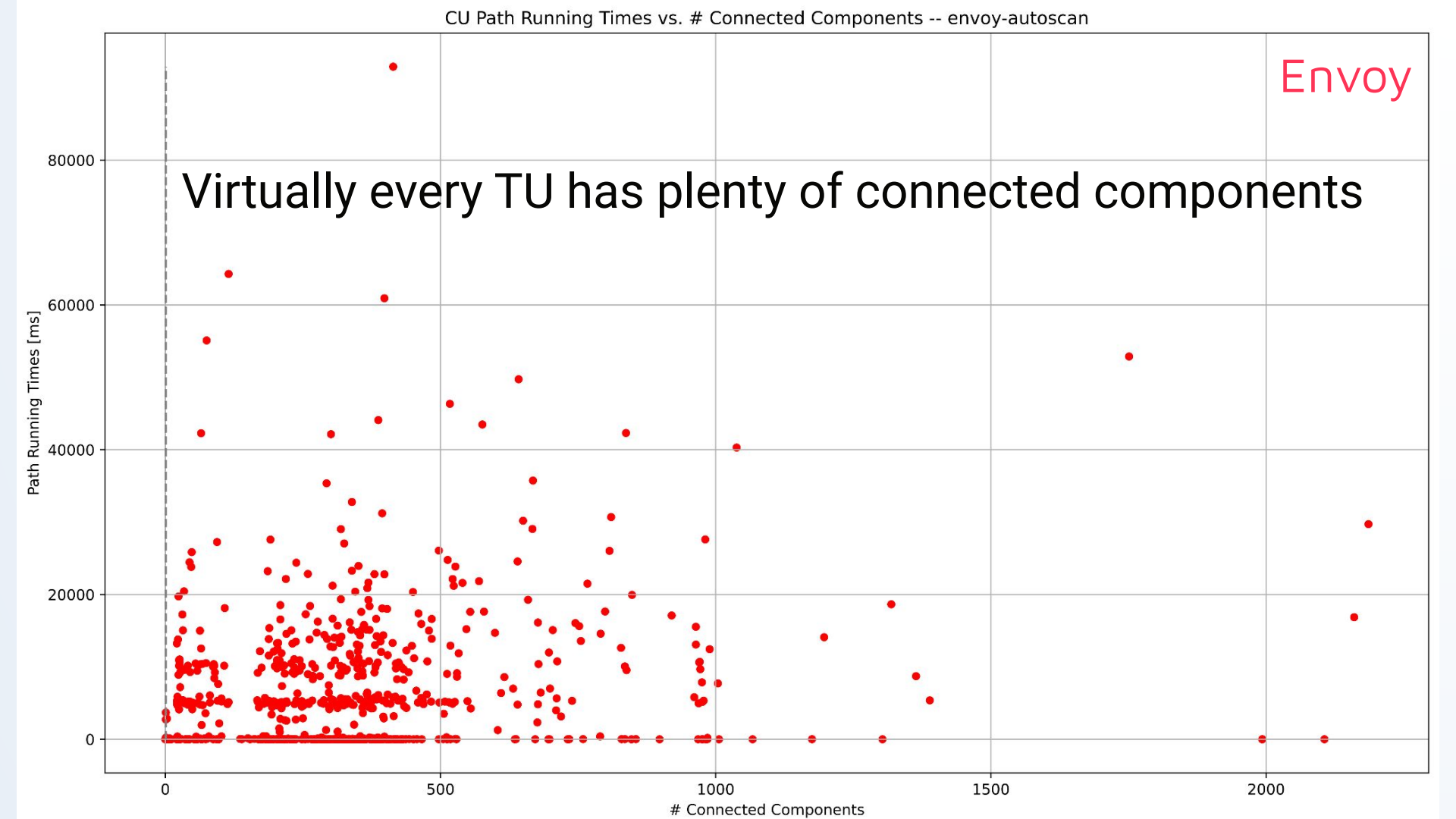


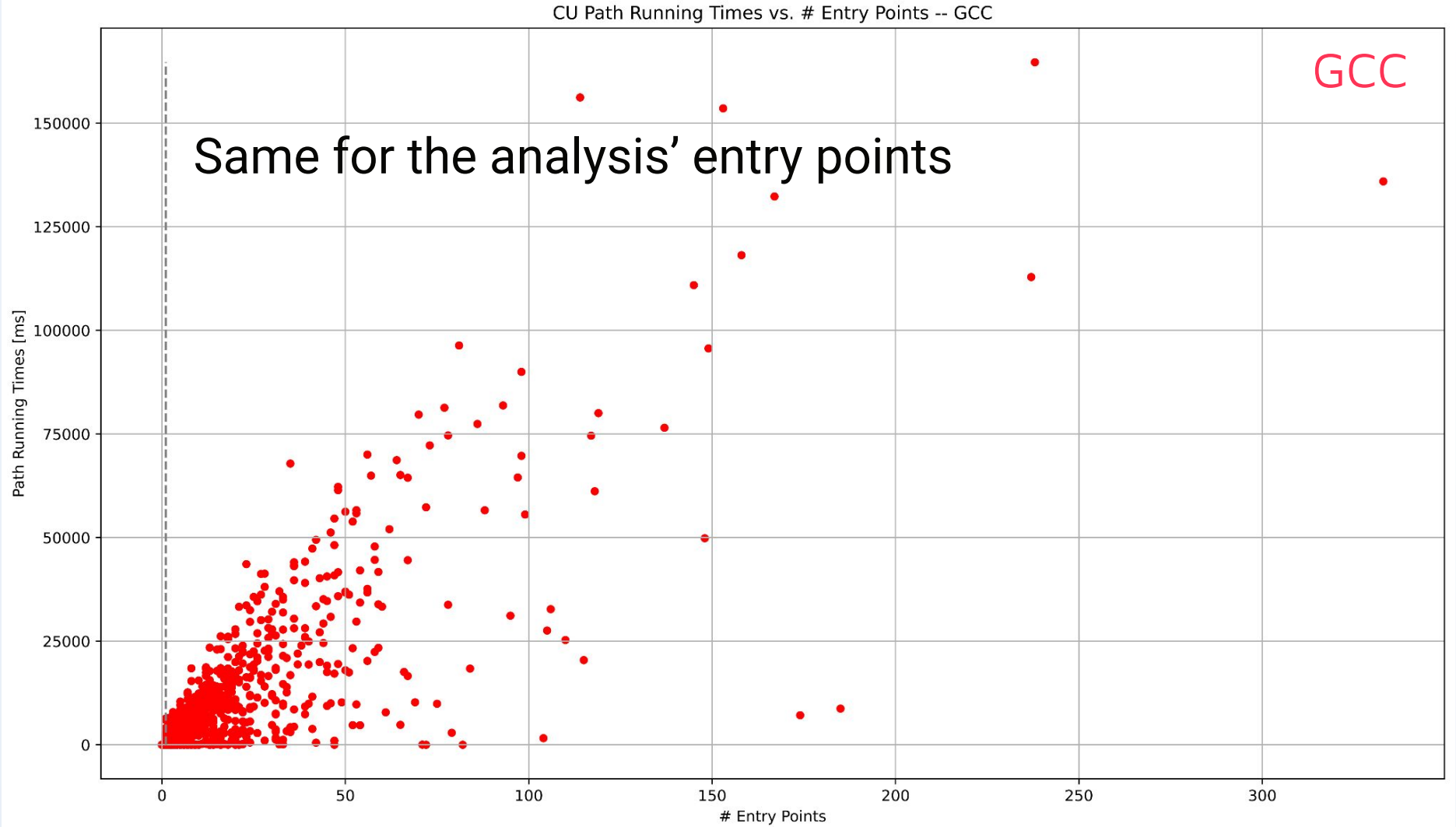
Call graph

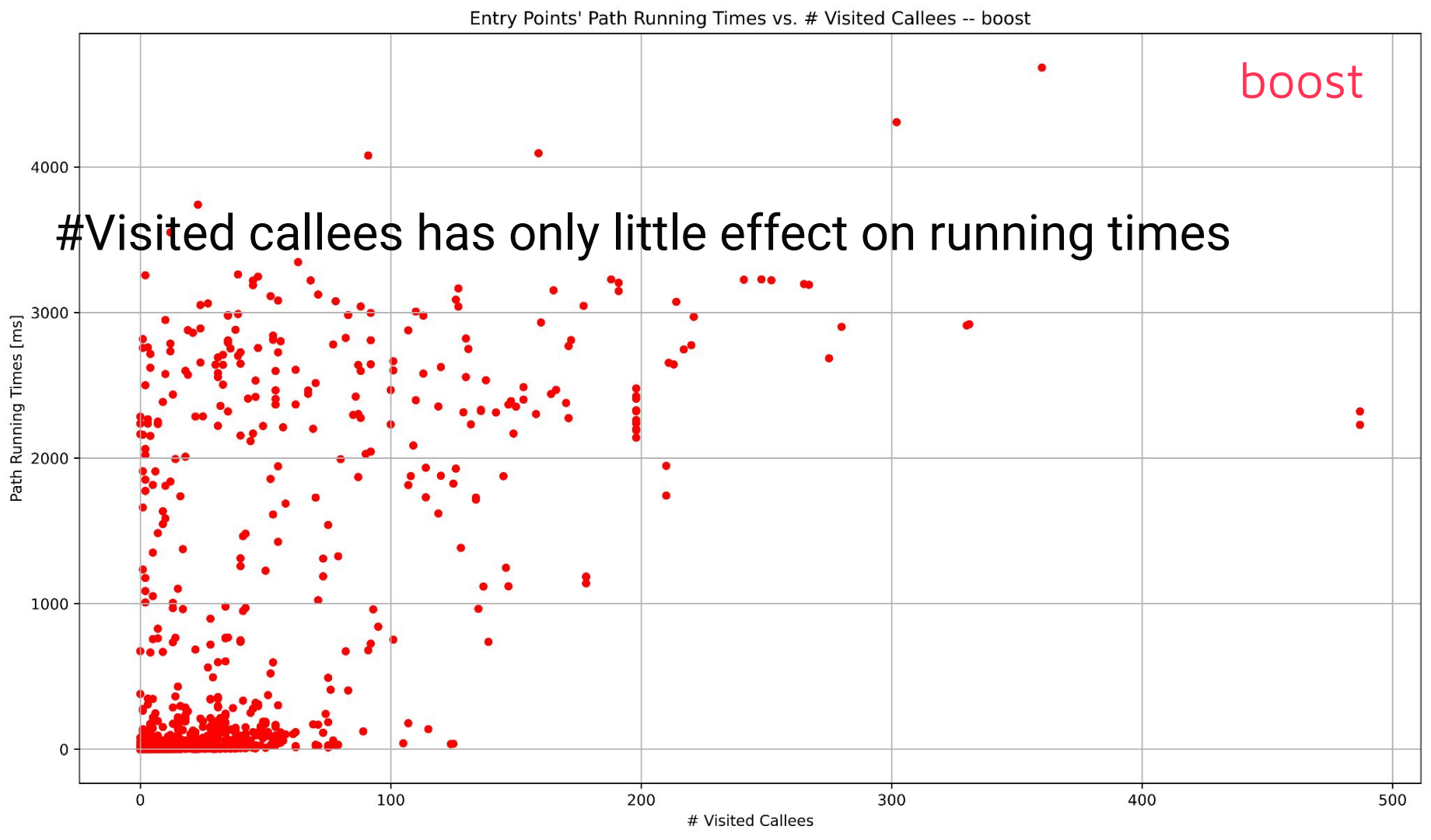


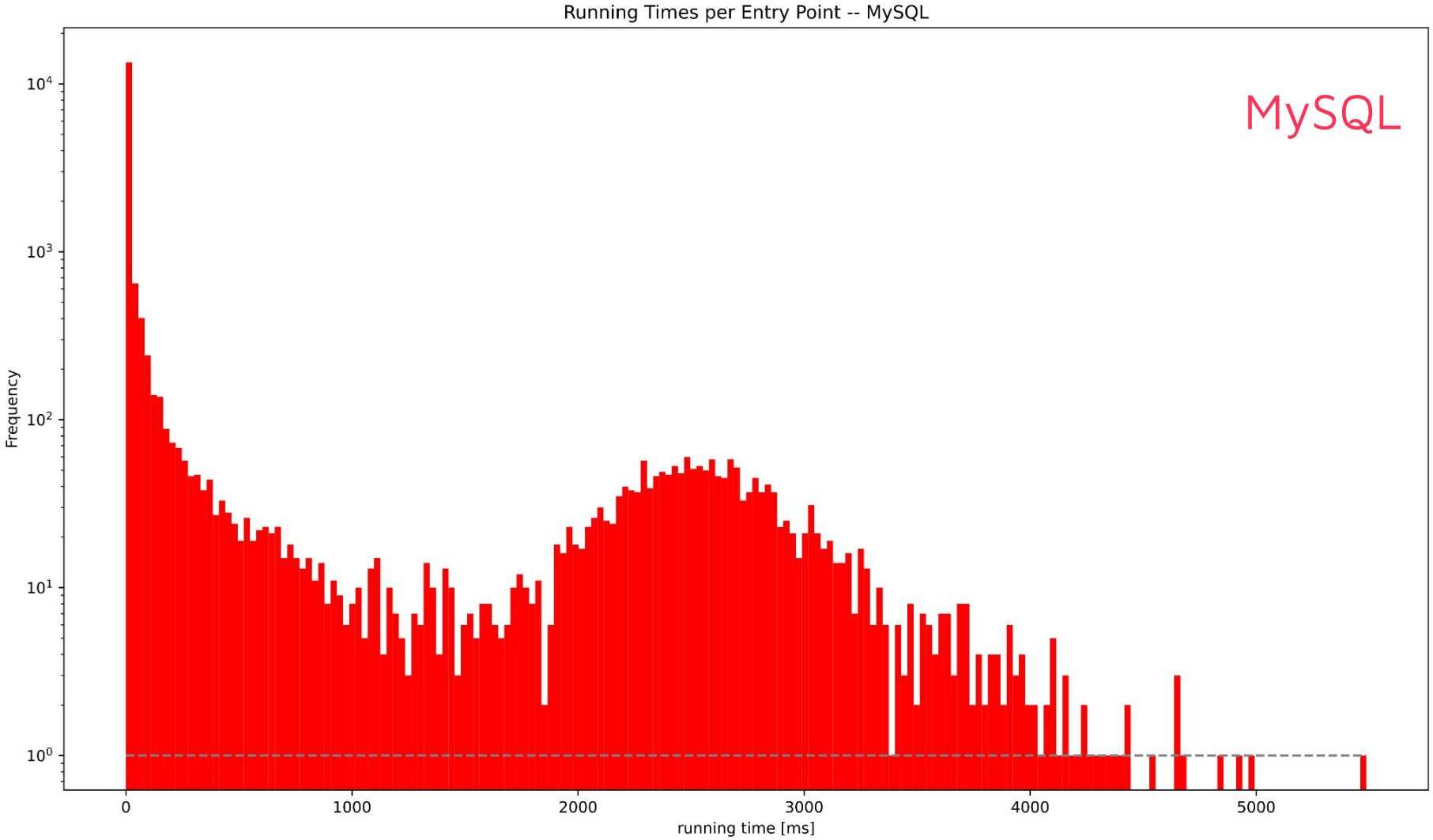
Call graph

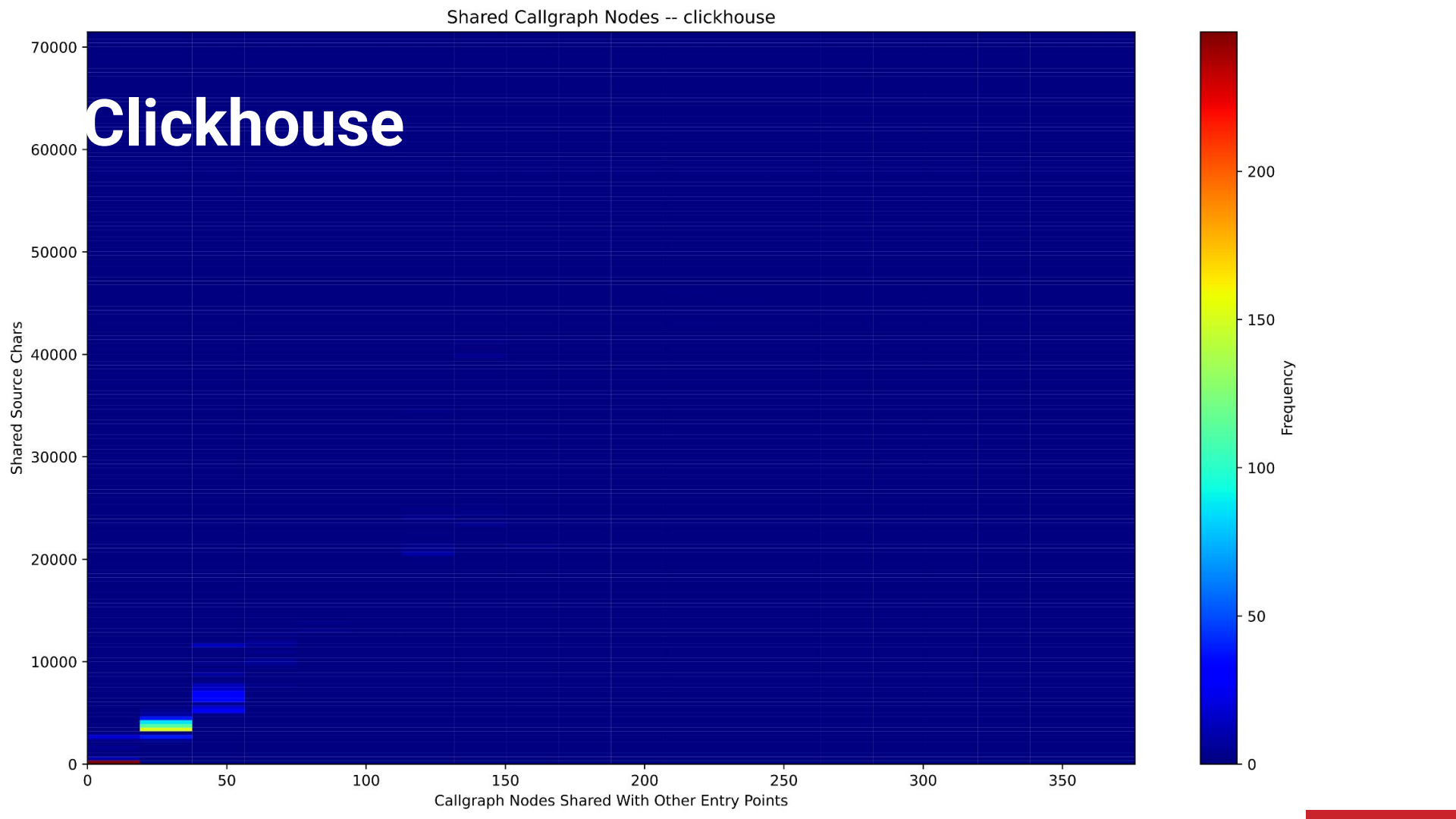


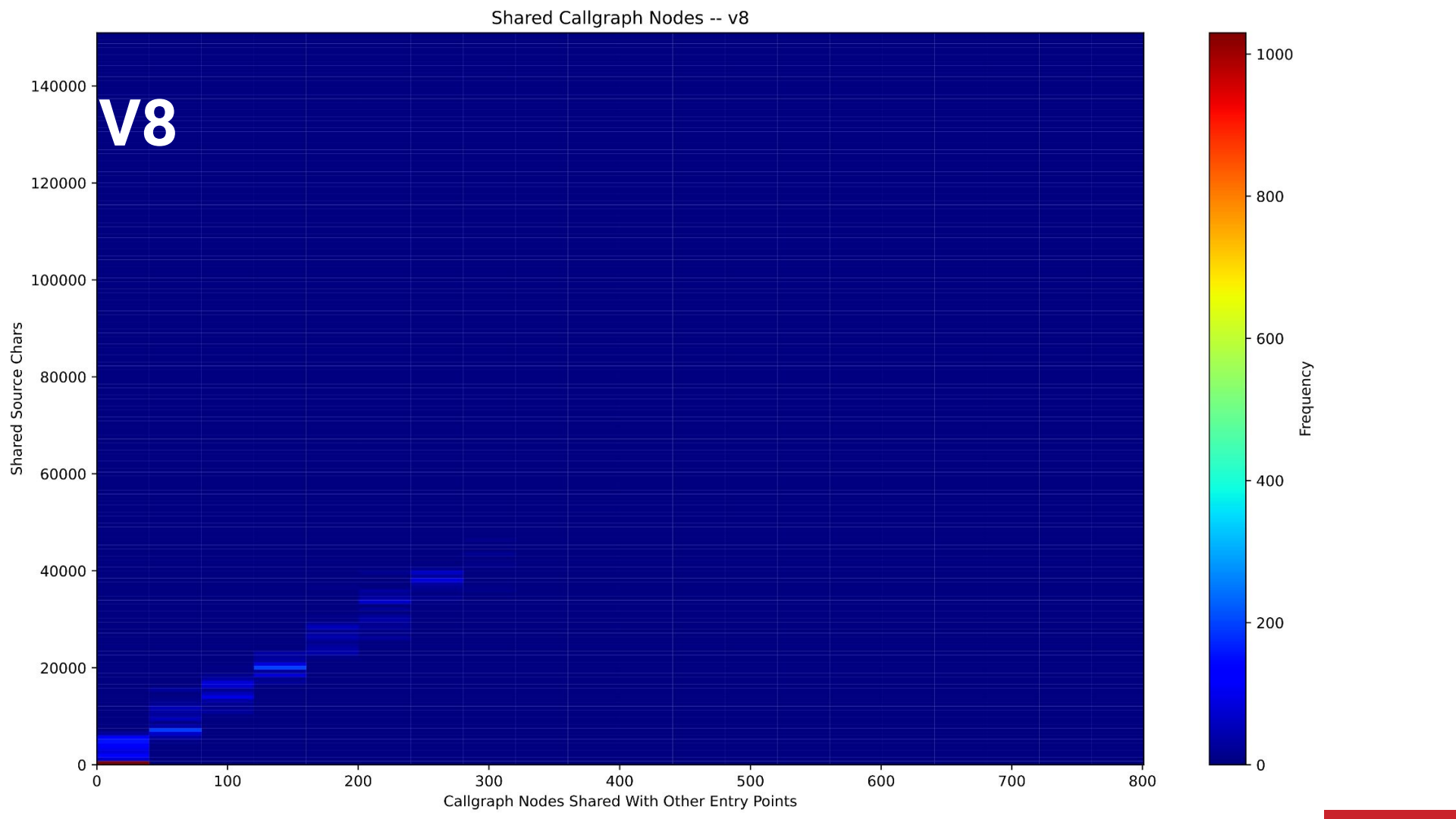

CU Path Running Times vs. # Connected Components -- envoy-autoscan
Envoy
Virtually every TU has plenty of connected components
Path Running Times [ms]
80000
60000
40000
20000
0
0
500
1000
1500
2000
Connected Components

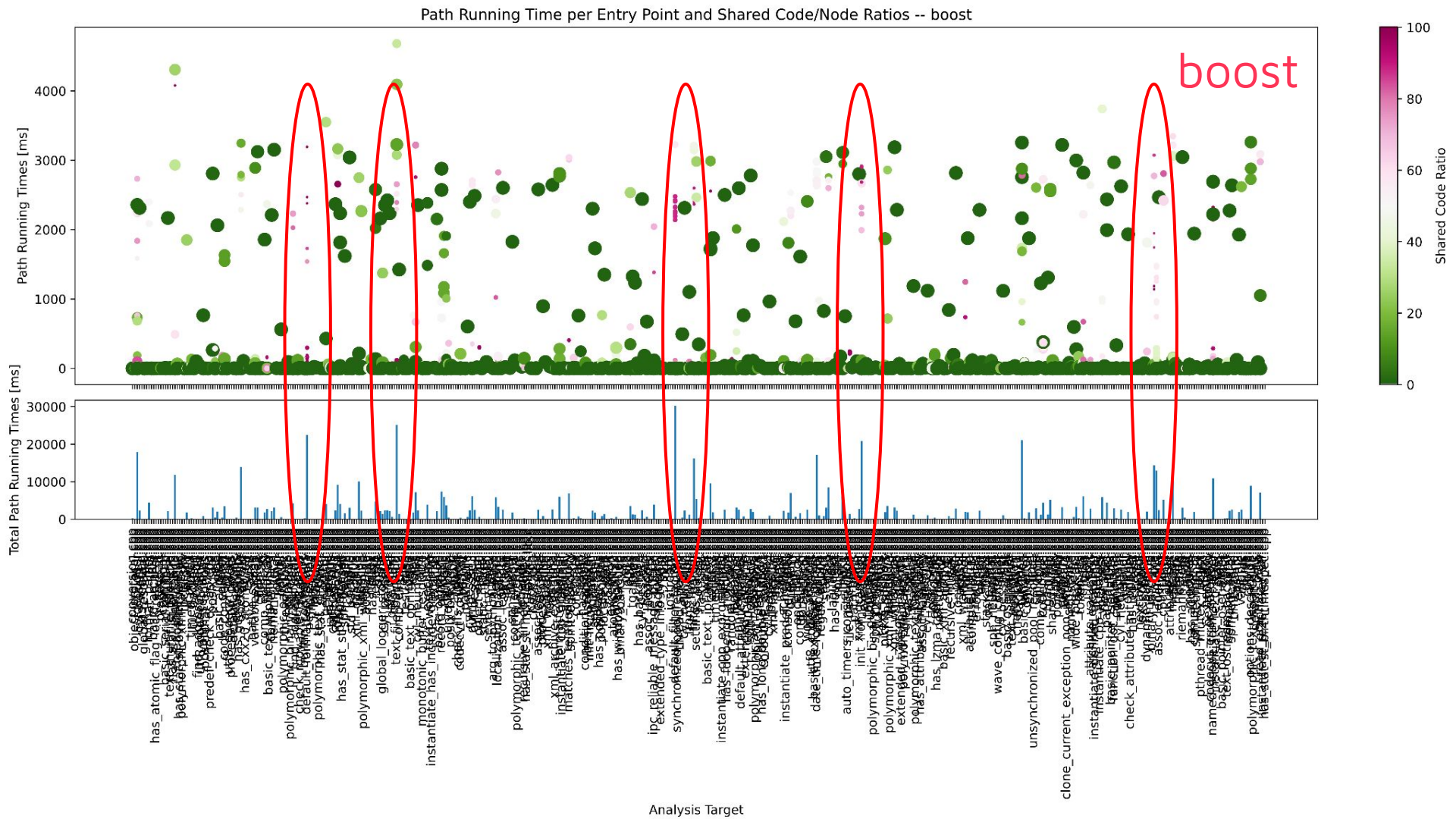

CU Path Running Times vs. # Entry Points -- GCC
GCC
Same for the analysis’ entry points
Path Running Times [ms]
Entry Points
0
25000
50000
75000
100000
125000
150000
0
50
100
150
200
250
300

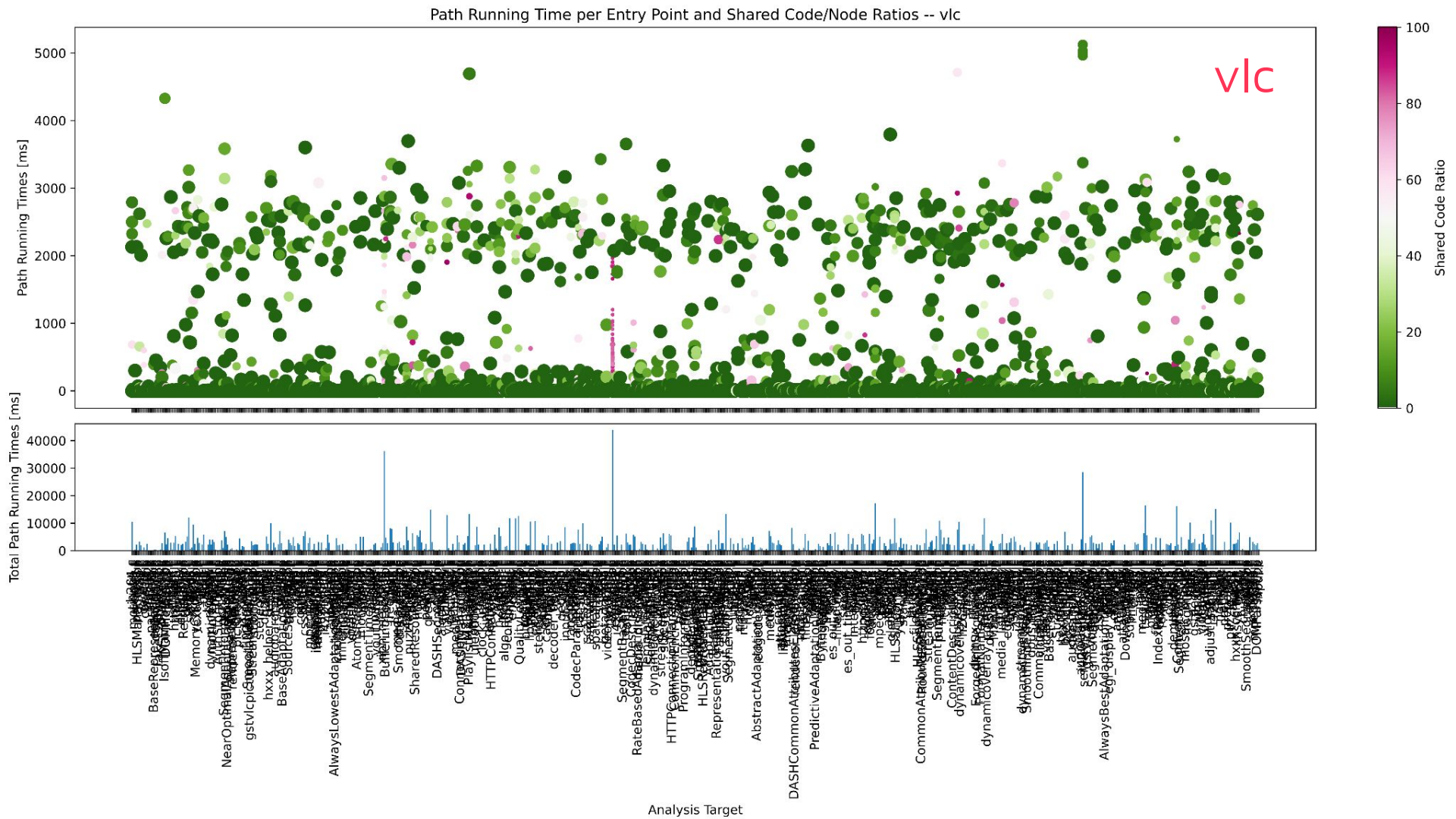

Entry Points' Path Running Times vs. # Visited Callees -- boost
boost
#Visited callees has only little effect on running times
Path Running Times [ms]
0
1000
2000
3000
4000
0
100
200
300
400
500
Visited Callees


Running Times per Entry Point -- MySQL
MySQL
Frequency
10^4
10^3
10^2
10^1
10^0
0
1000
2000
3000
4000
5000
running time [ms]

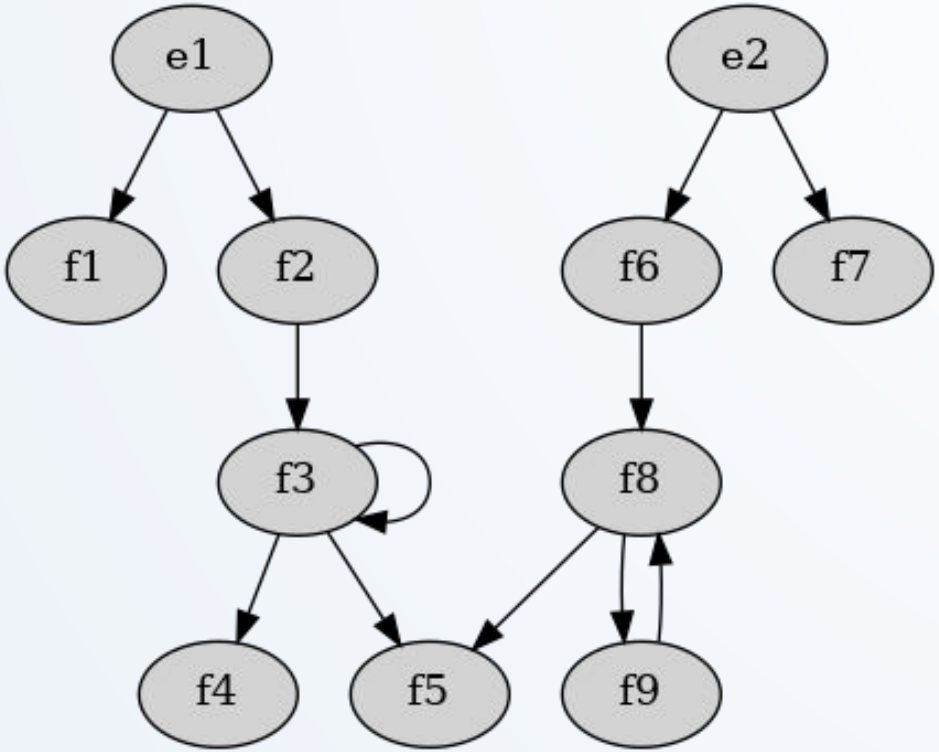

Shared Callgraph Nodes -- clickhouse
Clickhouse
Shared Source Chars
70000
60000
50000
40000
30000
20000
10000
0
0
50
100
150
200
250
300
350
Callgraph Nodes Shared With Other Entry Points
Frequency
200
150
100
50
0


Shared Callgraph Nodes -- v8
V8
Shared Source Chars
0
20000
40000
60000
80000
100000
120000
140000
Callgraph Nodes Shared With Other Entry Points
0
100
200
300
400
500
600
700
800
Frequency
0
200
400
600
800
1000


Path Running Time per Entry Point and Shared Code/Node Ratios -- boost
boost
Path Running Times [ms]
Total Path Running Times [ms]
Shared Code Ratio
Analysis Target

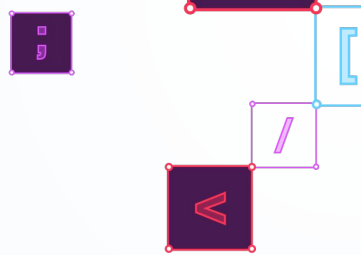

Path Running Time per Entry Point and Shared Code/Node Ratios -- vlc
vlc
Path Running Times [ms]
Shared Code Ratio
Total Path Running Times [ms]
Analysis Target

Call graph



CSA pipeline

- Pre-analysis
- Worklist
- Diagnostic construction
- Diagnostic consumption



Guiding principles

- Minimal
- Self-contained
- Significant speedup for the “usual” cases
- Fail early

Prototype architecture

- Analysis cache
- Oracle
- Report replayer
- Report recorder

Relocatable diagnostics

- Anchor decl USR
- AST index sequence
- Getter function
- Message

```
int someFunc(int x) {
    int result = x / 42;
    return result;
}
```

```
`-FunctionDecl 0x7fecf7067830 <test.c:2:1, line:5:1> line:2:5 someFunc 'int (int)'
  |-ParmVarDecl 0x7fecf7017538 <col:14, col:18> col:18 used x 'int'
  `-CompoundStmt 0x7fecf7067a80 <col:21, line:5:1>
    |-DeclStmt 0x7fecf7067a10 <line:3:3, col:22>
    | `-VarDecl 0x7fecf7067928 <col:3, col:20> col:7 used result 'int' cinit
    |   `-BinaryOperator 0x7fecf70679e8 <col:16, col:20> 'int' '/'
    |     |-ImplicitCastExpr 0x7fecf70679d0 <col:16> 'int' <LValueToRValue>
    |     | `-DeclRefExpr 0x7fecf7067988 <col:16> 'int' lvalue ParmVar 0x7fecf701753
    |     `-IntegerLiteral 0x7fecf70679b0 <col:20> 'int' 42
    `-ReturnStmt 0x7fecf7067a68 <line:4:3, col:10>
      `-ImplicitCastExpr 0x7fecf7067a50 <col:10> 'int' <LValueToRValue>
        `-DeclRefExpr 0x7fecf7067a28 <col:10> 'int' lvalue Var 0x7fecf7067928 'resul
```


sonar

Results

Preliminary Local Evaluation

Default Analysis

```
inflate.c,2159
---
analysisTarget: 'inflate.c'
config: 2
parsing: 48
annotation: 2
astRules: 20
symbolicExecution: 1984
total: 2056
...
```

Incremental Analysis

```
inflate.c,211
---
analysisTarget: 'inflate.c'
config: 1
parsing: 62
annotation: 3
astRules: 30
symbolicExecution: 42
total: 138
...
```

Runtime Overall (ii)

Runtime Distribution (iii)

Commit Data (i)

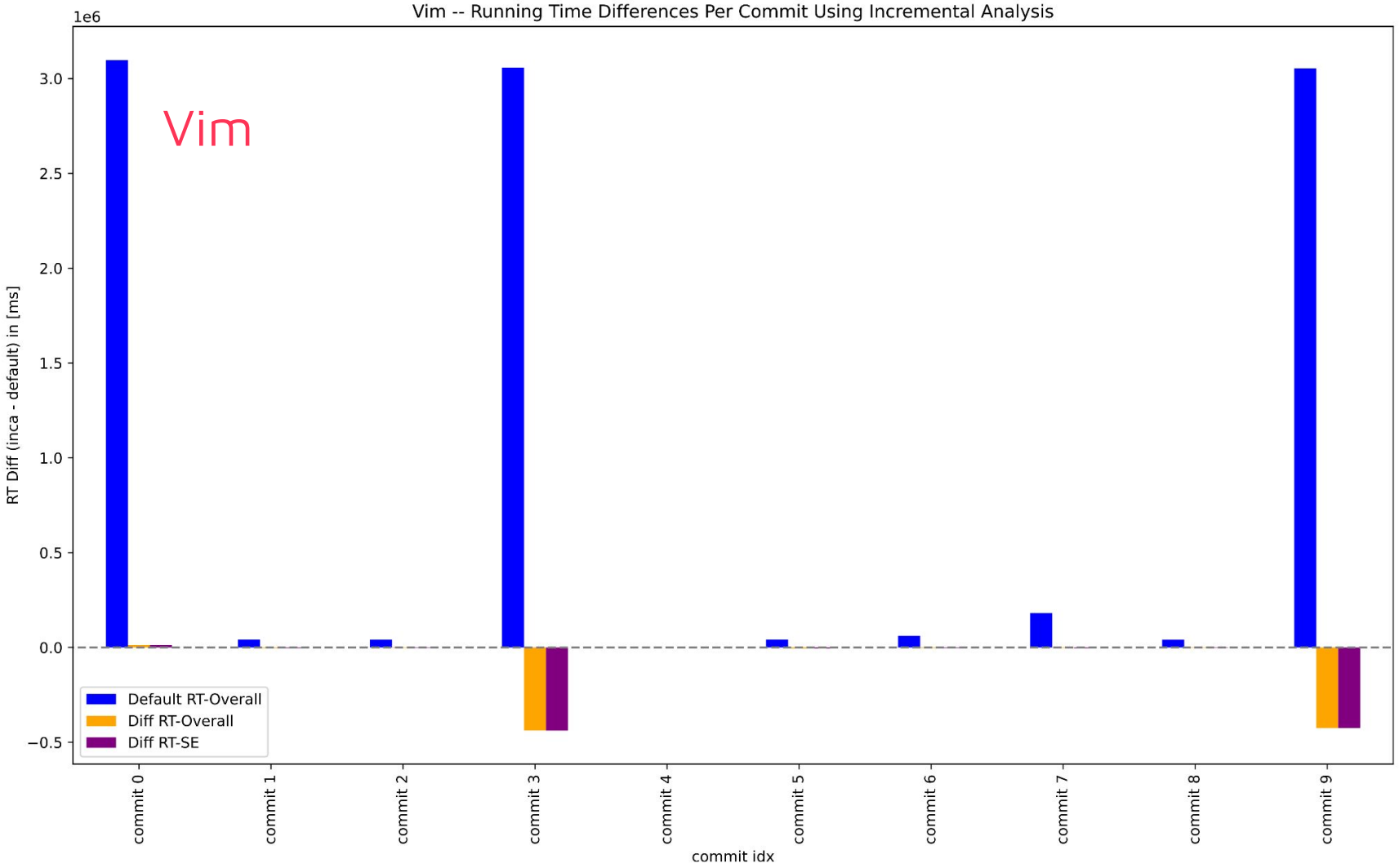
```
Refs: v1.12-30-g00f4ade
Author: Jim Meyering <meyering@fb.com>
AuthorDate: Wed Dec 28 08:36:27 2022 -0800
Commit: Jim Meyering <meyering@fb.com>
CommitDate: Wed Dec 28 08:36:27 2022 -0800

maint: SPC-indent inflate.c

* inflate.c: Indent with spaces, not TABs.
---
inflate.c | 16 ++++++++--------
1 file changed, 8 insertions(+), 8 deletions(-)
```

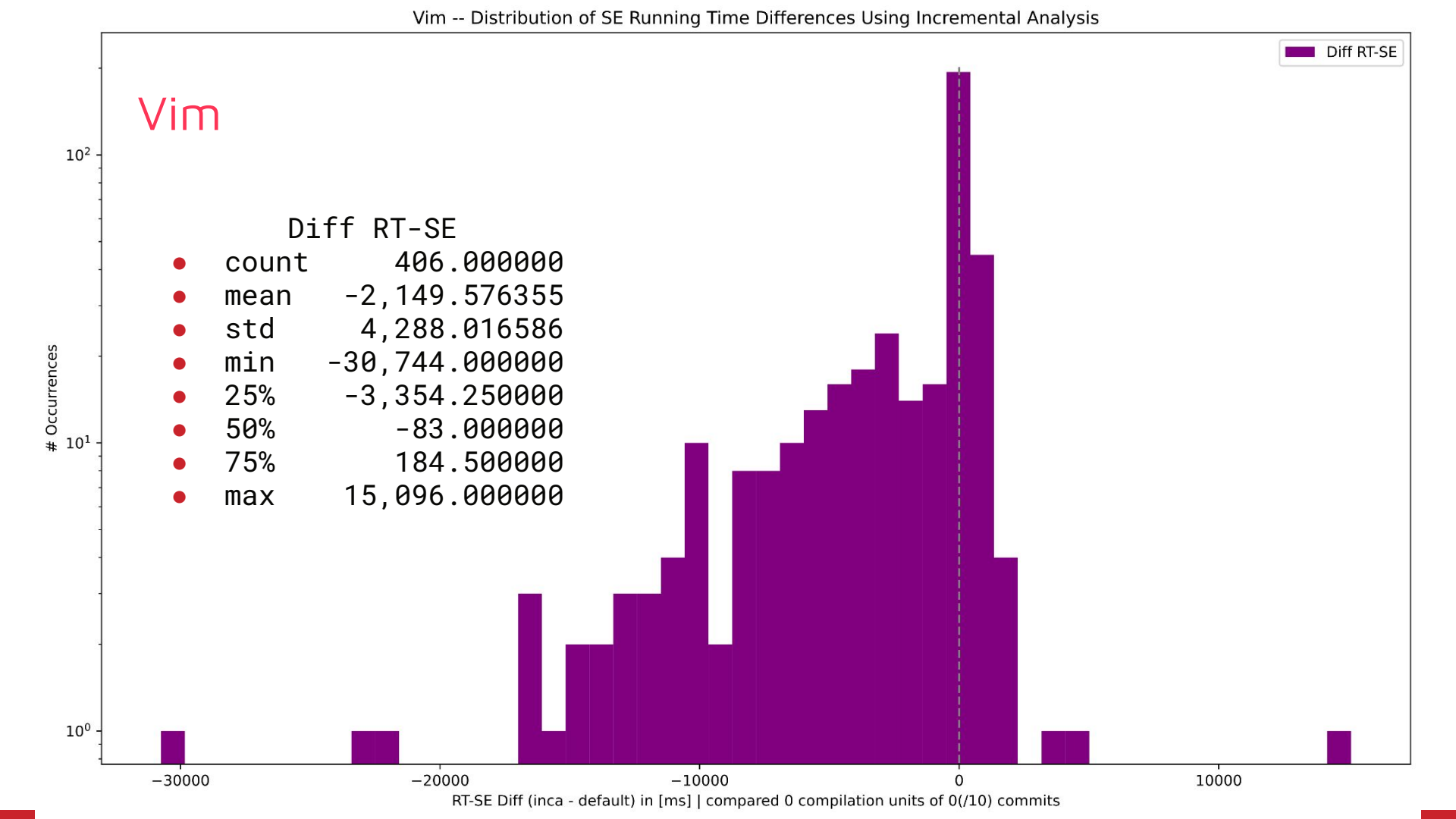
Analyzer Stats (iv)

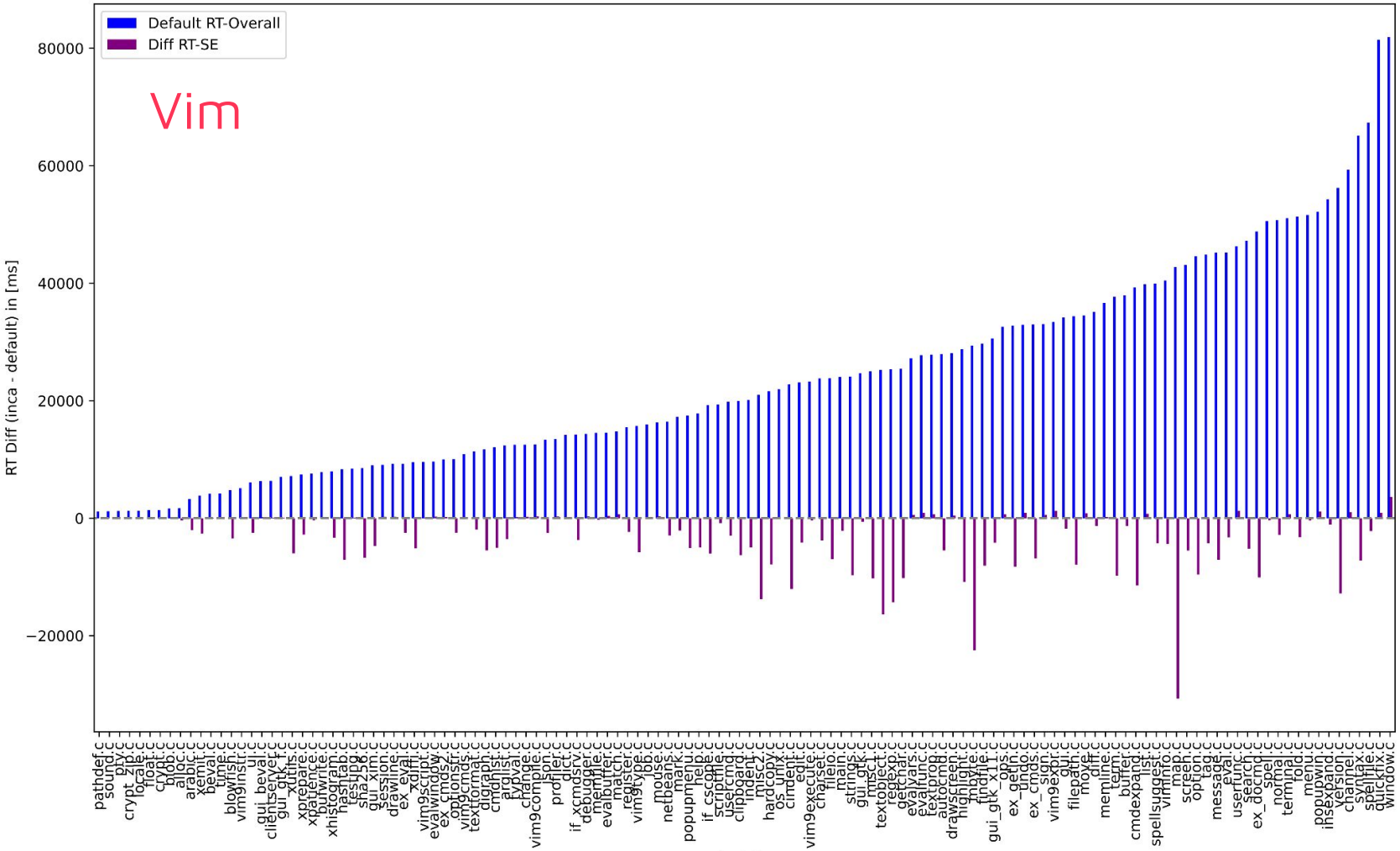

```
"AnalysisConsumer.NumFunctionTopLevel" : 14,
"AnalysisConsumer.NumFunctionsAnalyzed" : 7,
"time.analyzer.exprengine.wall" : 1.9754426479339600e+00 ,
"ConnectedCallGraphComponents" :
{ "Components" : 14, "TotalSourceChars" : 18770, "NumEntryPoints" : 7, "EntryPoints" : [
{ "Function" : "huft free" , "Location" : "inflate.c:498:1" , "PathRunningTime" : 0,
"CacheRunningTime" : 0, "CacheHit" : 0, "SourceChars" : 255, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate" , "Location" : "inflate.c:958:1" , "PathRunningTime" : 925,
"CacheRunningTime" : 0, "CacheHit" : 0, "SourceChars" : 3667, "VisitedCallees" : 3,
"SharedSourceChars" : 0, "SharedBy" : [{ "SharedByOther" : 1, "Freq" : 3 }] },
{ "Function" : "inflate block" , "Location" : "inflate.c:911:12" , "PathRunningTime" : 1,
"CacheRunningTime" : 0, "CacheHit" : 0, "SourceChars" : 712, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate codes" , "Location" : "inflate.c:520:1" , "PathRunningTime" : 893,
"CacheRunningTime" : 0, "CacheHit" : 0, "SourceChars" : 2933, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate dynamic" , "Location" : "inflate.c:735:1" , "PathRunningTime" :
61, "CacheRunningTime" : 0, "CacheHit" : 0, "SourceChars" : 4188, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate_fixed" , "Location" : "inflate.c:685:1" , "PathRunningTime" : 0,
[...]
```

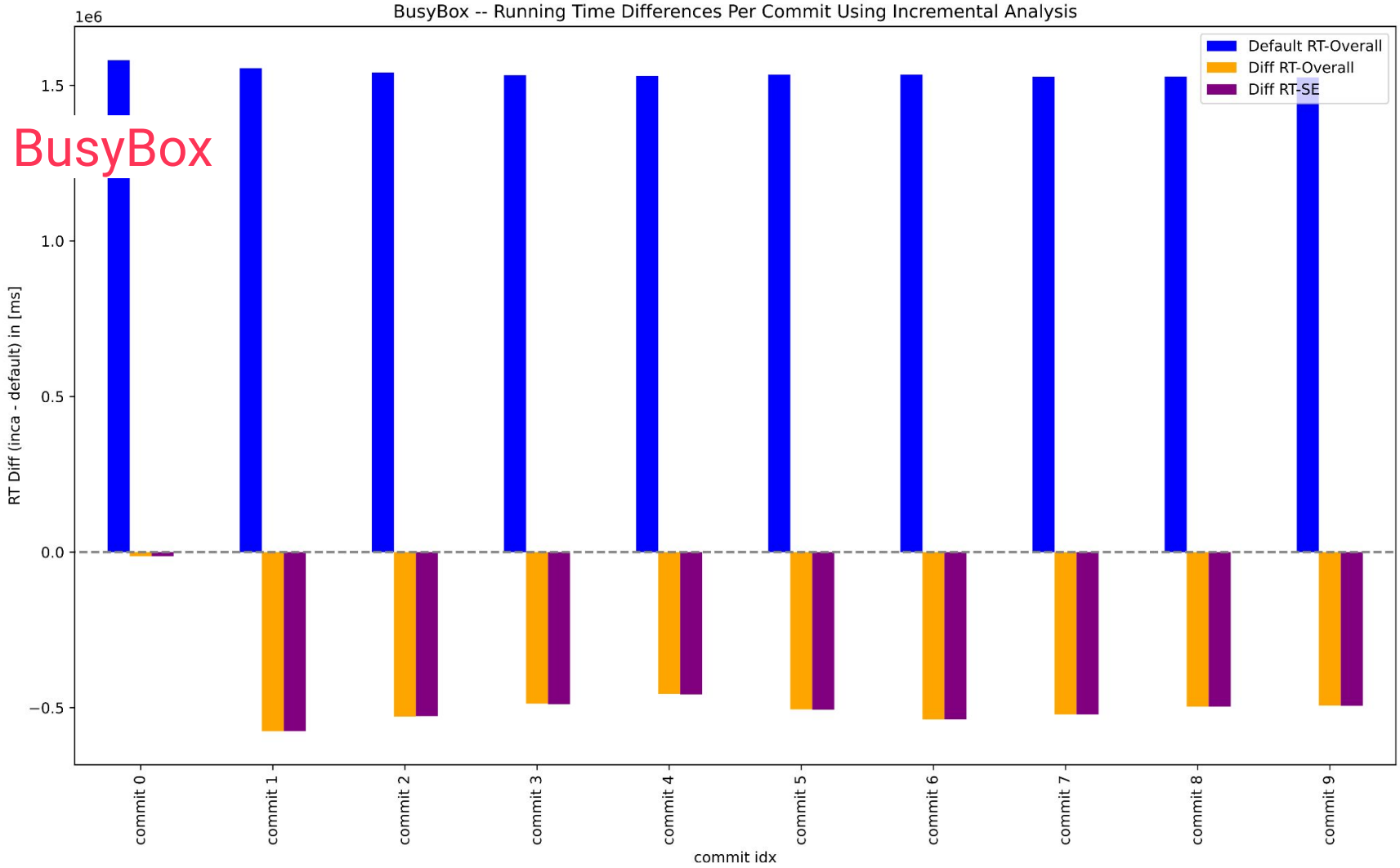


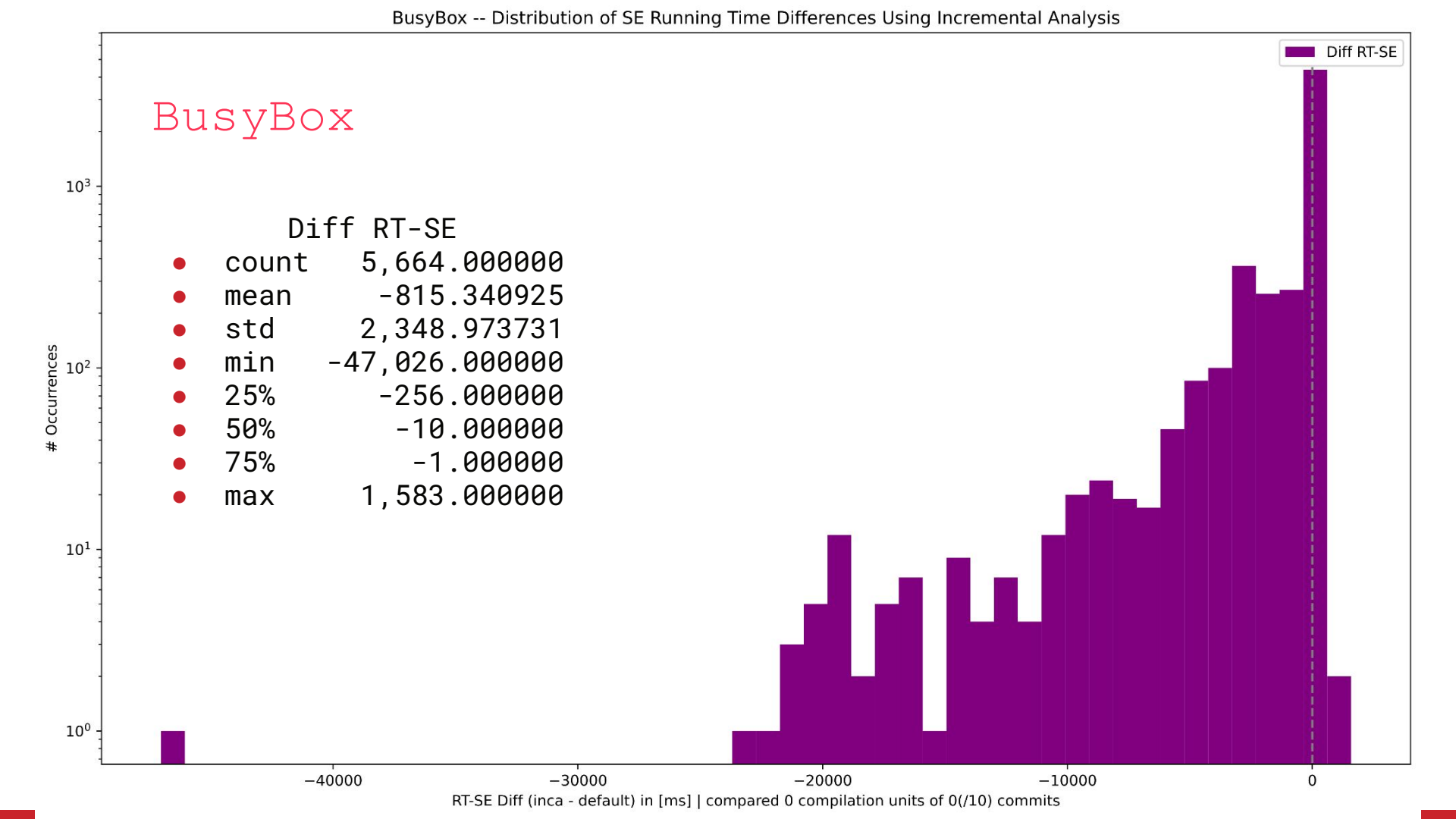
```
"AnalysisConsumer.NumFunctionTopLevel" : 14,
"AnalysisConsumer.NumFunctionsAnalyzed" : 7,
"AnalysisConsumer.NumIncrementalAnalysisCacheHits" : 7,
"time.analyzer.cachedecision.wall" : 3.8180351257324219e-02 ,
"ConnectedCallGraphComponents" :
{ "Components" : 14, "TotalSourceChars" : 18770, "NumEntryPoints" : 7, "EntryPoints" : [
{ "Function" : "huft free" , "Location" : "inflate.c:498:1" , "PathRunningTime" : 0,
"CacheRunningTime" : 0, "CacheHit" : 1, "SourceChars" : 255, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate" , "Location" : "inflate.c:958:1" , "PathRunningTime" : 0,
"CacheRunningTime" : 10, "CacheHit" : 1, "SourceChars" : 3667, "VisitedCallees" : 3,
"SharedSourceChars" : 0, "SharedBy" : [{ "SharedByOther" : 1, "Freq" : 3 }] },
{ "Function" : "inflate block" , "Location" : "inflate.c:911:12" , "PathRunningTime" : 0,
"CacheRunningTime" : 10, "CacheHit" : 1, "SourceChars" : 712, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate codes" , "Location" : "inflate.c:520:1" , "PathRunningTime" : 0,
"CacheRunningTime" : 3, "CacheHit" : 1, "SourceChars" : 2933, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate dynamic" , "Location" : "inflate.c:735:1" , "PathRunningTime" : 0,
"CacheRunningTime" : 9, "CacheHit" : 1, "SourceChars" : 4188, "VisitedCallees" : 0,
"SharedSourceChars" : 0, "SharedBy" : [] },
{ "Function" : "inflate_fixed" , "Location" : "inflate.c:685:1" , "PathRunningTime" : 0,
[...]
```

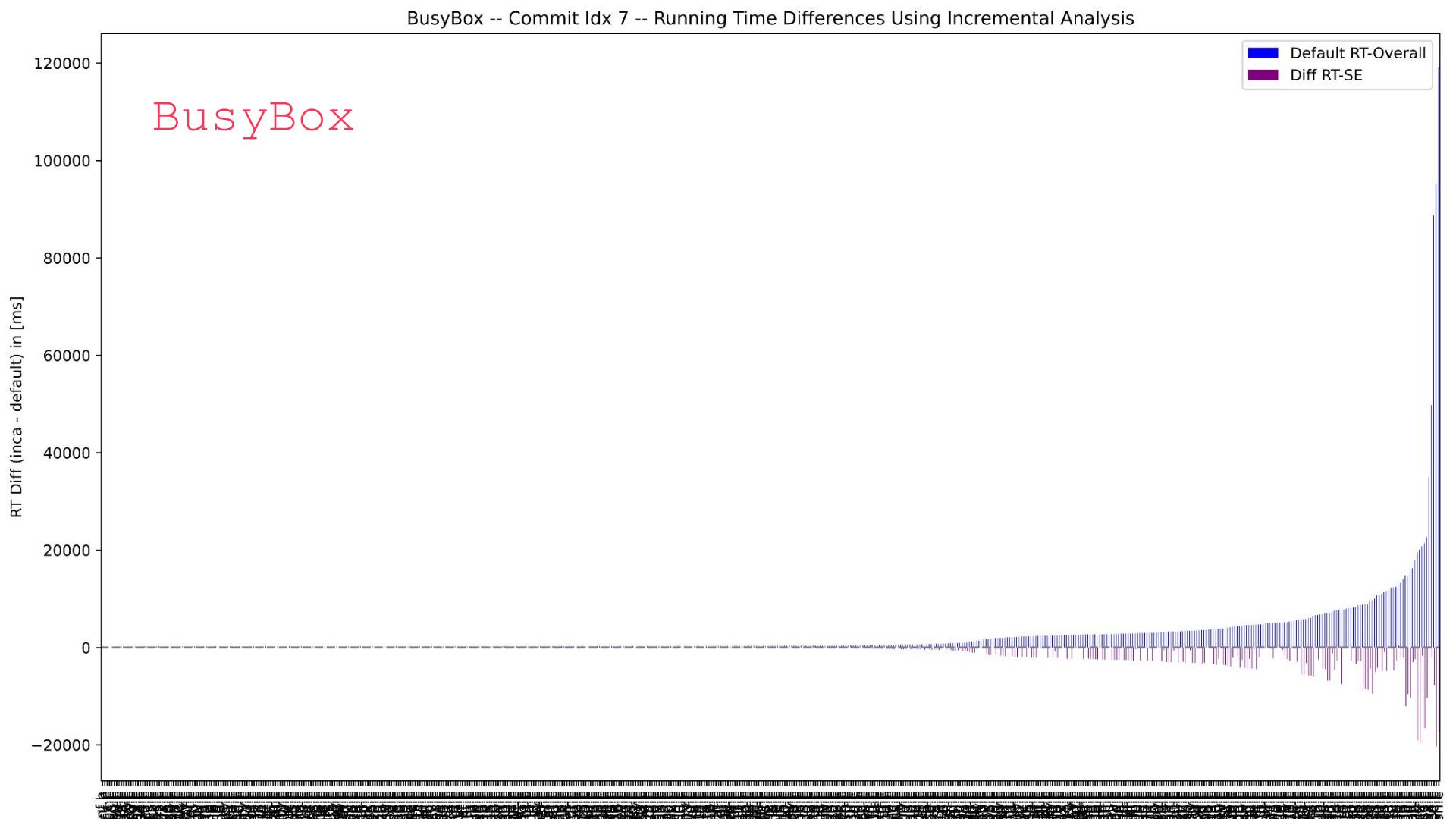

Vim -- Running Time Differences Per Commit Using Incremental Analysis
1e6
Vim
RT Diff (inca - default) in [ms]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Default RT-Overall
Diff RT-Overall
Diff RT-SE
commit 0
commit 1
commit 2
commit 3
commit 4
commit 5
commit 6
commit 7
commit 8
commit 9
commit idx


Vim -- Distribution of SE Running Time Differences Using Incremental Analysis
Diff RT-SE
Vim
Diff RT-SE
count 406.000000
mean -2,149.576355
std 4,288.016586
min -30,744.000000
25% -3,354.250000
50% -83.000000
75% 184.500000
max 15,096.000000
Occurrences
10^2
10^1
10^0
−30000
−20000
−10000
0
10000
RT-SE Diff (inca - default) in [ms] | compared 0 compilation units of 0(/10) commits


Vim -- Commit Idx 9 -- Running Time Differences Using Incremental Analysis
Default RT-Overall
Diff RT-SE
Vim
RT Diff (inca - default) in [ms]
80000
60000
40000
20000
0
−20000
analysisTarget


BusyBox -- Running Time Differences Per Commit Using Incremental Analysis
1e6
Default RT-Overall
Diff RT-Overall
Diff RT-SE
BusyBox
RT Diff (inca - default) in [ms]
1.5
1.0
0.5
0.0
−0.5
commit 0
commit 1
commit 2
commit 3
commit 4
commit 5
commit 6
commit 7
commit 8
commit 9
commit idx


BusyBox -- Distribution of SE Running Time Differences Using Incremental Analysis
Diff RT-SE
BusyBox
Diff RT-SE
count 5,664.000000
mean -815.340925
std 2,348.973731
min -47,026.000000
25% -256.000000
50% -10.000000
75% -1.000000
max 1,583.000000
Occurrences
10^3
10^2
10^1
10^0
−40000
−30000
−20000
−10000
0
RT-SE Diff (inca - default) in [ms] | compared 0 compilation units of 0(/10) commits


BusyBox -- Commit Idx 7 -- Running Time Differences Using Incremental Analysis
Default RT-Overall
Diff RT-SE
BusyBox
RT Diff (inca - default) in [ms]
120000
100000
80000
60000
40000
20000
0
−20000

Challenges

- Preprocessor and tokens
- Diagnostic relocation (abs line refs)
- Side effects (fn summaries, z3 refutation nondet)