

Analyzing LHC experiment software in terms of obsolete memory utilization with a focus on ROOT objects

Nathalie Rauschmayr, Sami Kama

IT-SDC-OL, Southern Methodist University

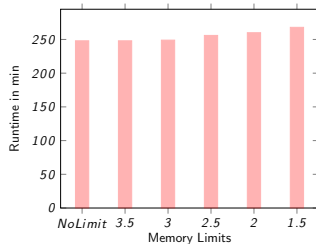
18. September 2015

Introduction

- Many LHC experiments are facing serious memory footprint problems

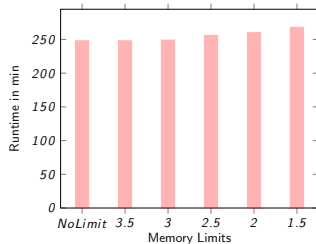
Introduction

- Many LHC experiments are facing serious memory footprint problems



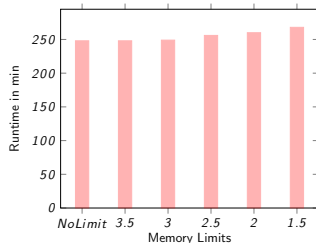
Introduction

- Many LHC experiments are facing serious memory footprint problems
- Which objects are used and which ones unused (=obsolete)?



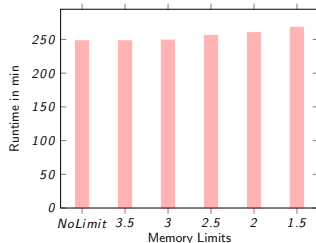
Introduction

- Many LHC experiments are facing serious memory footprint problems
- Which objects are used and which ones unused (=obsolete)?
- What is the contribution of ROOT to this memory consumption?
- Can ROOT help experiments to improve memory footprint/performance?



Introduction

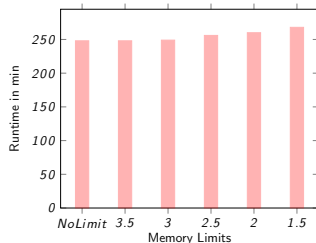
- Many LHC experiments are facing serious memory footprint problems
- Which objects are used and which ones unused (=obsolete)?
- What is the contribution of ROOT to this memory consumption?
- Can ROOT help experiments to improve memory footprint/performance?



New tool is necessary:

Introduction

- Many LHC experiments are facing serious memory footprint problems
- Which objects are used and which ones unused (=obsolete)?
- What is the contribution of ROOT to this memory consumption?
- Can ROOT help experiments to improve memory footprint/performance?



New tool is necessary:

- **unused/obsolete memory != memory leak**
- Allocated objects can be unused and then freed by Python, ROOT etc.

FOM Tools (Find Obsolete Memory)

```
LD_PRELOAD=malloc-hook.so my_process
```


FOM Tools (Find Obsolete Memory)

`LD_PRELOAD=malloc-hook.so my_process`

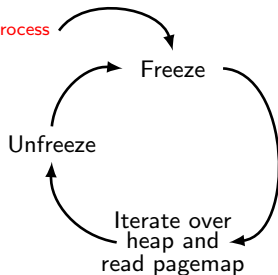


`/cgroup/memory`
`/cgroup/freezer`

FOM Tools (Find Obsolete Memory)

LD_PRELOAD=malloc-hook.so my_process

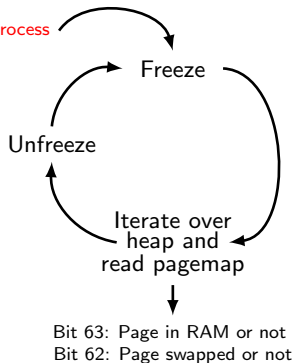
↓
/cgroup/memory
/cgroup/freezer



FOM Tools (Find Obsolete Memory)

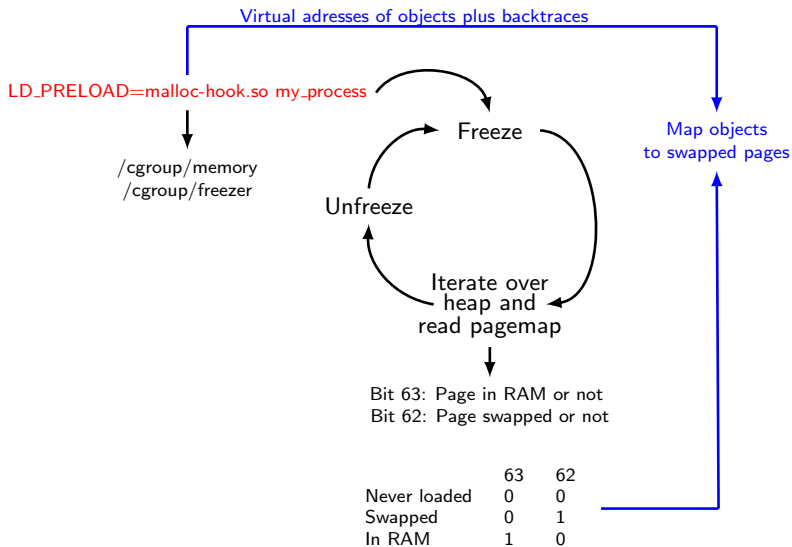
LD_PRELOAD=malloc-hook.so my_process

↓
/cgroup/memory
/cgroup/freezer



	63	62
Never loaded	0	0
Swapped	0	1
In RAM	1	0

FOM Tools (Find Obsolete Memory)



FOM Tools (Find Obsolete Memory)

- How many pages in RAM, Swap or never loaded
- Memory utilization patterns
- Memory hotspots
- Allocation profiles
- Which fraction of large datasets remain in RAM/Swap
- etc....

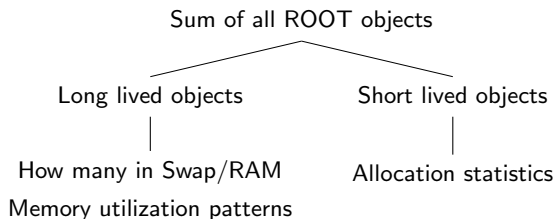
<https://twiki.cern.ch/twiki/bin/view/ITSDC/FomTool>

FOM Tools (Find Obsolete Memory)

- **How many pages in RAM, Swap or never loaded**
- **Memory utilization patterns**
- Memory hotspots
- **Allocation profiles**
- Which fraction of large datasets remain in RAM/Swap
- etc....

<https://twiki.cern.ch/twiki/bin/view/ITSDC/FomTool>

Benchmark tests with a focus on ROOT objects



Benchmark tests with a focus on ROOT objects

Time: 584984.636 Allocation Type: 1 PageBegin: 0x145b000 PageEnd: 0x145bfff Address: 0x145b020 Size: 385 :

```
0 malloc+0x6d in /afs/cern.ch/sw/lcg/contrib/gcc/4.9.1/x86_64-slc6/include/c++/4.9.1/bits/atomic_base.h:308
18 _Znwm+0x1d in ../../../../gcc-4.8.4/libstdc++-v3/libsupc++/new_op.cc:52
425 _ZNSs4_Rep9_S_createEmmRKsAlcE+0x59 in ./libstdc++-v3/include/bits/basic_string.tcc:609
551 _ZNSs4_Rep8_M_cloneERKsAlcEm+0x1b in ./libstdc++-v3/include/bits/basic_string.tcc:630
552 _ZNSs7reserveEm+0x34 in ./libstdc++-v3/include/bits/basic_string.tcc:511
653 _ZNSs6appendEPKcm+0x4f in ./libstdc++-v3/include/bits/basic_string.tcc:308
654 _ZL11DynamicPathPKcb+0x3e5 in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4520
634 _ZN11TUnixSystem14GetDynamicPathEv+0x1b in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4560
635 _ZN11TUnixSystem18FindDynamicLibraryER7TStringb+0xb4 in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4586
636 _ZN7TSystem15DynamicPathNameEPKcb+0x4a in ./ROOT/6.02.12/core/base/src/TSystem.cxx:1906
637 _ZN5TROOT15InitInterpreterEv+0xdf in ./ROOT/6.02.12/core/base/src/TROOT.cxx:1683
638 _ZN4ROOT8GetROOT2Ev+0x2b in ./ROOT/6.02.12/core/base/src/TROOT.cxx:341
82 _ZN4ROOT7GetROOTEv+0xd in ./ROOT/6.02.12/core/base/src/TROOT.cxx:349
639 _ZN12TApplicationC1EPKcPiPPcPvi+0x2c3 in ./ROOT/6.02.12/core/base/src/TApplication.cxx:133
640 _ZN12TApplication17CreateApplicationEv+0x8b in ./ROOT/6.02.12/core/base/src/TApplication.cxx:1253
641 _ZN4xAOD4InitEPKcPiPPc+0xf5 in ??:0
642 _ZN4xAOD4InitEPKc+0xd in ??:0
643 main+0xd2 in ??:0
644 __libc_start_main+0xfd in ??:0
645 _start+0x29 in ??:0
```


Benchmark tests with a focus on ROOT objects

Time: 584984.636 Allocation Type: 1 PageBegin: 0x145b000 PageEnd: 0x145bfff Address: 0x145b020 Size: 385 :

```
0 malloc+0x6d in /afs/cern.ch/sw/lcg/contrib/gcc/4.9.1/x86_64-slc6/include/c++/4.9.1/bits/atomic_base.h:308
18 _Znwm+0x1d in ../../../../gcc-4.8.4/libstdc++-v3/libsupc++/new_op.cc:52
425 _ZNSs4_Rep9_S_createEmmRKsAlcE+0x59 in ./libstdc++-v3/include/bits/basic_string.tcc:609
551 _ZNSs4_Rep8_M_cloneERKsAlcEm+0x1b in ./libstdc++-v3/include/bits/basic_string.tcc:630
552 _ZNSs7reserveEm+0x34 in ./libstdc++-v3/include/bits/basic_string.tcc:511
653 _ZNSs6appendEPKcm+0x4f in ./libstdc++-v3/include/bits/basic_string.tcc:308
654 _ZL11DynamicPathPKcb+0x3e5 in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4520
634 _ZN11TUnixSystem14GetDynamicPathEv+0x1b in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4560
635 _ZN11TUnixSystem18FindDynamicLibraryER7TStringb+0xb4 in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4586
636 _ZN7TSystem15DynamicPathNameEPKcb+0x4a in ./ROOT/6.02.12/core/base/src/TSystem.cxx:1906
637 _ZN5TROOT15InitInterpreterEv+0xdf in ./ROOT/6.02.12/core/base/src/TROOT.cxx:1683
638 _ZN4ROOT8GetROOT2Ev+0x2b in ./ROOT/6.02.12/core/base/src/TROOT.cxx:341
82 _ZN4ROOT7GetROOTEv+0xd in ./ROOT/6.02.12/core/base/src/TROOT.cxx:349
639 _ZN12TApplicationC1EPKcPiPPcPvi+0x2c3 in ./ROOT/6.02.12/core/base/src/TApplication.cxx:133
640 _ZN12TApplication17CreateApplicationEv+0x8b in ./ROOT/6.02.12/core/base/src/TApplication.cxx:1253
641 _ZN4xAOD4InitEPKcPiPPc+0xf5 in ??:0
642 _ZN4xAOD4InitEPKc+0xd in ??:0
643 main+0xd2 in ??:0
644 __libc_start_main+0xfd in ??:0
645 _start+0x29 in ??:0
```

Benchmark tests with a focus on ROOT objects

Time: 584984.636 Allocation Type: 1 PageBegin: 0x145b000 PageEnd: 0x145bfff Address: 0x145b020 Size: 385 :

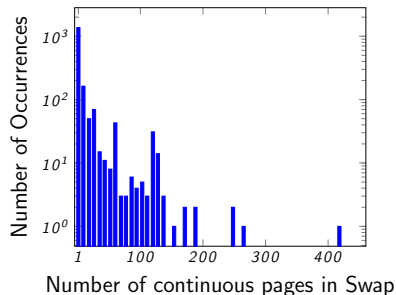
```
0 malloc+0x6d in /afs/cern.ch/sw/lcg/contrib/gcc/4.9.1/x86_64-slc6/include/c++/4.9.1/bits/atomic_base.h:308
18 _Znwm+0x1d in ../../../../gcc-4.8.4/libstdc++-v3/libsupc++/new_op.cc:52
425 _ZNs4_Rep9_S_createEmmRKsAlcE+0x59 in ./libstdc++-v3/include/bits/basic_string.tcc:609
551 _ZNs4_Rep8_M_cloneERKSAlcEm+0x1b in ./libstdc++-v3/include/bits/basic_string.tcc:630
552 _ZNs7reserveEm+0x34 in ./libstdc++-v3/include/bits/basic_string.tcc:511
653 _ZNs6appendEPKcm+0x4f in ./libstdc++-v3/include/bits/basic_string.tcc:308
654 _ZL11DynamicPathPKcb+0x3e5 in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4520 Terminator
634 _ZN11TUnixSystem14GetDynamicPathEv+0x1b in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4560
635 _ZN11TUnixSystem18FindDynamicLibraryER7TStringb+0xb4 in ./ROOT/6.02.12/core/unix/src/TUnixSystem.cxx:4586
636 _ZN7TSystem15DynamicPathNameEPKcb+0x4a in ./ROOT/6.02.12/core/base/src/TSystem.cxx:1906
637 _ZN5TROOT15InitInterpreterEv+0xdf in ./ROOT/6.02.12/core/base/src/TROOT.cxx:1683
638 _ZN4ROOT8GetROOT2Ev+0x2b in ./ROOT/6.02.12/core/base/src/TROOT.cxx:341
82 _ZN4ROOT7GetROOTEv+0xd in ./ROOT/6.02.12/core/base/src/TROOT.cxx:349
639 _ZN12TApplicationC1EPKcPiPPcPvi+0x2c3 in ./ROOT/6.02.12/core/base/src/TApplication.cxx:133
640 _ZN12TApplication17CreateApplicationEv+0x8b in ./ROOT/6.02.12/core/base/src/TApplication.cxx:1253 Initiator
641 _ZN4xAOD4InitEPKcPiPPc+0xf5 in ??:0
642 _ZN4xAOD4InitEPKc+0xd in ??:0
643 main+0xd2 in ??:0
644 __libc_start_main+0xfd in ??:0
645 _start+0x29 in ??:0
```

Analysis Job

- **Vmem:** 1 GB, **Rss:** 870 MB, **Heap Vmem:** 532 MB
- Long lived ROOT objects:
 - ≈ 260 MB ($\approx 48\%$ of the heap)
 - Most of the time in Swap: ≈ 22 k pages
(33% of long lived ROOT objects)
 - Always in Ram: ≈ 12 k pages
(18% of long lived ROOT objects)

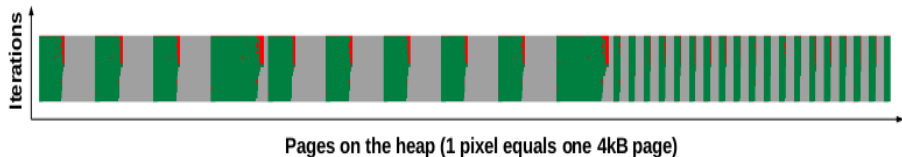
Analysis Job

- **Vmem:** 1 GB, **Rss:** 870 MB, **Heap Vmem:** 532 MB
- Long lived ROOT objects:
 - ≈ 260 MB ($\approx 48\%$ of the heap)
 - Most of the time in Swap: ≈ 22 k pages (33% of long lived ROOT objects)
 - Always in Ram: ≈ 12 k pages (18% of long lived ROOT objects)



Analysis Job

Memory Utilization Patterns:



Red pixel: Page is in Swap and contains a ROOT object

Green pixel: Page is in RAM and contains a ROOT object

Gray pixel: Page was never loaded into RAM and contains a ROOT object

Analysis Job

Memory Utilization Patterns:



Analysis Job

Memory Utilization Patterns:



- Read TTree from file (progressively)
- ROOT creates TBuffer
- In this case: size of allocations above MMAP_THRESHOLD
 - malloc will use anonymous mapping (page is not in RAM until process tries to access it for the first time)
- Efficiency ratio $\leq 50\%$ (What is allocated and actually needed for reading)

Analysis Job

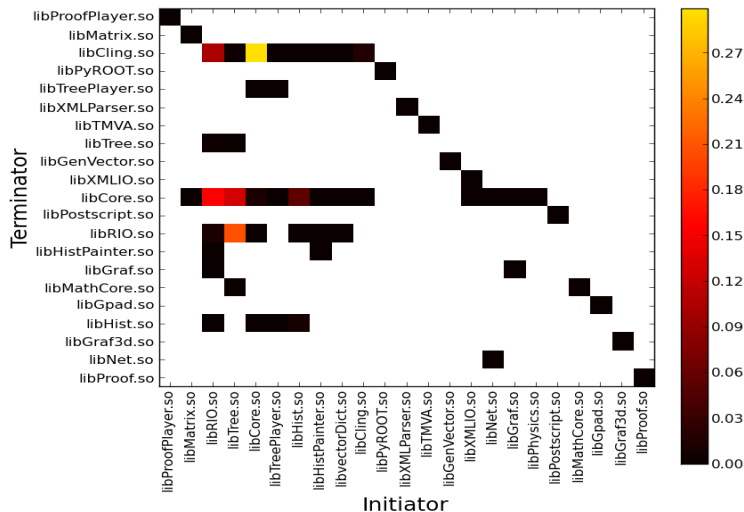


Figure: Percentage of allocations (size) involving ROOT libraries (excluding TObject Allocator)

Analysis Job

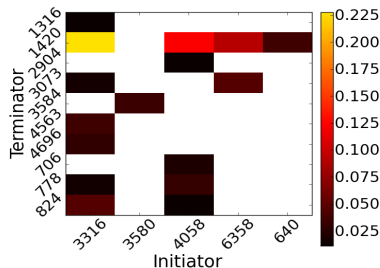


Figure: Percentage of allocations (size) initiated by different function calls and terminating in libCling

```
3316: libCore.so _ZN6TClass8GetClassEPKcbb+0x6e3 in ./ROOT/6.02.12/core/meta/src/TClass.cxx:2881
4058: libRIO.so _ZN5TFile4OpenEPKcS1_S1_ii+0xcc0 in ./ROOT/6.02.12/io/io/src/TFile.cxx:3977
6358: libCore.so _ZN11TMethodCall17InitWithPrototypeEP6TClassPKcS3_bN4ROOT18EFunctionMatchModeE+0xbb in
      ./ROOT/6.02.12/core/meta/src/TMethodCall.cxx:344

1420: libCling.so _ZNK5clang15DeclarationName11getAsStringEv+0x69
```

Analysis Job

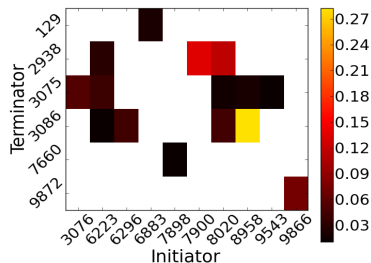


Figure: Percentage of allocations (size) initiated by different function calls and terminating in libCore

```
8958: libTree.so _ZN14TBranchElement8GetEntryExi+0x222
      in ./ROOT/6.02.12/tree/tree/src/TBranchElement.cxx:2310
8020: libRIO.so _ZN14TDirectoryFile3GetEPKc+0xf9
      in ./ROOT/6.02.12/io/io/src/TDirectoryFile.cxx:866
7900: libHist.so _ZNK3TH15CloneEPKc+0x96 in ./ROOT/6.02.12/hist/hist/src/TH1.cxx:2568
3086: libCore.so R__unzip+0x2f5 in ./ROOT/6.02.12/core/zip/src/ZInflate.c:1198
2938: libCore.so _ZN6TExMapC2Ei+0xb3 in ./ROOT/6.02.12/core/cont/src/TExMap.cxx:46
```

Analysis Job

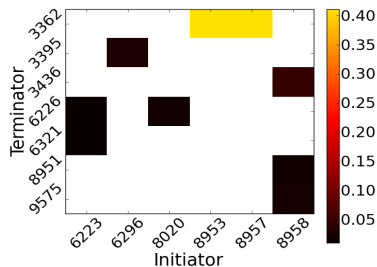


Figure: Percentage of allocations (size) initiated by different function calls and terminating in libRIO

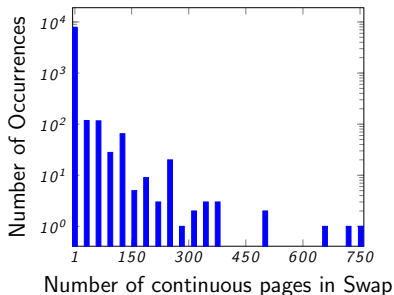
```
8953: libTree.so _ZN14TBranchElement12SetMakeClassEb+0xb3
      in ./ROOT/6.02.12/tree/tree/src/TBranchElement.cxx:4987
8957: libTree.so _ZN14TBranchElement12SetMakeClassEb+0xbf
      in ./ROOT/6.02.12/tree/tree/src/TBranchElement.cxx:4989
3362: libRIO.so _ZN9__gnu_cxx13new_allocatorIN20TStreamerInfoActions17TConfiguredActionEE8allocateEmPKv+0x40
      in ./gcc/4.8.4/x86_64-slc6/include/c++/4.8.4/ext/new_allocator.h:105
```

ESD to AOD

- **Vmem:** 8.2 GB, **Rss:** 5.4 GB, **Heap Vmem:** 6.35
- Long lived ROOT objects ($> 4\text{kB}$):
 - $\approx 1.2\text{ GB} \approx 19\%$ of the heap
 - Most of the time in Swap: $\approx 93\text{k}$ pages
(27% of long lived ROOT objects)
 - Always in Ram: $\approx 52\text{k}$ pages
(15% of long lived ROOT objects)

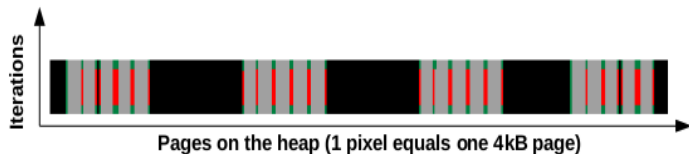
ESD to AOD

- **Vmem:** 8.2 GB, **Rss:** 5.4 GB, **Heap Vmem:** 6.35
- Long lived ROOT objects ($> 4\text{kB}$):
 - $\approx 1.2\text{ GB} \approx 19\%$ of the heap
 - Most of the time in Swap: $\approx 93\text{k}$ pages (27% of long lived ROOT objects)
 - Always in Ram: $\approx 52\text{k}$ pages (15% of long lived ROOT objects)



ESD to AOD

Memory Utilization Patterns:



- Red pixel:** Page is in Swap and contains a ROOT object
- Green pixel:** Page is in RAM and contains a ROOT object
- Gray pixel:** Page was never loaded into RAM and contains a ROOT object
- Black pixel:** Page does not contain a ROOT object

ESD to AOD

Memory Utilization Patterns:



- Monitoring during event loop
- It creates: histogram for data and a TTree for metadata (4 branches)
- Only tiny amounts of the tree are filled
 - 35 KB allocated but only filled with few bytes

ESD to AOD

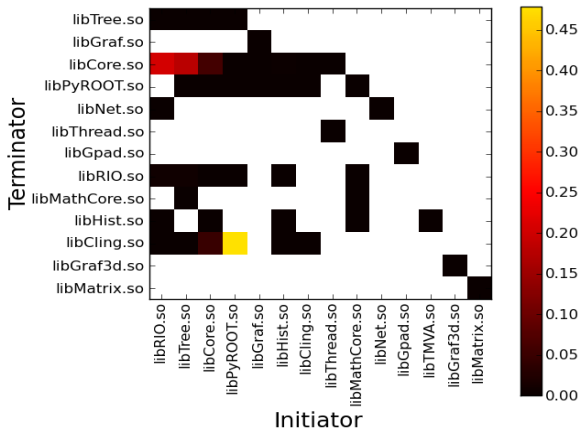


Figure: Percentage of allocations (size) involving ROOT libraries (excluding TObject Allocator)

ESD to AOD

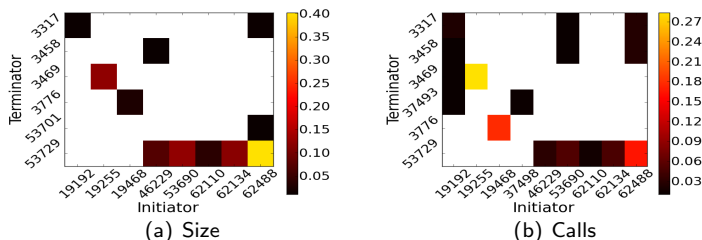


Figure: Percentage of allocations (size/calls) initiated by different function calls and terminating in libCore

62488: libRIO.so _ZN5TFile5WriteEPKcii+0x1e7 in ./ROOT/6.02.08/io/io/src/TFile.cxx:2221

53729: libCore.so R__zipMultipleAlgorithm+0x3c2 in ./ROOT/6.02.08/core/zip/src/Bits.h:678

19255: libTree.so _ZN7TBranch8GetEntryExi+0x289 in ./ROOT/6.02.08/tree/tree/src/TBranch.cxx:1241

3469: libCore.so R__unzip+0x2f5 in ./ROOT/6.02.08/core/zip/src/ZInflate.c:1198

ESD to AOD

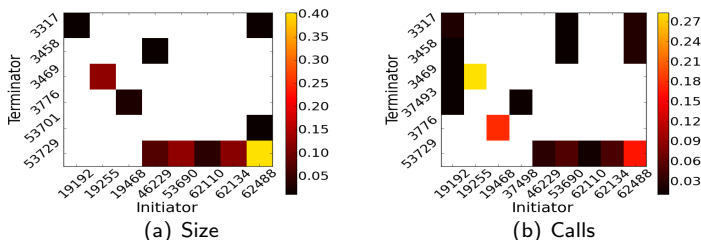


Figure: Percentage of allocations (size/calls) initiated by different function calls and terminating in libCore

```
62488: libRIO.so _ZN5TFile5WriteEPKcii+0x1e7 in ./ROOT/6.02.08
```

```
53729: libCore.so R__zipMultipleAlgorithm+0x3c2 in ./ROOT/6.02.08
```

```
19255: libTree.so _ZN7TBranch8GetEntryExi+0x289 in ./ROOT/6.02.08/tree/tree/src/TBranch.cxx:1241
```

```
3469: libCore.so R__unzip+0x2f5 in ./ROOT/6.02.08/core/zip/src/ZInflate.c:1198
```

Buffers are allocated/freed for each request		
	Size	Occurrences
Top 1	65 kB	105.000
Top 2	11 kB	74.000
Top 3	65 kB	72.000

ESD to AOD

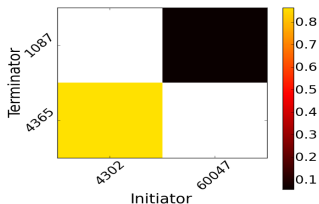


Figure: Percentage of allocations (size) initiated by different function calls and terminating in libCling

```
4302: libPyROOT.so _ZN6PyROOT12_GLOBAL__N_17mp_callEPNS_11MethodProxyEP7_objectS4_+0x138  
      in ./ROOT/6.02.08/bindings/pyroot/src/MethodProxy.cxx:464
```

```
4365: libCling.so _ZNK4TClingCallFunc4execEPvS0_+0x116c  
      in ./ROOT/6.02.08/core/meta/src/TClingCallFunc.cxx:1882
```

Summary

- FOM Tools can analyze:
 - Obsolete allocations

Summary

- FOM Tools can analyze:
 - Obsolete allocations
 - Allocation profiles and correlations

Summary

- FOM Tools can analyze:
 - Obsolete allocations
 - Allocation profiles and correlations
- Some unexpected code behaviors:

Summary

- FOM Tools can analyze:
 - Obsolete allocations
 - Allocation profiles and correlations
- Some unexpected code behaviors:
 - Zip Buffers
 - Initial buffer sizes

Backup Slides

AliRoot PbPbBench Reconstruction Step

- **Vmem:** 4 GB, **Rss:** 2.9 GB, **Heap Vmem:** 2.8 GB
- Long lived ROOT objects ($> 4\text{kB}$):
 - 70k Objects
 - ≈ 2.1 GB ($\approx 75\%$ of the heap)
 - Most of the time in Swap: $\approx 101\text{k}$ pages
(18% of long lived ROOT objects)
 - Always in Ram: $\approx 135\text{k}$ pages
(24% of long lived ROOT objects)

AliRoot PbPbBench Reconstruction Step

Long lived ROOT objects which move to Swap and remain there - an example:

- Read bunch of classes with pointers to each other causing a lot TExMap entries (3000 pages)
- During reconstruction: large TMatrix allocations (5800 pages)
- Many streamers buffers for reading (1000 pages)