



Improving Performance with `async-profiler`

2020

Andrei Pangin
Odnoklassniki

About me



 AndreiPangin

- Principal Engineer @ Odnoklassniki
- JVM enthusiast
- Top #jvm answerer on Stack Overflow
- Author of async-profiler

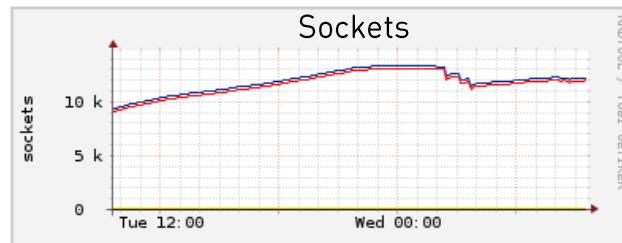
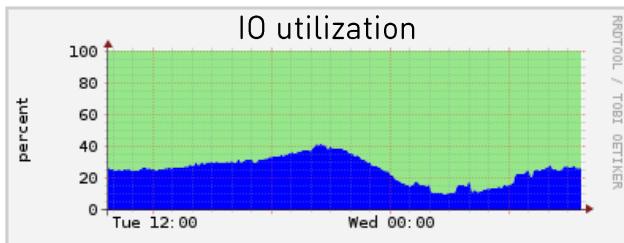
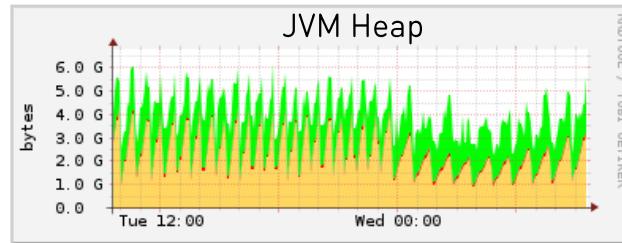
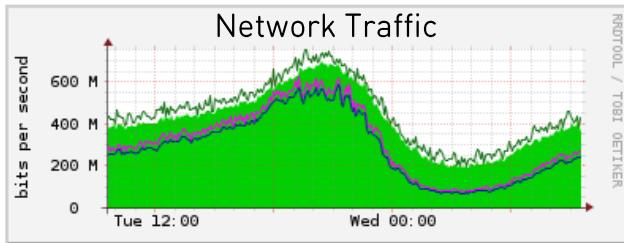
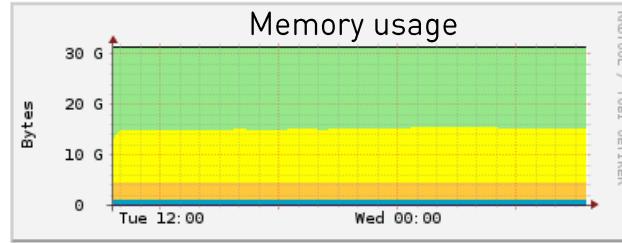
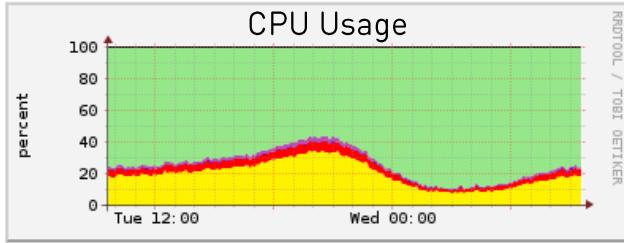
<https://github.com/jvm-profiling-tools/async-profiler>



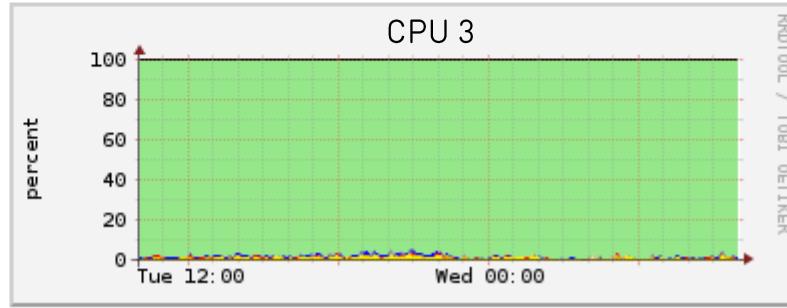
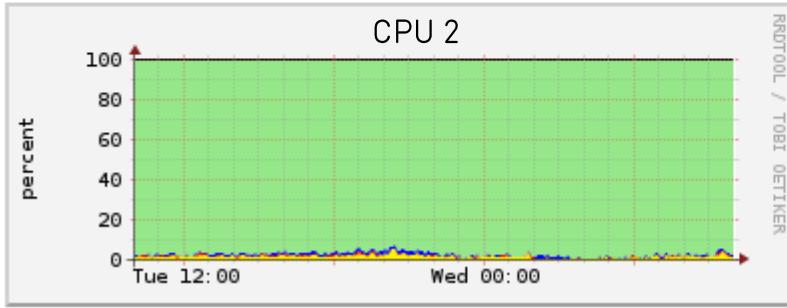
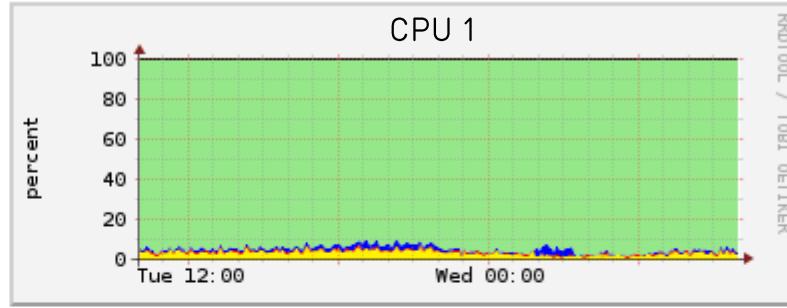
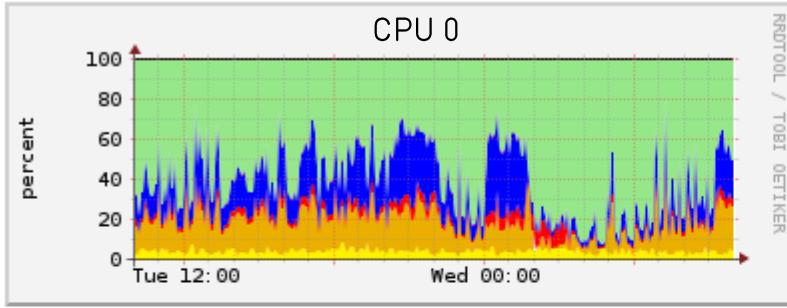
Agenda

1. Profiling theory
2. What's wrong with the classical approach
3. Introduction to async-profiler
4. Allocation profiling
5. Advanced techniques

What to optimize?



To profile or not to profile?





CPU profiling



How to profile

Instrumenting

Trace method transitions

- Sloooow

Sampling

Periodic snapshots

- Production ready

Thread Dump



Java API

`Thread.getAllStackTraces()`



`Map<Thread, StackTraceElement[]>`

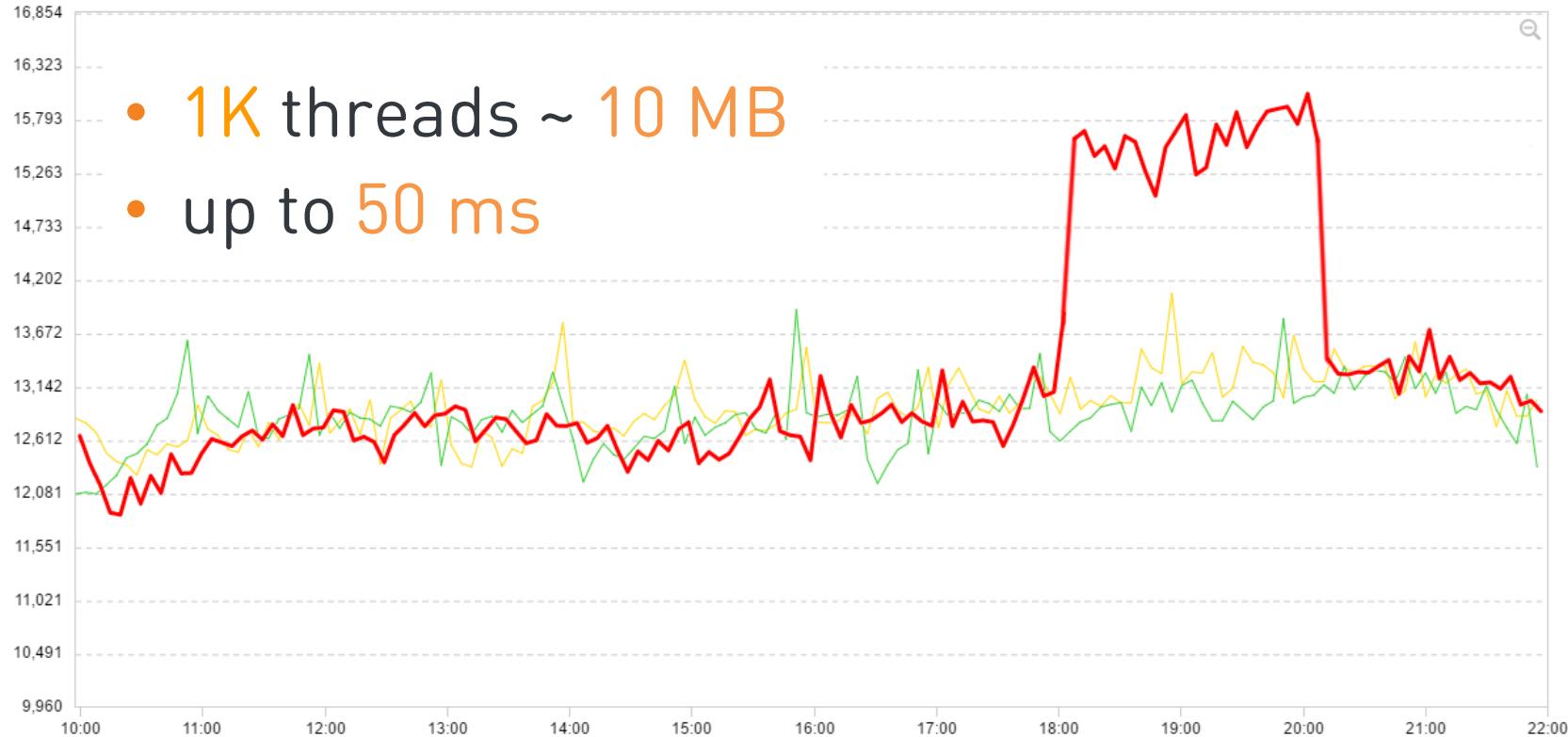
Native (JVMTI)

`GetAllStackTraces()`

Overhead



- 1K threads ~ 10 MB
- up to 50 ms





Advantages

- Simple
 - All Java platforms
 - No extra JVM options required
-
- ✓ Supported by many profilers



Replying to [@AndreiPangin](#)

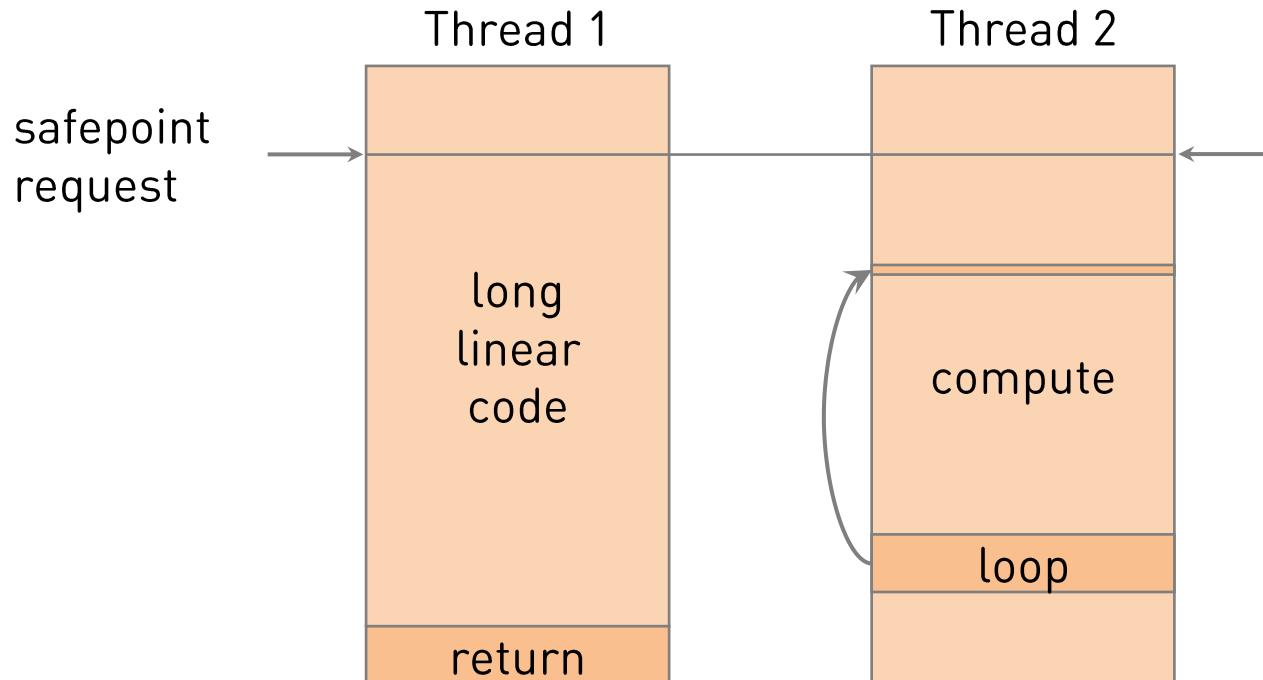
I'm happy with Profiler ... no need for another tool



Demo

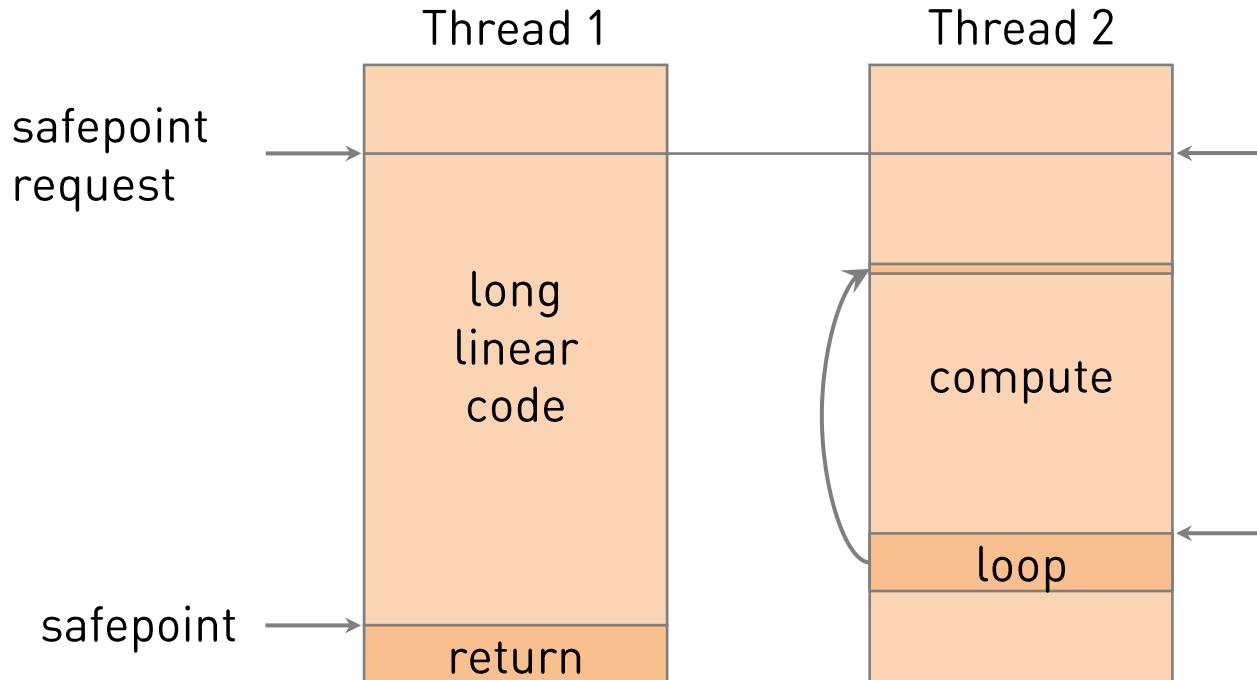


Safepoint





Safepoint





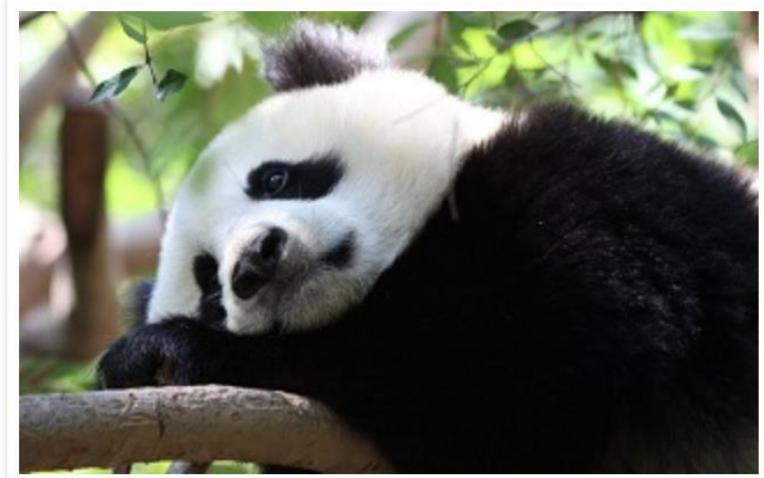
Wednesday, 24 February 2016

Why (Most) Sampling Java Profilers Are F███ Terrible

This post builds on the basis of a [previous post on safepoints](#). If you've not read it you might feel lost and confused. If you have read it, and still feel lost and confused, and you are certain this feeling is related to the matter at hand (as opposed to an existential crisis), please ask away.

So, now that we've established what safepoints are, and that:

1. Safepoint polls are dispersed at fairly arbitrary points (depending on execution mode, mostly at uncounted loop back edge or method return/entry).
2. Bringing the JVM to a global safepoint is high cost



<http://psy-lob-saw.blogspot.ru/2016/02/why-most-sampling-java-profilers-are.html>



Wednesday, 24 February 2016

*Why (Most) Sampling Java Profilers Are F*** Terrible*

because of Safepoint bias
...not the profiler's fault...

<http://psy-lob-saw.blogspot.ru/2016/02/why-most-sampling-java-profilers-are.html>

Native methods



```
Socket s = new Socket(host, port);

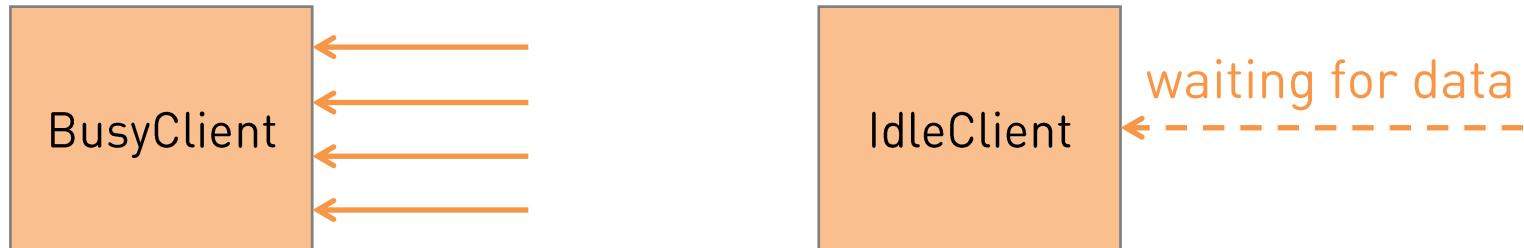
InputStream in = s.getInputStream();
while (in.read(buf) >= 0) {
    // keep reading
}
```

Native methods

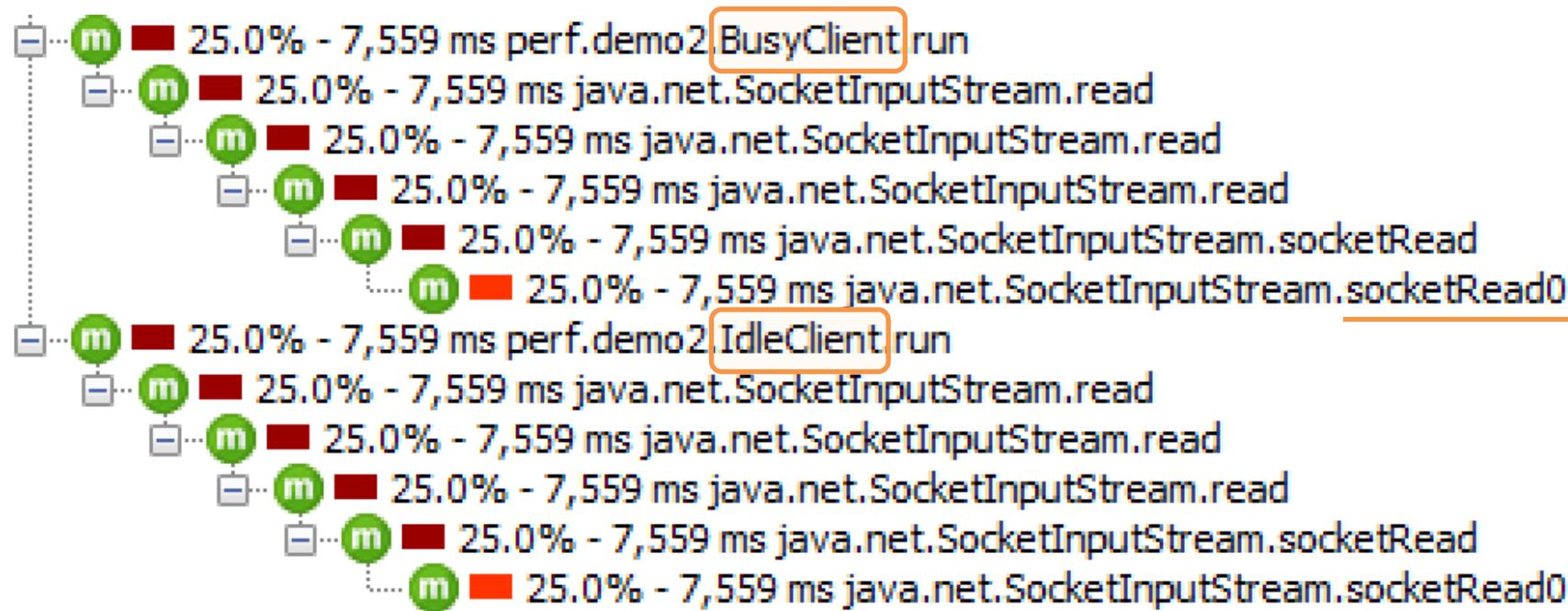


```
Socket s = new Socket(host, port);

InputStream in = s.getInputStream();
while (in.read(buf) >= 0) {
    // keep reading
}
```



Native methods



Solving GetAllStackTraces problems



- Avoid safepoint bias
- Skip idle threads
- Profile native code



AsyncGetCallTrace

How it works



```
AsyncGetCallTrace(ASGCT_CallTrace *trace,  
                  jint depth,  
                  void* ucontext)
```



from signal handler

- Oracle Developer Studio
- github.com/jvm-profiling-tools/honest-profiler
- github.com/jvm-profiling-tools/async-profiler

honest-profiler example



Method	Total %	Self # % ▲
▼ Thread-1		
perf.demo2.IdleClient.run	0,20 %	0,00 %
▼ Thread-2		
▼ perf.demo2.BusyClient.run	96,22 %	0,00 %
▼ java.net.SocketInputStream.read	95,82 %	0,20 %
▼ java.net.SocketInputStream.read	95,62 %	0,00 %
▼ java.net.SocketInputStream.read	95,62 %	0,00 %
▼ java.net.SocketInputStream.socketRead	94,62 %	0,60 %
java.net.SocketInputStream.socketRead0	94,62 %	0,00 %

AsyncGetCallTrace



- Pros
 - Only active threads
 - No safepoint bias
 - XX:+DebugNonSafepoints
- Cons
 - ✖ No native or JVM code
 - ✖ Windows support?



The bad part

Method	Total %	Self # %
▼ main	98,07 %	0,00 %
<u>AGCT.UnknownJavaErr5</u>	95,22 %	95,22 %
▼ perf.demo1.ArrayListTest.main	2,64 %	0,29 %
▶ sun.launcher.LauncherHelper.checkAndLoadMain	0,21 %	0,00 %
▼ Unknown Thread(s)	1,85 %	0,00 %
AGCT.UnknownNotJava3	1,85 %	1,85 %

The bad part



```
double avg(Number... numbers) {  
    double sum = 0;  
    for (Number n : numbers) {  
        sum += n.doubleValue();  
    }  
    return sum / numbers.length;  
}  
  
x = avg(123, 45.67, 890L, 33.3f, 999, 787878L);
```



The bad part

Method	Total %	Self # %
▼ main	99,79 %	0,00 %
<u>AGCT.UnknownJavaErr5</u>	55,16 %	55,16 %
▼ perf.demo3.Numbers.main	44,42 %	0,00 %
▼ perf.demo3.Numbers.loop	44,42 %	10,67 %
▼ perf.demo3.Numbers.avg	33,75 %	24,35 %
java.lang.Long.doubleValue	6,88 %	6,88 %
java.lang.Integer.doubleValue	1,33 %	1,33 %
java.lang.Float.doubleValue	0,70 %	0,70 %
java.lang.Double.doubleValue	0,49 %	0,49 %

Problems



```
enum {
    ticks_no_Java_frame          =  0,
    ticks_no_class_load          = -1,
    ticks_GC_active              = -2,
    ticks_unknown_not_Java       = -3,
    ticks_not_walkable_not_Java = -4,
    ticks_unknown_Java           = -5,
    ticks_not_walkable_Java      = -6,
    ticks_unknown_state          = -7,
    ticks_thread_exit            = -8,
    ticks_deopt                  = -9,
    ticks_safepoint              = -10
};
```

[src/share/vm/prims/forte.cpp](#)



JDK / JDK-8178287



AsyncGetCallTrace fails to traverse valid Java stacks

Details

Type:	Bug	Status:	OPEN
Priority:	P4	Resolution:	Unresolved
Affects Version/s:	8u121, 9, 10	Fix Version/s:	tbd
Component/s:	hotspot		
Labels:	AsyncGetCallTrace analyzer		
Subcomponent:	svc		
CPU:	x86_64		

People

Assignee:



Reporter:



Votes:

[0](#) Vote for this issue

Watchers:

[3](#) Start watching this issue

Description

There is a number of cases (observed in real applications) when profiling with AsyncGetCallTrace is useless due to HotSpot inability to walk through Java stack. Here is the analysis of such cases.

#1. An application performs many System.arraycopy() and spends a lot of time inside

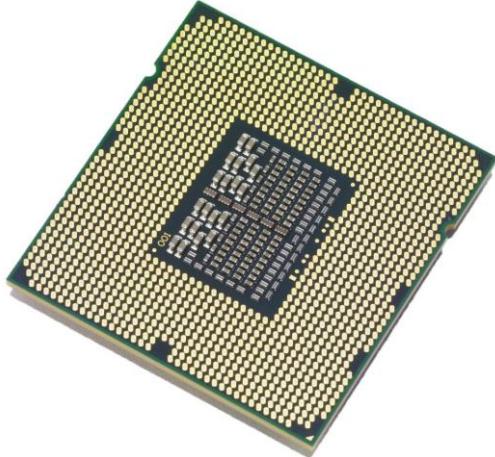
Dates

Created:

2017-04-06 17:32



Perf Events



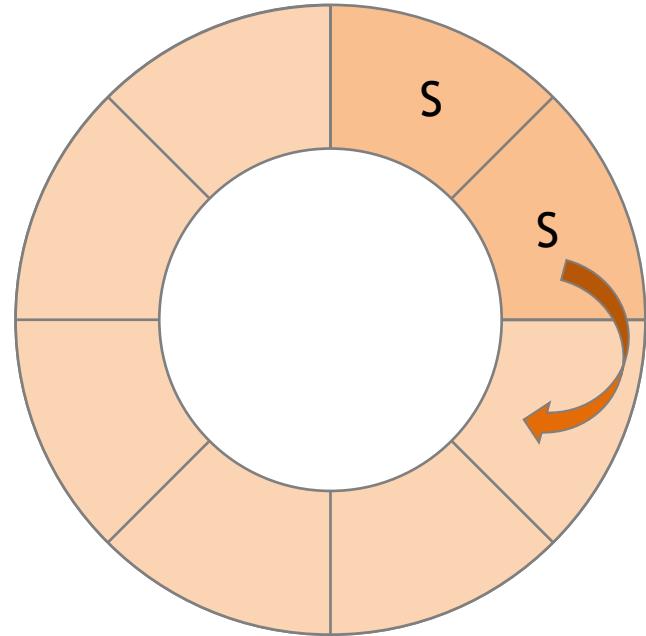
- Hardware counters
 - Cycles, instructions
 - Cache misses
 - Branch misses
 - etc.

Counter overflow → HW interrupt



perf_event_open

- Linux syscall
 - Subscribe to HW/OS events
- Samples
 - pid, tid
 - CPU registers
 - Call chain (user + kernel)





```
$ perf record -F 1009 java ...
$ perf report
```

perf.wiki.kernel.org/index.php/Tutorial

```
$ perf record -F 1009 java ...
```

```
$ perf report
```

4.70%	java	[kernel.kallsyms]	[k]	clear_page_c
2.10%	java	libpthread-2.17.so	[.]	pthread_cond_wait
1.97%	java	libjvm.so	[.]	Unsafe_Park
1.40%	java	libjvm.so	[.]	Parker::park
1.31%	java	[kernel.kallsyms]	[k]	try_to_wake_up
1.31%	java	perf-18762.map	[.]	0x00007f8510e9e757
1.21%	java	perf-18762.map	[.]	0x00007f8510e9e89e
1.17%	java	perf-18762.map	[.]	0x00007f8510e9cc17

perf.wiki.kernel.org/index.php/Tutorial

perf-map-agent



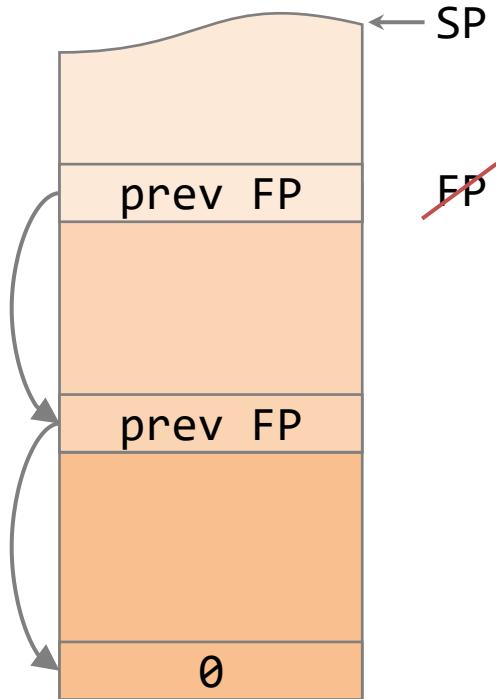
github.com/jvm-profiling-tools/perf-map-agent

```
$ java -agentpath:/usr/lib/libperfmap.so ...
```



7fe0e91175e0	140	java.lang.String::hashCode
7fe0e9117900	20	java.lang.Math::min
7fe0e9117ae0	60	java.lang.String::length
7fe0e9117d20	180	java.lang.String::indexOf

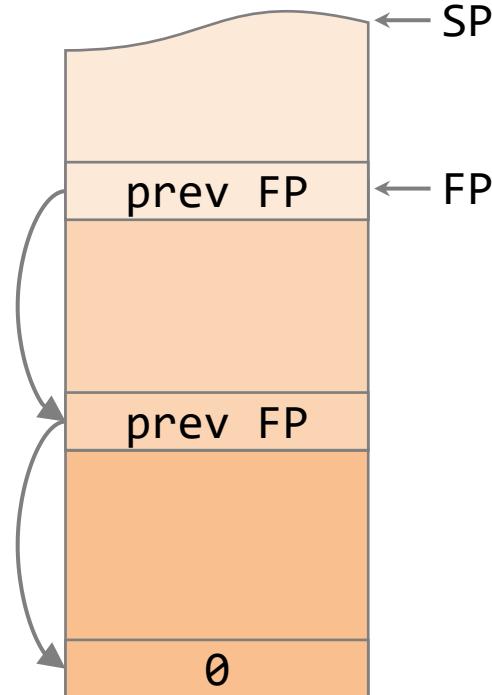
JVM support for perf



JVM support for perf



- -XX:+PreserveFramePointer
- JDK ≥ 8u60
- Permanent overhead (< 5%)



perf + FlameGraph

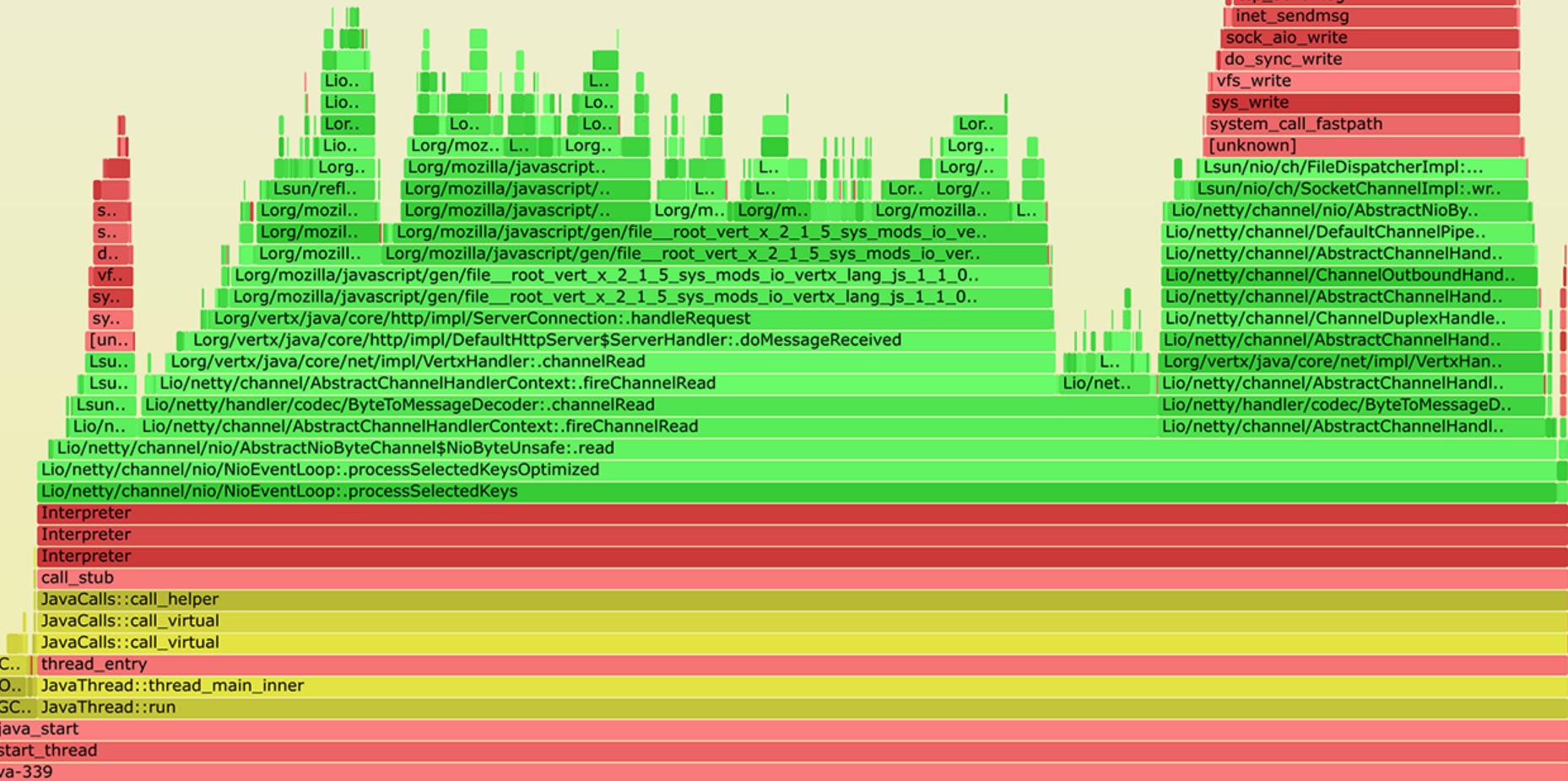


1. perf record
2. perf script
3. FlameGraph/stackcollapse-perf.pl
4. FlameGraph/flamegraph.pl

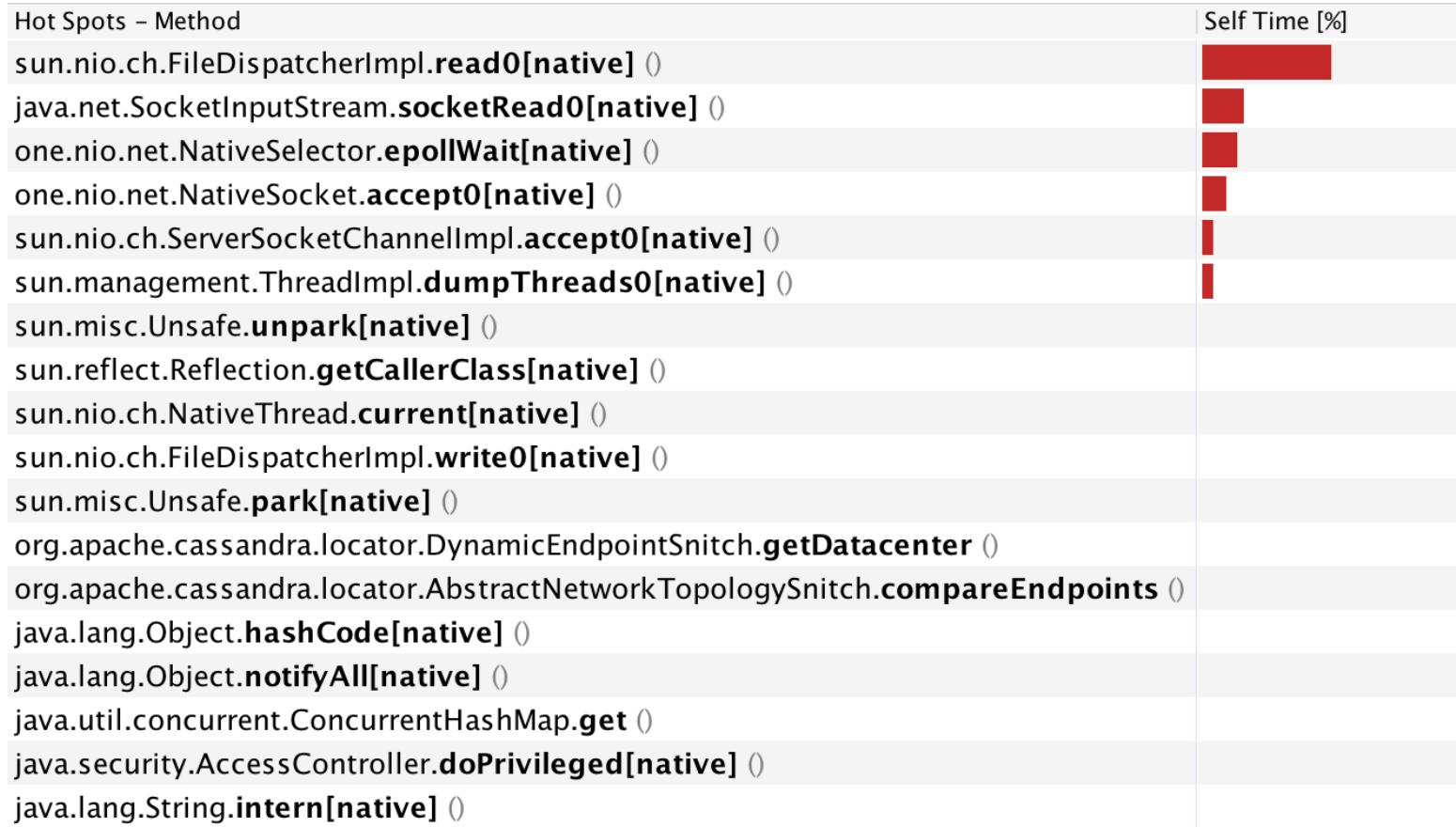


github.com/brendangregg/FlameGraph

Flame Graph



Flat profile



Tree view



Call Tree - Method

Thread-6

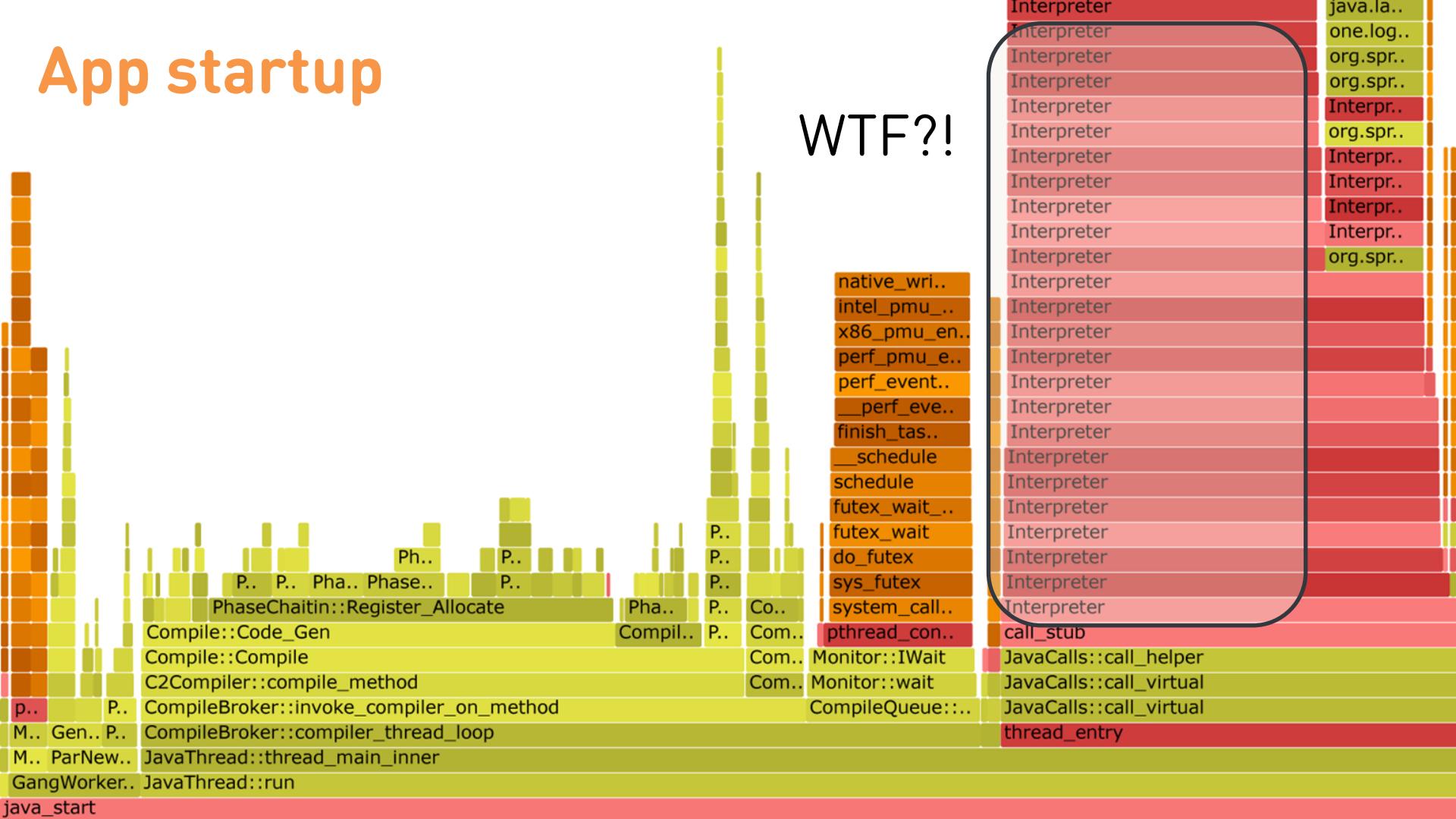
```
    ▼ org.apache.cassandra.net.IncomingTcpConnection.run ()  
      ▼ org.apache.cassandra.net.IncomingTcpConnection.handleModernVersion ()  
        ▼ java.io.DataInputStream.readInt ()  
          ▼ com.google.common.io.CountingInputStream.read ()  
            ▼ java.io.BufferedInputStream.read ()  
              ▼ java.io.BufferedInputStream.fill ()  
                ▼ sun.nio.ch.ChannelInputStream.read ()  
                  ▼ sun.nio.ch.SocketAdaptor$SocketInputStream.read ()  
                    ▼ sun.nio.ch.SocketChannelImpl.read ()  
                      ▼ sun.nio.ch.IOUtil.read ()  
                        ▼ sun.nio.ch.IOUtil.readIntoNativeBuffer ()  
                          ▼ sun.nio.ch.SocketDispatcher.read ()  
                            sun.nio.ch.FileDispatcherImpl.read0[native] ()
```



Demo

App startup

WTF?!



Perf disadvantages



- No interpreted frames
- perf-map-agent for symbols
- -XX:+PreserveFramePointer
- JDK ≥ 8u60
- /proc/sys/kernel/perf_event_paranoia
- /proc/sys/kernel/perf_event_max_stack = 127
- «Big Data»

Mixed approach

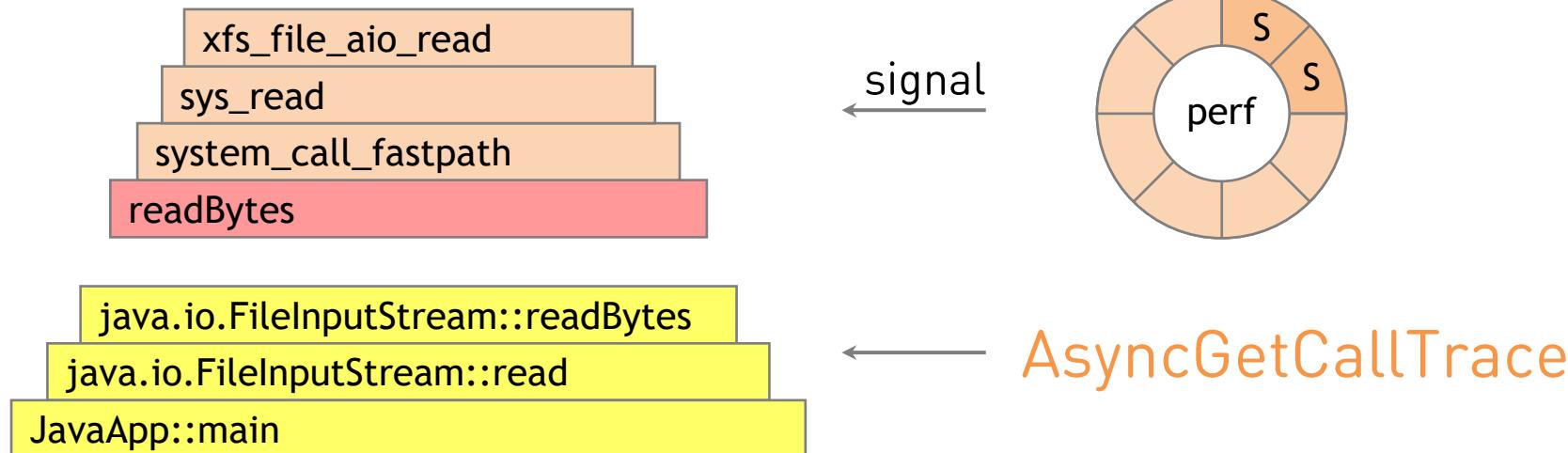


perf_event_open

- Kernel + native stacks
 - HW counters
-
- Entire Java
 - Fast and precise

AsyncGetCallTrace

How it works in async-profiler





Introducing async-profiler



Search or jump to...

/

Pulls

Issues

Marketplace

Explore



jvm-profiling-tools / async-profiler

Unwatch

180

Star

3.4k

Fork

388

About



Sampling CPU and HEAP profiler for Java featuring AsyncGetCallTrace + perf_events

github.com/jvm-profiling-tools/async-profiler

[Readme](#)

[Apache-2.0 License](#)

Releases

17

Maintenance release Latest
8 days ago

Contributors

33





Download

Latest release (1.8.2):

- Linux x64 (glibc): [async-profiler-1.8.2-linux-x64.tar.gz](#)
- Linux x86 (glibc): [async-profiler-1.8.2-linux-x86.tar.gz](#)
- Linux x64 (musl): [async-profiler-1.8.2-linux-musl-x64.tar.gz](#)
- Linux ARM: [async-profiler-1.8.2-linux-arm.tar.gz](#)
- Linux AArch64: [async-profiler-1.8.2-linux-aarch64.tar.gz](#)
- macOS x64: [async-profiler-1.8.2-macos-x64.tar.gz](#)

Supported platforms

- Linux / x64 / x86 / ARM / AArch64
- macOS / x64

Package contents



`async-profiler`

 └── `build`

 └── `async-profiler.jar`

 └── `converter.jar`

 └── `jattach`

 └── `libasyncProfiler.so`

 └── `CHANGELOG.md`

 └── `LICENSE`

 └── `profiler.sh`

 └── `README.md` !

Java API

Standalone converters

Dynamic Attach helper

Core profiler agent

CLI



Demo

Case study: file reading

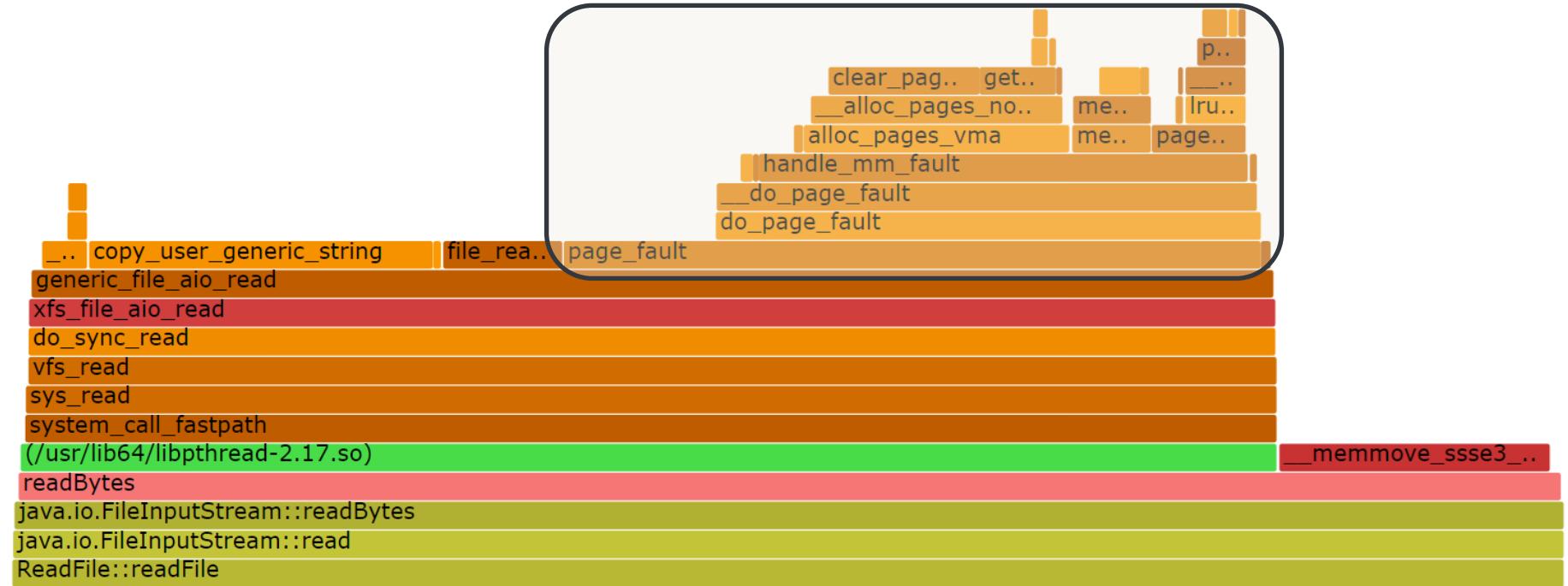


```
byte[] buf = new byte[bufSize];  
  
try (FileInputStream in = new FileInputStream(fileName)) {  
    int bytesRead;  
    while ((bytesRead = in.read(buf)) > 0) {  
        ...  
    }  
}
```

Buffer size?

- 64 K
- 1 M
- 4 M
- 16 M
- 32 M

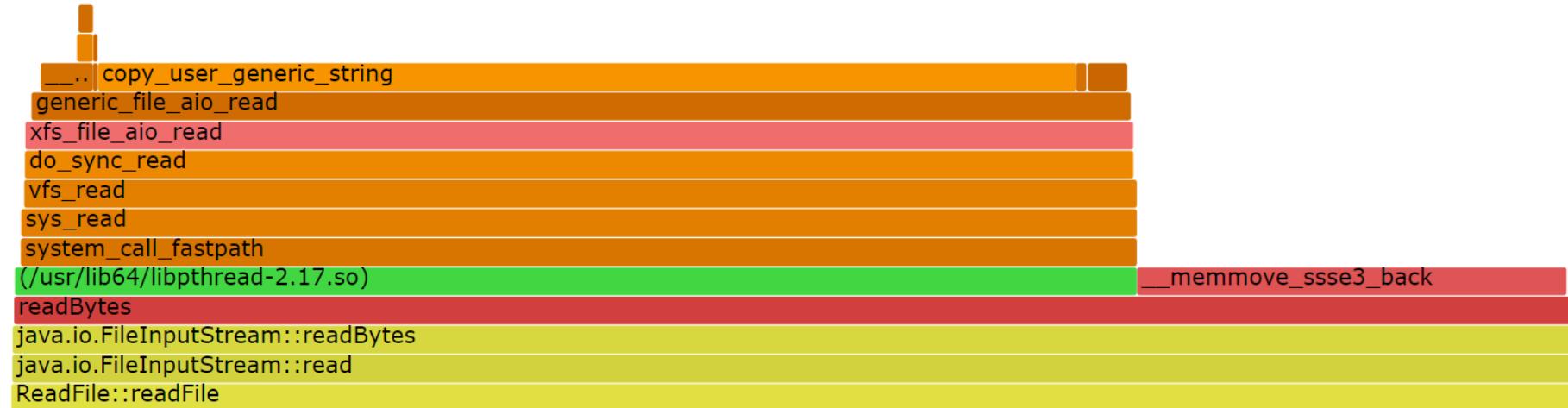
Down to Linux Kernel



Down to Linux Kernel



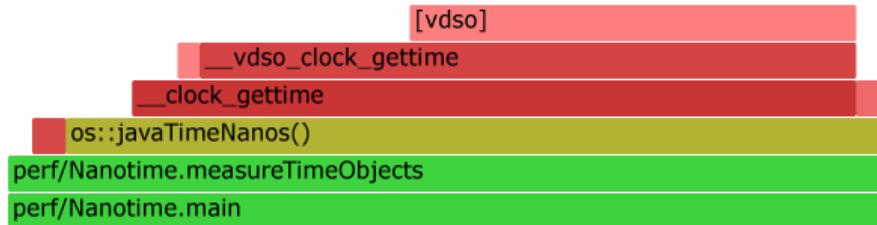
Buffer size: 32M => 31M





Case study: nanoTime

System.nanoTime() latency: 40 ns → 10 000 ns



"good" nanoTime



"bad" nanoTime

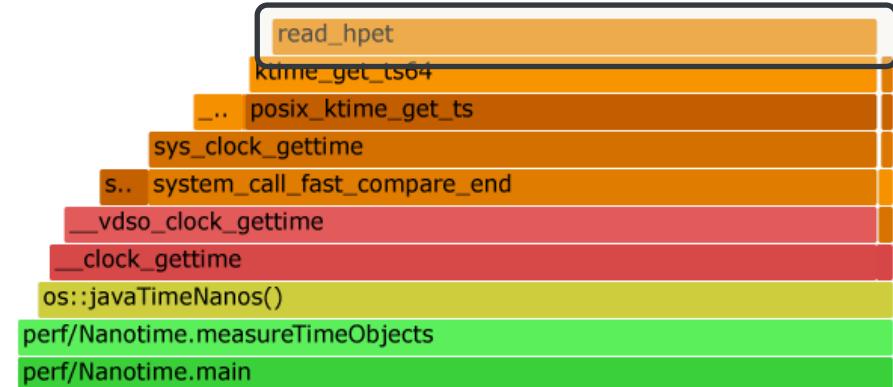
Case study: nanoTime



System.nanoTime() latency: 40 ns → 10 000 ns



"good" nanoTime



"bad" nanoTime

```
# cat /sys/devices/system/clocksource/*/available_clocksource
tsc hpet acpi_pm
# echo tsc > /sys/devices/system/clocksource/*/current_clocksource
```

Clock source on EC2 Linux instances



[https://aws.amazon.com/premiumsupport/
knowledge-center/manage-ec2-linux-clock-source/](https://aws.amazon.com/premiumsupport/knowledge-center/manage-ec2-linux-clock-source/)

```
$ cat /sys/devices/system/clocksource/clocksource0/current_clocksource  
xen
```

Case study: RandomAccessFile.setLength



Products

Home PUBLIC Stack Overflow Tags Users FIND A JOB Jobs Companies TEAMS What's this?

RandomAccessFile.setLength much slower on Java 10 (Centos)

Asked 2 years, 5 months ago Active 2 years, 5 months ago Viewed 621 times

The following code

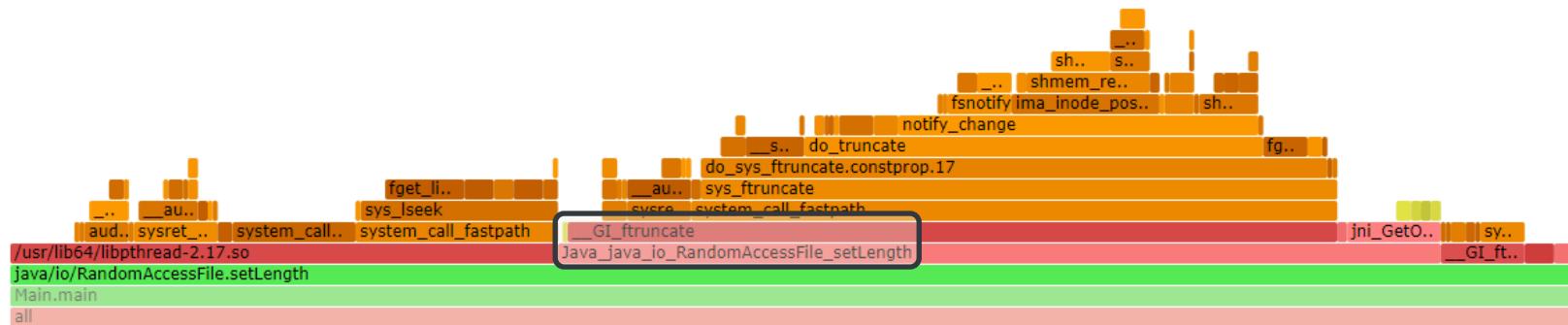
20 5

```
public class Main {
    public static void main(String[] args) throws IOException {
        File tmp = File.createTempFile("deleteme", "dat");
        tmp.deleteOnExit();
        RandomAccessFile raf = new RandomAccessFile(tmp, "rw");
        for (int t = 0; t < 10; t++) {
            long start = System.nanoTime();
            int count = 5000;
            for (int i = 1; i < count; i++)
                raf.setLength((i + t * count) * 4096);
        }
    }
}
```

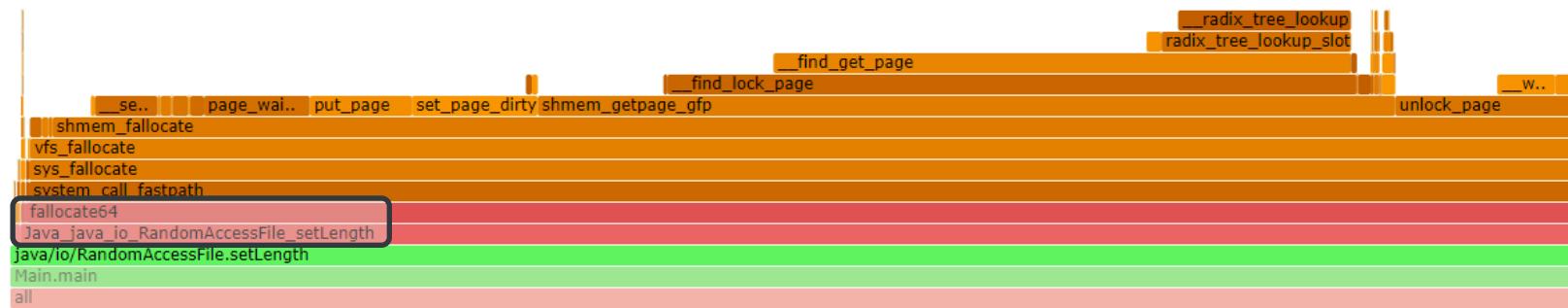
Case study: RandomAccessFile.setLength



JDK 8



JDK 10





Replying to [@AndreiPangin](#)

I'm happy with  Profiler ... no need for another tool



	JFR	perf	async-profiler
Java stack	Yes	No interpreted	Yes
Native stack	No	Yes	Yes
Kernel stack	No	Yes	Yes
JDK support	11+, 8u262+	8u60+	6+
OS support	All	Linux only	Linux, macOS*
Permanent overhead	0	1-5%	0
System-wide profiling	No	Possible	No



Wall clock profiling

Waste time doing nothing



- Thread.sleep
- Object.wait / Condition.await
- Wait to acquire a lock / semaphore / etc.
- Wait for I/O
 - Socket read
 - DB query
 - Disk I/O

Waste time doing nothing



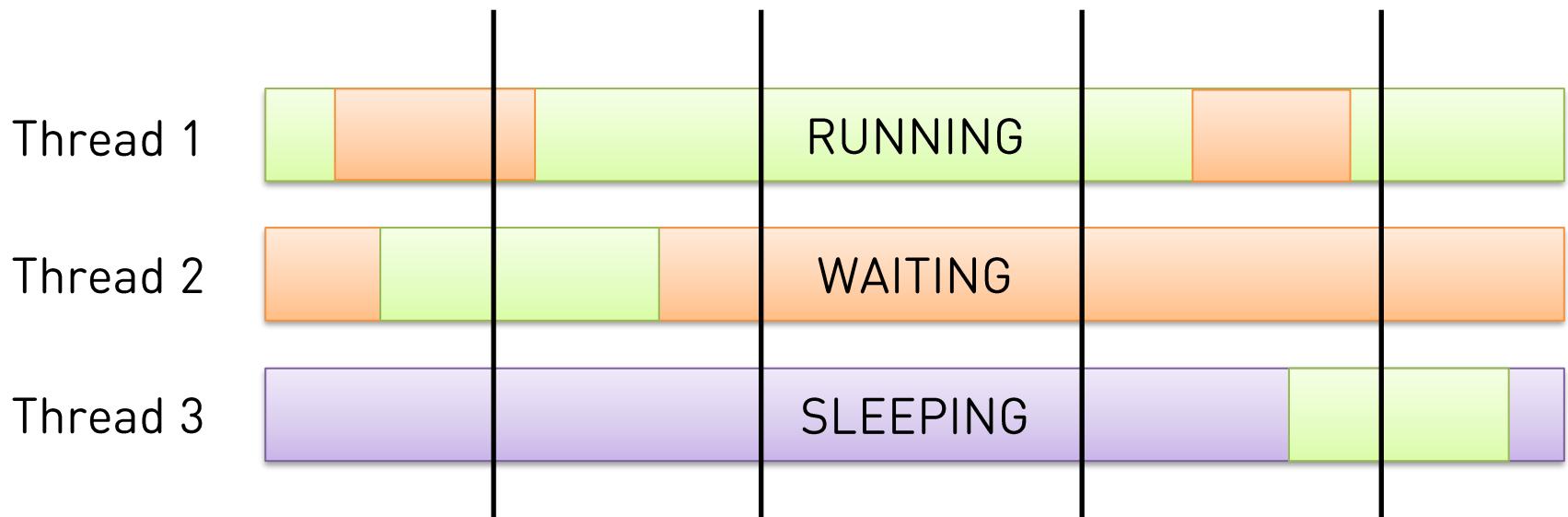
- Thread.sleep
- Object.wait / Condition.await
- Wait to acquire a lock / semaphore / etc.
- Wait for I/O
 - Socket read
 - DB query
 - Disk I/O



CPU profiling is useless

A large, orange, starburst-shaped graphic with a jagged edge, containing the text "CPU profiling is useless".

Wall clock profiling



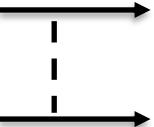
Wall clock mode in async-profiler



CLI option	Agent option
-e wall	event=wall
-t	threads
-i 10ms	interval=10ms



Lock contention

wait time  synchronized (obj) {

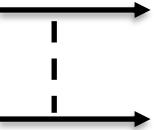
Profiler option

-e lock

-o flamegraph=total



Lock contention

wait time  ReentrantLock.lock()

Profiler option

-e lock

-o flamegraph=total



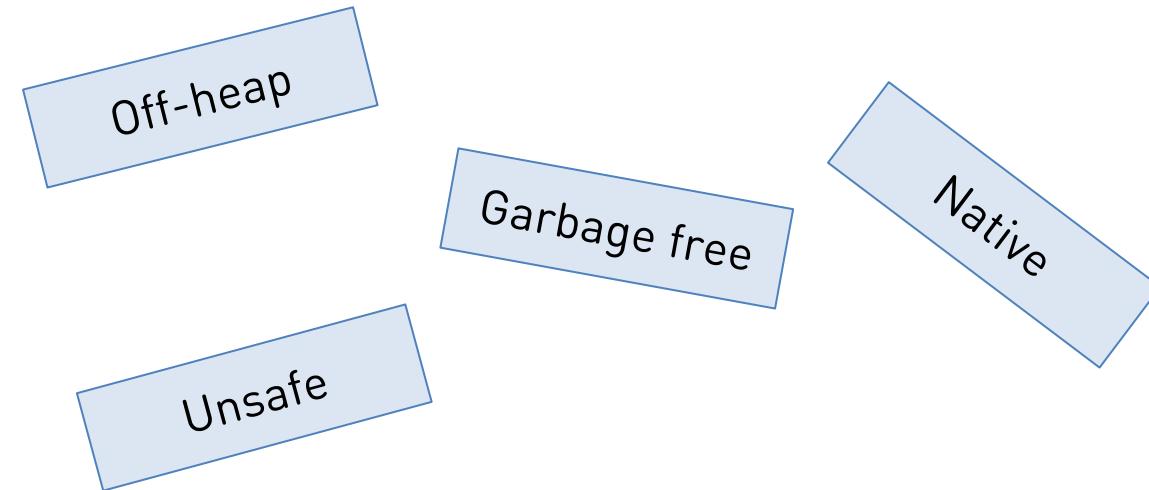
Demo



Allocation profiling



«**new** is the root of all evil»





We want full-featured Java!

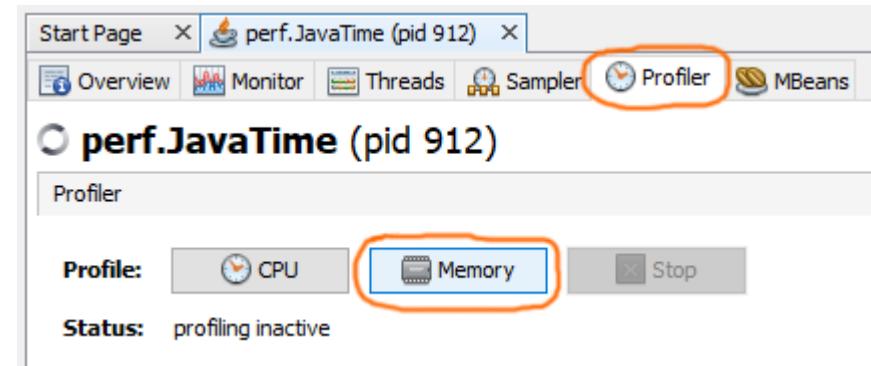
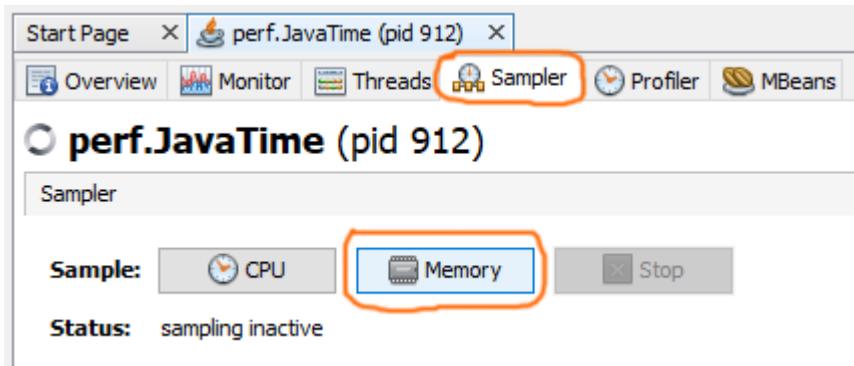


How many objects created?



```
LocalDateTime deadline =  
    LocalDateTime.now().plusSeconds(10);  
  
while (LocalDateTime.now().isBefore(deadline)) {  
    iterations++;  
}  
}
```

Profiling with VisualVM





Sampler

- Walk through heap
- Collect class histogram

Class Name	Bytes [%] ▼	Bytes	Instances	
java.util.concurrent.ConcurrentHashMap		15,441,856 (20.7%)	241,279 (12.2%)	▲
sun.util.calendar.ZoneInfo		13,510,112 (18.1%)	241,252 (12.2%)	☰
java.time.zone.ZoneRules		9,649,880 (12.9%)	241,247 (12.2%)	
java.time.LocalTime		5,790,672 (7.7%)	241,278 (12.2%)	
java.time.LocalDateTime		5,790,120 (7.7%)	241,255 (12.2%)	
java.time.LocalDate		5,790,120 (7.7%)	241,255 (12.2%)	
java.time.ZoneOffset[]		5,789,928 (7.7%)	241,247 (12.2%)	
java.time.ZoneRegion		5,789,928 (7.7%)	241,247 (12.2%)	

Profiler



- Bytecode instrumentation

```
→ Tracer.recordAllocation(ArrayList.class, 24);
List<String> list = new ArrayList<>();
```

- getStackTrace() each Nth sample

DTrace / SystemTap



- -XX:+DTraceAllocProbes
- Slow path allocations

```
$ stap -e '
    probe hotspot.object_alloc { log(probeestr) }
'
```

<https://docs.oracle.com/javase/8/docs/technotes/guides/vm/dtrace.html>
<https://epickrram.blogspot.ru/2017/09/heap-allocation-flamegraphs.html>



Other tools

- Aprof
<https://code.devexperts.com/display/AProf>
- Allocation Instrumenter
<https://github.com/google/allocation-instrumenter>

Overhead



Profiler	x10 ⁶ ops	Slowdown
No profiling	33,2	
VisualVM Sampler	28,2	-15%
	25,6	-23%
	18,1	-45%
VisualVM Profiler	6,8	-80%
JProfiler	2,8	-92%
DTraceAllocProbes	13,3	-60%
Aprof	18,9	-43%
Allocation Instrumenter	22,8	-31%

JMC / Flight Recorder



Allocation Profile

Stack Trace	Average Object Size	Total TLAB size	Pressure
perf.JavaTime.main(String[])	37 bytes	22,75 GB	99,94%
perf.JavaTime.measureTimeObjects(int)	37 bytes	22,75 GB	99,94%
java.time.LocalDateTime.now()	37 bytes	22,75 GB	99,94%
java.time.Clock.systemDefaultZone()	39 bytes	19,33 GB	84,91%
java.time.LocalDateTime.now(Clock)	24 bytes	3,42 GB	15,02%
java.time.LocalDateTime.ofEpochSecond(long, int, ZoneOffset)	24 bytes	3,42 GB	15,02%
java.time.LocalDate.ofEpochDay(long)	24 bytes	1,25 GB	5,51%
java.time.LocalTime.ofNanoOfDay(long)	24 bytes	1,02 GB	4,46%
java.time.LocalTime.create(int, int, int, int)	24 bytes	1,02 GB	4,46%

JMC / Flight Recorder



Allocation Profile

Stack Trace

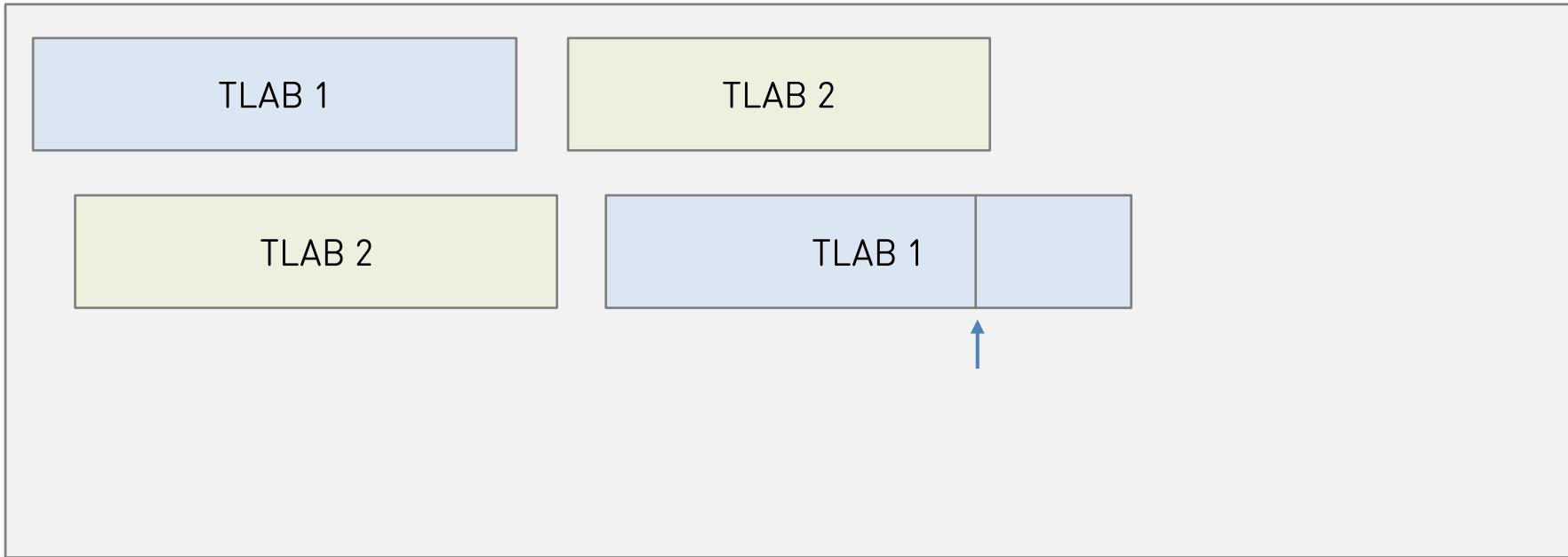
```
perf.JavaTime.main(String[])
  perf.JavaTime.measureTimeObjects(int)
    java.time.LocalDateTime.now()
      java.time.Clock.systemDefaultZone()
        java.time.LocalDateTime.now(Clock)
          java.time.LocalDateTime.ofEpochSecond(long, int, ZoneOffset)
            java.time.LocalDate.ofEpochDay(long)
              java.time.LocalTime.ofNanoOfDay(long)
                java.time.LocalTime.create(int, int, int, int)
```

Average Object Size	Total TLAB size	Pressure
37 bytes	22,75 GB	99,94%
37 bytes	22,75 GB	99,94%
37 bytes	22,75 GB	99,94%
39 bytes	19,33 GB	84,91%
24 bytes	3,42 GB	15,02%
24 bytes	3,42 GB	15,02%
24 bytes	1,25 GB	5,51%
24 bytes	1,02 GB	4,46%
24 bytes	1,02 GB	4,46%

😊 Open source since JDK 11, backported to 8u262

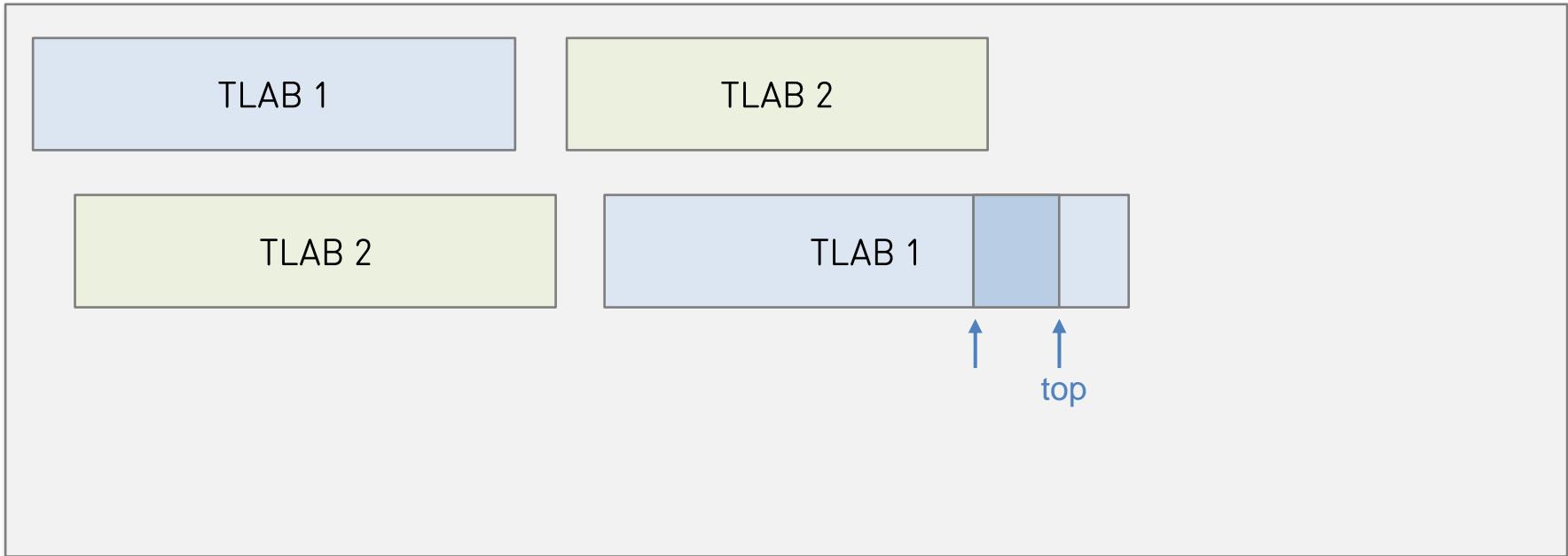
😊 Overhead < 5%

Thread Local Allocation Buffer



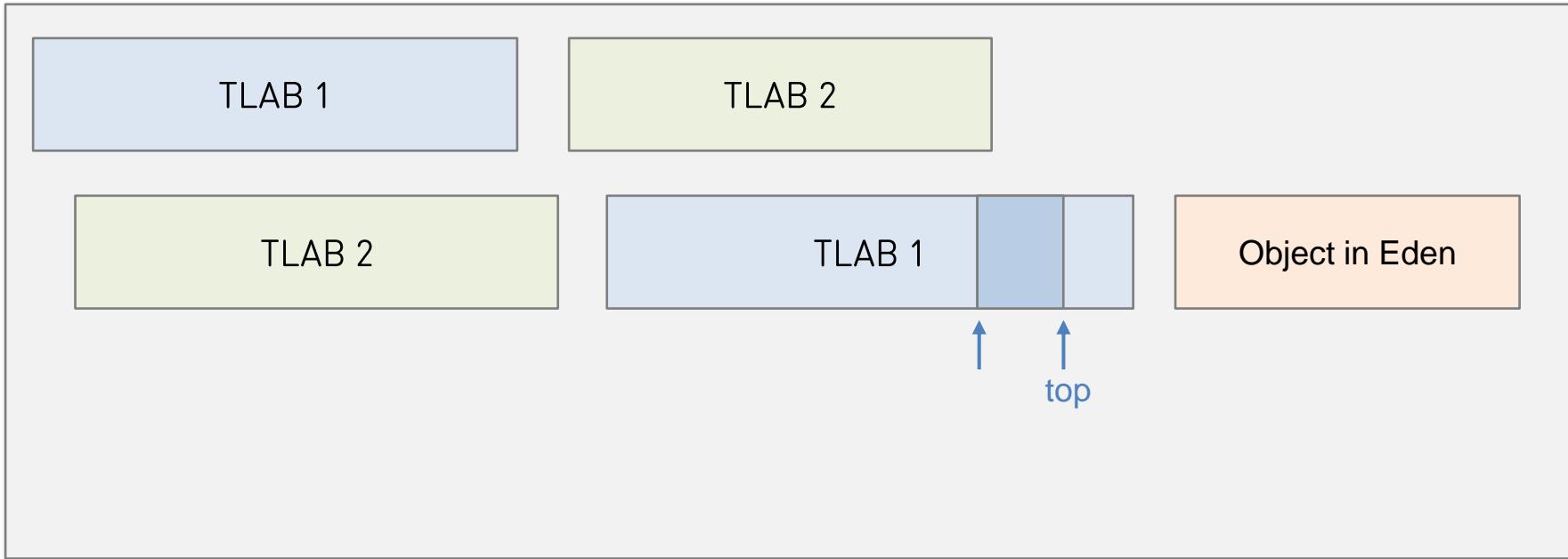
Eden

Thread Local Allocation Buffer



Eden

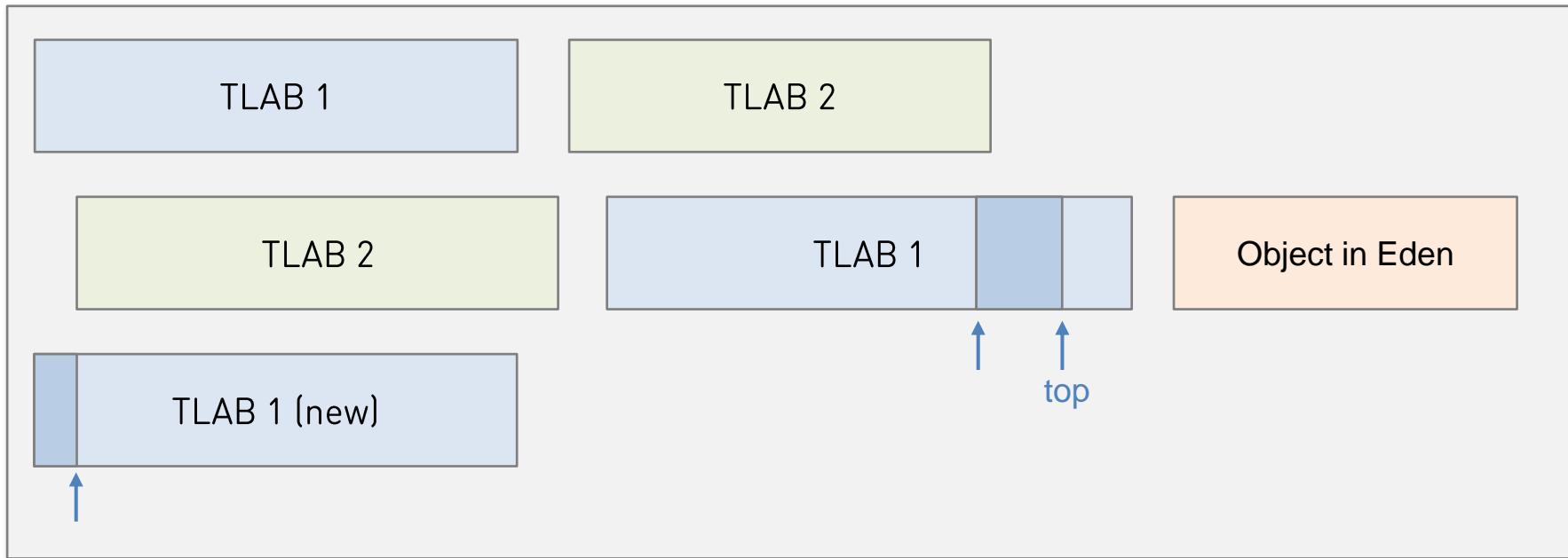
Thread Local Allocation Buffer



Eden



Thread Local Allocation Buffer



Eden



Allocation path

- Fast: inlined
 - Inside TLAB
- Slow: call to VM Runtime
 - Outside TLAB
 - Allocation of new TLAB



Profile



Allocation sampler in async-profiler

1. Intercept slow allocations in JVM runtime
2. Record allocated object + stack trace
- ...
3. PROFIT!



Allocation sampler in async-profiler

1. Intercept slow allocations in JVM runtime
2. Record allocated object + stack trace
- ...
3. PROFIT!



Works with OpenJDK 7u40+



Does not rely on commercial features



Appeared in JDK 11

- JEP 331: Low-Overhead Heap Profiling

```
SampledObjectAlloc(jvmtiEnv *jvmti_env,
                    JNIEnv* jni_env,
                    jthread thread,
                    jobject object,
                    jclass object_klass,
                    jlong size)
```

```
SetHeapSamplingInterval(jint sampling_interval)
```



Demo

async-profiler events



- Built-in
 - cpu (perf events or timer based)
 - wall
 - alloc
 - lock
 - itimer

async-profiler events



- Kernel events
 - page-faults
 - context-switches
- Tracepoints
 - syscalls:sys_enter_* syscalls:sys_exit_*
 - sched:sched_wakeup
 - filemap:mm_filemap_add_to_page_cache
 - see “perf list”



async-profiler events

- PMU
 - cycles, bus-cycles
 - instructions
 - cache-references, cache-misses
 - LLC-load-misses, dTLB-load-misses
 - branches, branch-misses

async-profiler events



- Function breakpoints
 - malloc
 - pthread_start
- VM functions
 - VMThread::execute
 - Deoptimization::uncommon_trap
 - G1CollectedHeap::humongous_obj_allocate
 - java_lang_Throwable::fill_in_stack_trace

async-profiler events



- Java methods
 - MyClass.methodName
 - MyClass.methodName(signature)V
 - MyClass.*

Myths vs. Facts



- ✖ Need to install `perf`
 - ✓ Relies on `perf_event_open` syscall
- ✖ Requires root privileges
 - ✓ `perf_event_paranoid=1` for kernel stack traces or 2 for user-space only
- ✖ Cannot be used in a container
 - ✓ Adjust `seccomp` or fallback to `itimer`

Contribute to async-profiler



- Try it out
- Provide feedback
 - <https://www.baeldung.com/java-async-profiler>
 - <https://hackernoon.com/profiling-java-applications-with-async-profiler-049s2790>

Contribute to async-profiler



- Questions, issues, feature requests

A screenshot of a GitHub repository page. At the top, there is a dark header with the GitHub logo, a search bar containing "Search or jump to...", and navigation links for "Pull requests" and "Issues". Below the header, the repository name "jvm-profiling-tools / async-profiler" is displayed. Underneath the repository name, there are four navigation links: "Code", "Issues 14", "Pull requests 5", and "Actions".

A screenshot of the Stack Overflow question creation interface. At the top, the Stack Overflow logo and a "Products" link are visible. A search bar with the placeholder "Search..." is also at the top. Below the search bar, the word "Tags" is displayed, followed by the instruction "Add up to 5 tags to describe what your question is about". A text input field contains the tag "async-profiler" with a small "X" icon next to it.

Pull Requests are welcome



- Documentation updates
- Bug fixes
- Features?
 - Open an issue for discussion
 - Who will benefit from the feature?
 - Who will support it?



Thank you

@AndreiPangin
<https://pangin.pro>