Binary code browser

Student: Alin Mindroc (Romania)
Mentor: Dr. Sandro Wenzel

Main goals:

-Create two projects: web app and Eclipse plugin which could assist developers in the process of browsing/analyzing binary code

-Create an abstract layer so that the web app / Eclipse plugin (Java) can communicate to ParseApi (C++)

-Generate call graphs for executables

-Generate histograms for assembly instructions

-Provide a “diff” view so that you could easily compare two functions

-Use a source to source parser to easily generate JNI - ready C++ sources

-Generate a mapping view (C/C++ -> assembly)

Architecture

Input data
(executable files, object files, static libraries, shared objects)

```
400548:       0f 1f 84 00 00 00 00    nopl   0x0(%rax,%rax,1)
40054f:       00
400550:       4c 89 ea                mov    %r13,%rdx
400553:       4c 89 f6                mov    %r14,%rsi
400556:       44 89 ff                mov    %r15d,%edi
400559:       41 ff 14 dc             callq  *(%r12,%rbx,8)
40055d:       48 83 c3 01             add    $0x1,%rbx
400561:       48 39 eb                cmp    %rbp,%rbx
400564:       75 ea                   jne    400550 <__libc_csu_init+0x40>
400566:       48 83 c4 08             add    $0x8,%rsp
```

Parsing layer

Dyninst
C++
ParseAPI
InstructionAPI
SymtabAPI

JNI Wrapper
Java

Dino Web app

Backend (Java)

Apache Tomcat

Parser cache

Frontend (AngularJS)

- Histogram view
- Diff view
- Assembly listing
- Function browsing
- Sorting

Dino Eclipse plugin
(Java, Swing)

Parser cache

Eclipse
Views

Diff View:

- Lists Functions in executables
- Shows assembly for a function
- Generates diffs between any two functions in two executables

Source View
Maps a line of source code to a sequence of assembly addresses and vice versa

Dino Webapp:

Interactive web app which lets the user upload executable files and list functions, assembly code, generates histograms and diff views between different functions' assembly.

The input files can be categorized as:

1. Executable files, shared objects (**.so**) : big list of (address -> instruction) mapping, with some addresses labeled as functions
2. Static archives (**.a**) : contain more object files (**.o**) which contain address -> instruction mappings, so function names are not unique in a static archive, one function is also identified by the object file where it is defined

Function lists can be sorted by name / address / size + object name for static archive files, can be searched.

Why *"Dino"* **:** Dyninst (**Dyn**amic **Inst**rumentation) -> Dyno -> *Dino*

Demo time!

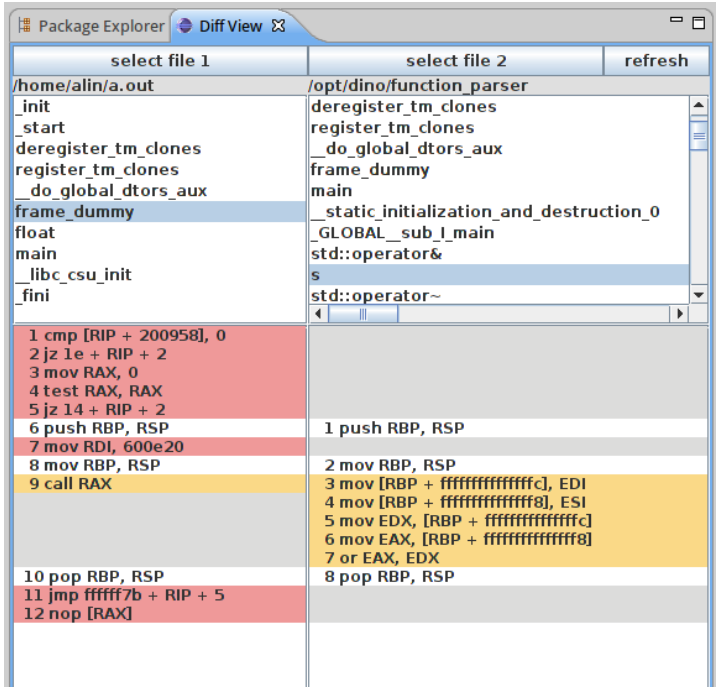
http://gsoc1.cern.ch:8080/dino

Dino plugin:

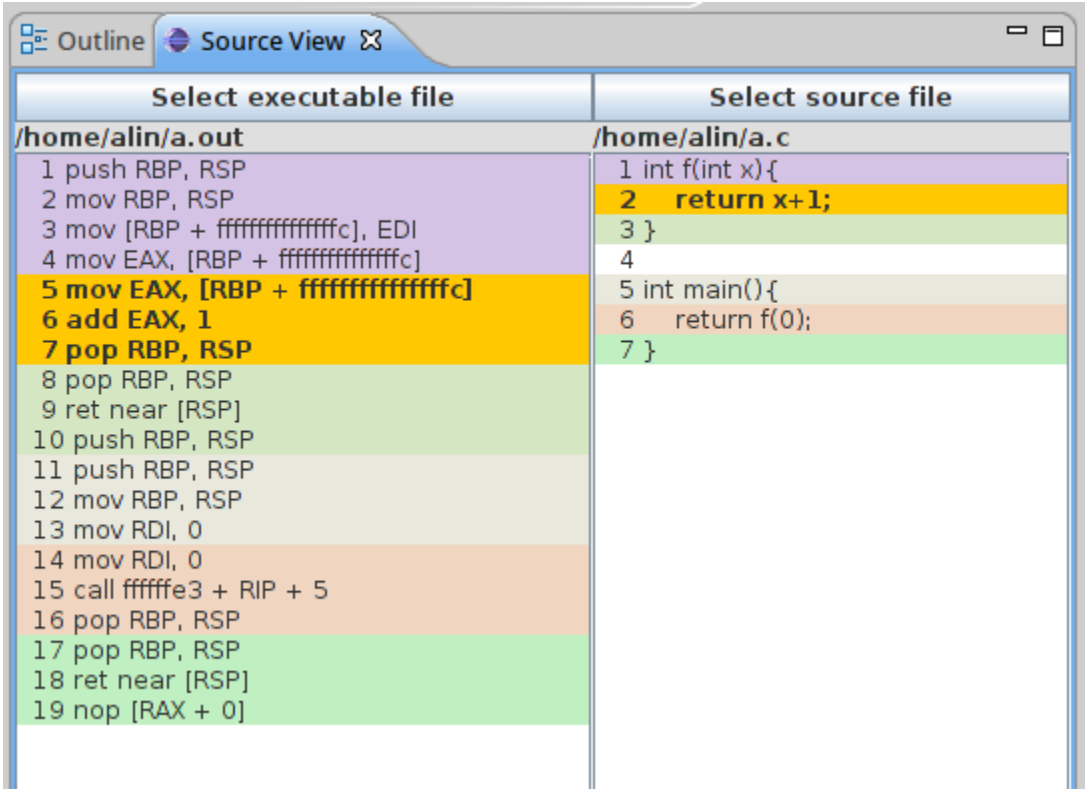
Eclipse plugin which implements some of the web app's functionality in the Eclipse IDE.

It contains two views:

1. **Diff view**: offers the possibility to get a diff view between two function's assembly code, it can also be used to browse the contents of an executable file



2. **Source view**: offers a mapping between assembly and source code for an executable file:



Overall, this project proved to be more of a “software engineering” one, requiring:

- planning on what technologies to use
- learning how to use a tool only from its documentation and the support from its little community - Dyninst framework
- time management between working on the web-app and the plugin
- having to abandon some of the initial goals, based on how the project evolved and on the Dyninst framework limitations (call graph)
- adding new functionalities which were not discussed initially (source to assembly mapping)

ASM Parser

ASM Upload

Binaries

First Previous 1 Next Last

search **8** files matched

sublime-text

three-functions

libscheduler.so

TorusBenchmark

TorusBenchmark2Mac

libvecgeom.a

test

assembly

Functions in **TorusBenchmark**

search **1044** functions matched

Sort by address ascending

4261b0	vecgeom::cxx::VUnplacedVolume::~VUnplacedVolume (14 bytes)
4261c0	vecgeom::cxx::Vector3D<double>::~Vector3D (14 bytes)
4261d0	vecgeom::cxx::VUnplacedVolume::~VUnplacedVolume (14 bytes)
4261f0	vecgeom::cxx::AlignedBase::~AlignedBase (14 bytes)
426210	vecgeom::cxx::Vector3D<double>::~Vector3D (14 bytes)
426230	**vecgeom::cxx::UnplacedTube::UnplacedTube (491 bytes)**
426450	vecgeom::cxx::GeoManager::Instance (196 bytes)
426520	main (2091 bytes)
426de0	_GLOBAL__sub_I_TorusBenchmark.cpp (80 bytes)
426e40	std::_Rb_tree<int, std::pair<int const, vecgeom::cxx::VPlacedVolume*>, std::_Select1st<std::pair<int const, vecgeom::cxx::VPlacedVolume*> >, std::less<int>, std::allocator<std::pair<int const, vecgeom::cxx::VPlacedVolume*> > >::_M_erase (448 bytes)

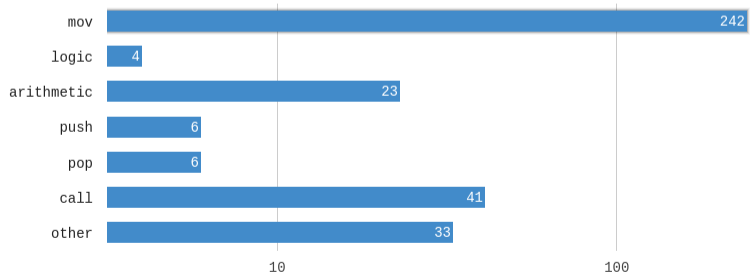
main Histogram

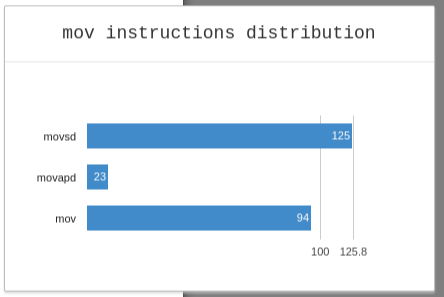
```
42694a : mov RDI, RBP
42694d : movsd XMM2, [RSP + 30]
426953 : movsd XMM3, [RSP + 38]
426959 : movsd [RSP + 458], XMM0
426962 : movapd XMM7, XMM2
426966 : movsd XMM0, [RSP + 48]
42696c : addsd XMM7, XMM3
426970 : movsd XMM1, [RSP + 28]
426976 : movsd [RSP + 460], XMM0
42697f : movapd XMM0, XMM3
426983 : movsd XMM4, [RSP + 8]
426989 : subsd XMM0, XMM2
42698d : movsd XMM3, [RSP + 10]
426993 : subsd XMM0, XMM1
426997 : addsd XMM1, XMM7
42699b : call vecgeom::cxx::UnplacedTube::UnplacedTube
4269a0 : movsd XMM0, [RSP + 190]
4269a9 : lea RDX, RSP + b0
4269b1 : lea RSI, RIP + 6ddc4
4269b8 : mov RDI, R13
4269bb : movsd [RSP + 4b0], XMM0
4269c4 : movsd XMM0, [RSP + 198]
4269cd : movsd [RSP + 4b8], XMM0
4269d6 : movsd XMM0, [RSP + 1a0]
4269df : movsd [RSP + 4c0], XMM0
4269e8 : movsd XMM0, [RSP + 1a8]
4269f1 : movsd [RSP + 4c8], XMM0
4269fa : movsd XMM0, [RSP + 1b0]
426a03 : movsd [RSP + 4d0], XMM0
426a0c : movsd XMM0, [RSP + 1b8]
426a15 : movsd [RSP + 4d8], XMM0
426a1e : movsd XMM0, [RSP + 1c0]
426a27 : movsd [RSP + 4e0], XMM0
426a30 : movsd XMM0, [RSP + 1c8]
426a39 : movsd [RSP + 4e8], XMM0
426a42 : movsd XMM0, [RSP + 1d0]
426a4b : movsd [RSP + 4f0], XMM0
426a54 : movsd XMM0, [RSP + 1d8]
426a5d : movsd [RSP + 4f8], XMM0
426a66 : movsd XMM0, [RSP + 1e0]
426a6f : movsd [RSP + 500], XMM0
```

vecgeom::cxx::UnplacedTube::UnplacedTube Histogram

```
426230 : push R12, RSP
426232 : push RBP, RSP
426233 : push RBX, RSP
426234 : mov RBX, RDI
426237 : lea RDI, RDI + d0
42623e : sub RSP, 20
426242 : mov RAX, [RIP + 7b2497]
426249 : movsd [RDI + fffff50], XMM0
426251 : movsd [RDI + fffff58], XMM1
426259 : mov [RDI + fffff78], 0
426264 : mov [RDI + ffffffffffff80], 0
42626c : movsd [RDI + fffff60], XMM2
426274 : mov [RDI + ffffffffffff88], 0
42627c : lea RDX, RAX + 10
426280 : add RAX, 68
426284 : movsd [RDI + fffff68], XMM3
42628c : mov [RDI + fffff40], RAX
426293 : movsd [RDI + fffff70], XMM4
42629b : mov [RDI + fffff30], RDX
4262a2 : movapd XMM1, XMM3
4262a6 : mov [RDI + ffffffffffff90], 0
4262ae : movapd XMM0, XMM4
4262b2 : mov [RDI + ffffffffffff98], 0
4262ba : mov [RDI + ffffffffffffa0], 0
4262c2 : mov [RDI + ffffffffffffa8], 0
4262ca : mov [RDI + ffffffffffffb0], 0
4262d2 : mov [RDI + ffffffffffffb8], 0
4262da : mov [RDI + ffffffffffffc0], 0
4262e2 : mov [RDI + ffffffffffffc8], 0
4262ea : mov [RDI + ffffffffffffd0], 0
4262f2 : call vecgeom::cxx::Wedge::Wedge
4262f7 : movsd XMM1, [RBX + 30]
4262fc : lea RBP, RSP + 18
426301 : movsd XMM0, [RIP + 6e537]
426309 : lea R12, RSP + 10
42630e : movapd XMM2, XMM1
426312 : mov RDI, RBP
426315 : addsd XMM1, XMM0
426319 : mov RSI, R12
42631c : subsd XMM2, XMM0
426320 : movsd [RBX + a0], XMM1
```

```
18. call ffffe949 + RIP + 5 ➜ _ZNSs4_Rep10_M_disposeERKSaIcE
19. call ffffeab7 + RIP + 5 ➜ _ZNSsC1EPKcRKSaIcE
20. call fffff2c1 + RIP + 5 ➜ getDoubleOpt
21. call ffffe8fd + RIP + 5 ➜ _ZNSs4_Rep10_M_disposeERKSaIcE
22. call 24dd5 + RIP + 5 ➜ vecgeom::cxx::Wedge::Wedge
23. call fffffa0f + RIP + 5 ➜ vecgeom::cxx::UnplacedTube::UnplacedTube
24. call ffffe735 + RIP + 5 ➜ sincos
25. call ffffe6ec + RIP + 5 ➜ sincos
26. call fffff890 + RIP + 5 ➜ vecgeom::cxx::UnplacedTube::UnplacedTube
27. call 935 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::LogicalVolume
28. call 923 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::LogicalVolume
29. call be4 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::PlaceDaughter
30. call b99 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::Place
31. call bc1 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::Place
32. call fffff799 + RIP + 5 ➜ vecgeom::cxx::GeoManager::Instance
33. call fffff790 + RIP + 5 ➜ vecgeom::cxx::GeoManager::Instance
34. call 29784 + RIP + 5 ➜ vecgeom::Benchmarker::Benchmarker
35. call 24dab + RIP + 5 ➜ vecgeom::Benchmarker::SetPoolMultiplier
36. call 2a5f0 + RIP + 5 ➜ vecgeom::Benchmarker::RunInsideBenchmark
37. call 2dcc8 + RIP + 5 ➜ vecgeom::Benchmarker::RunToInBenchmark
38. call 30040 + RIP + 5 ➜ vecgeom::Benchmarker::RunToOutBenchmark
39. call 29878 + RIP + 5 ➜ vecgeom::Benchmarker::~Benchmarker
40. call a00 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::~LogicalVolume
41. call 9f8 + RIP + 5 ➜ vecgeom::cxx::LogicalVolume::~LogicalVolume
```





OK

First | Previous | 1 | Next | Last

search **8** files matched

sublime
three-
libsche
TorusBe
TorusBe
libvecg
test
assembl

Functi

First

searc

nding

4003e0	
400440	_start (41 bytes)
400470	deregister_tm_clones (22 bytes)
4004a0	register_tm_clones (35 bytes)
4004e0	__do_global_dtors_aux (26 bytes)
400500	frame_dummy (40 bytes)
40052d	func1 (19 bytes)
400541	func2 (13 bytes)
40054f	**func3 (35 bytes)**
400573	funcrec (38 bytes)

Function diff

	func1		func3
1	push RBP, RSP	1	push RBP, RSP
2	mov RBP, RSP	2	mov RBP, RSP
3	mov [RBP + ffffffffffffff8], RDI	3	sub RSP, 8
4	mov RAX, [RBP + ffffffffffffff8]	4	mov [RBP + fffffffffffffffc], EDI
5	mov [RAX], 0	5	cmp [RBP + fffffffffffffffc], 0
6	pop RBP, RSP	6	jnz 7 + RIP + 2
		7	mov RAX, 0
		8	jmp a + RIP + 2
		9	mov EAX, [RBP + fffffffffffffffc]
		10	mov EDI, EAX
		11	call ffffffd0 + RIP + 5
		12	leave
7		13	

OK

Functions to diff

func1 in **three-functions**
func3 in **three-functions**

Diff | Hide

search | 8 files matched

sublime-text

three-

libsche

TorusBe

TorusBe

libvecc

test

assembl

Functi

First

searc

nding

4003e0

400440

400470

4004a0 register_tm_clones (35 bytes)

4004e0 __do_global_dtors_aux (26 bytes)

400500 frame_dummy (40 bytes)

40052d func1 (19 bytes)

400541 func2 (13 bytes)

40054f func3 (35 bytes)

400573 **funcrec** (38 bytes)

funcrec Histogram

Function diff

	func3		funcrec
1	push RBP, RSP	1	push RBP, RSP
2	mov RBP, RSP	2	mov RBP, RSP
3	sub RSP, 8	3	sub RSP, 10
4	mov [RBP + fffffffffffffffc], EDI	4	mov [RBP + fffffffffffffffc], EDI
5	cmp [RBP + fffffffffffffffc], 0	5	cmp [RBP + fffffffffffffffc], 0
6	jnz 7 + RIP + 2	6	jns 7 + RIP + 2
7	mov RAX, 0	7	mov RAX, 0
8	jmp a + RIP + 2	8	jmp d + RIP + 2
9	mov EAX, [RBP + fffffffffffffffc]	9	mov EAX, [RBP + fffffffffffffffc]
		10	sub EAX, 1
10	mov EDI, EAX	11	mov EDI, EAX
11	call ffffffd0 + RIP + 5	12	call ffffffdb + RIP + 5
12	leave	13	leave
13		14	

OK

Functions to diff

func3 in **three-functions**
funcrec in **three-functions**

Diff Hide