Idle Time Garbage Collection Scheduling

Hannes Payer, Google Germany

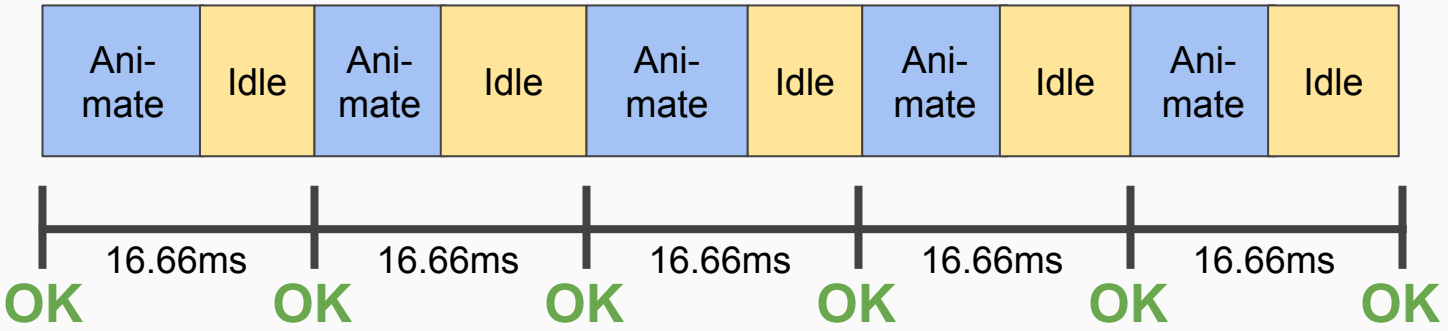
Ulan Degenbaev, Google Germany
Jochen Eisinger, Google Germany
Manfred Ernst, Google USA
Ross McIlroy, Google UK

Overview

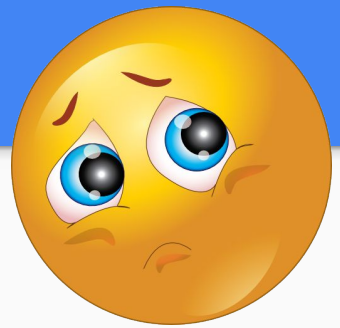
- Chrome & V8
- Idle Time Garbage Collection Scheduling for
  - Improving Latency
  - Improving Memory Consumption
- Experiments
- Related Work
- Conclusions

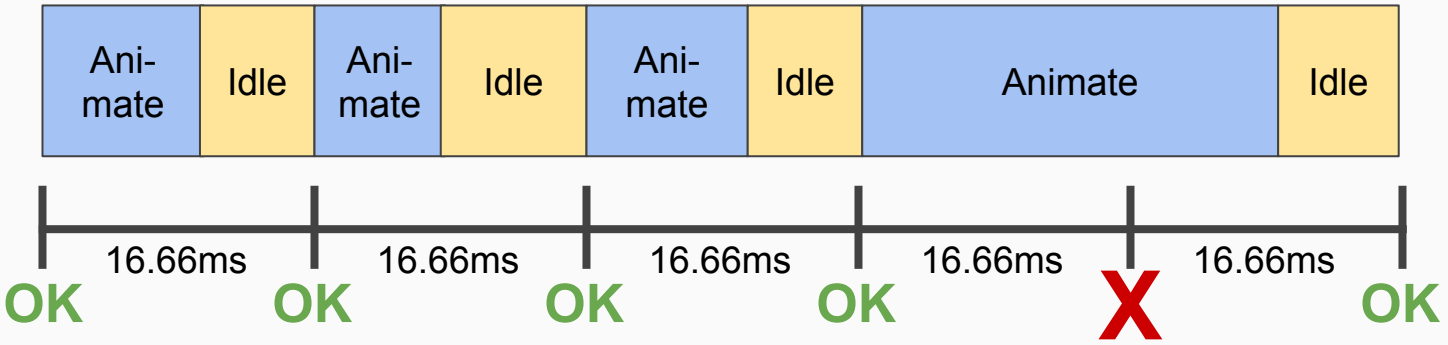
 + 60 fps = 

Animating at 60 fps



Dropping Frames

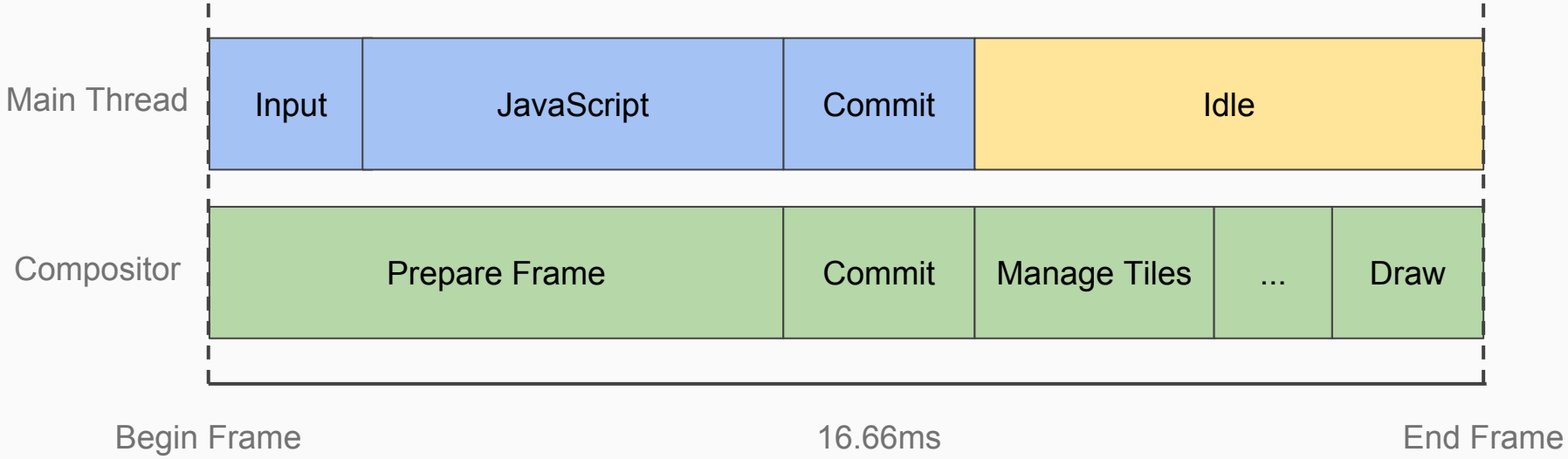




Five Fun Facts About Chrome & V8

- JavaScript is single threaded and runs on the *main thread*
- V8 is the JavaScript virtual machine used within Chrome
- DOM can be modified using JavaScript to perform animation
- The compositor is responsible for drawing DOM changes
- Task scheduler schedules tasks on main thread
  - Provides idle tasks

Chrome & V8


Main Thread
Input
JavaScript
Commit
Idle
Compositor
Prepare Frame
Commit
Manage Tiles
...
Draw
Begin Frame
16.66ms
End Frame

V8 Garbage Collection

- Weak generational hypothesis: “Most objects die young”
- V8 implements a generational garbage collector
  - Young generation (up to 16M)
  - Old generation (up to 1.4G)
- Dynamic allocation site based pretenuring

  Daniel Clifford, Hannes Payer, Michael Stanton, and Ben L. Titzer. 2015. Memento mori: dynamic allocation-site-based optimizations. In Proceedings of the 2015 International Symposium on Memory Management (ISMM '15). ACM, New York, NY, USA, 105-117.

Young Generation Garbage Collection

Cheney-style semi-space scavenger

- Triggers when a semi-spaces becomes full
- Copies live objects to the other semi-space or promotes objects
- Runtime is linear in number of live objects
- Can not be interrupted

Full Garbage Collection

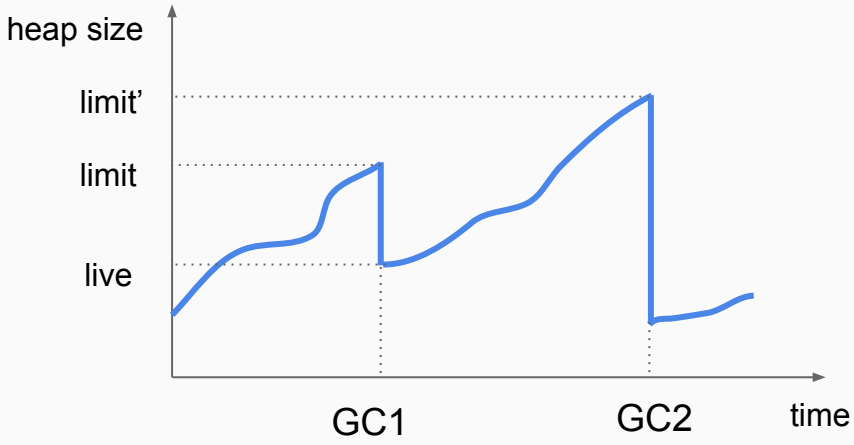
Mark-sweep collector with compaction

- Mark young and old generation
- Incremental marking
- Concurrent sweeping
- Young generation evacuation
- Compaction

Full Garbage Collection

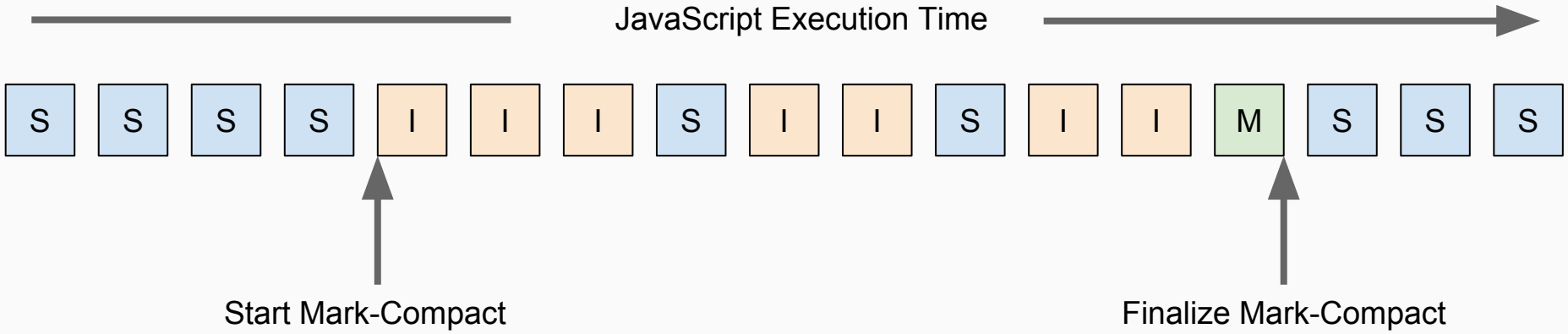
Mark-sweep collector with compaction

- Incremental marking starts close to heap limit determined by heap growing strategy
- limit’ = live object size x [1.1, 4]



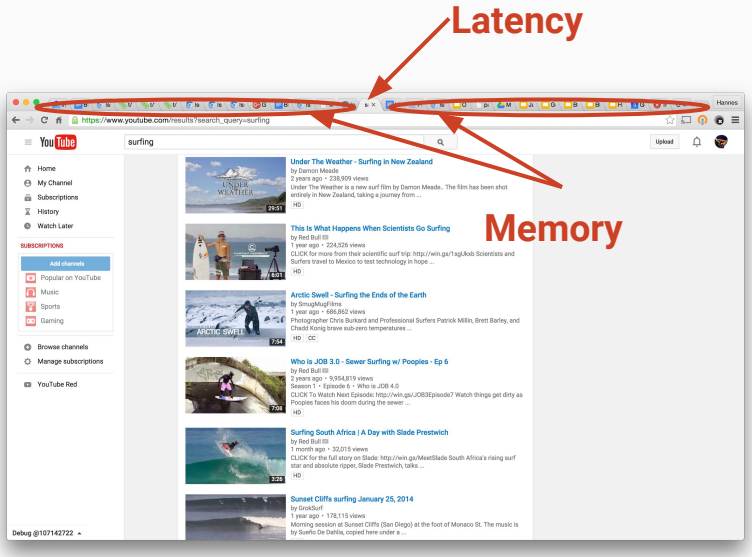
V8 Garbage Collection

- **S** Scavenger (~0-10 ms)
- **I** Incremental Marking (~0.01-CONFIGURABLE ms)
- **M** Finalization Mark-Compact Collection (~4-20 ms)



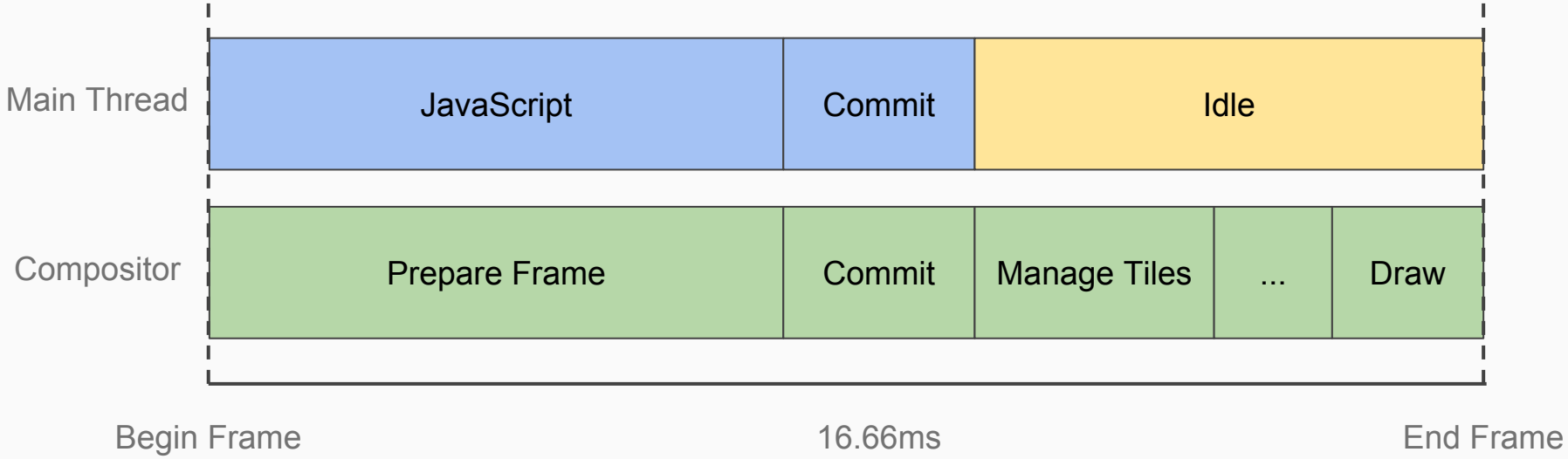
Chrome, a funky place for garbage collection

- Foreground tab
  - Latency matters
  - New frames are drawn every 16.66 ms when animation or scrolling happens
  - Reducing memory becomes important as soon as the tab becomes inactive
- Background tab
  - Reduce memory consumption
  - Latency is secondary

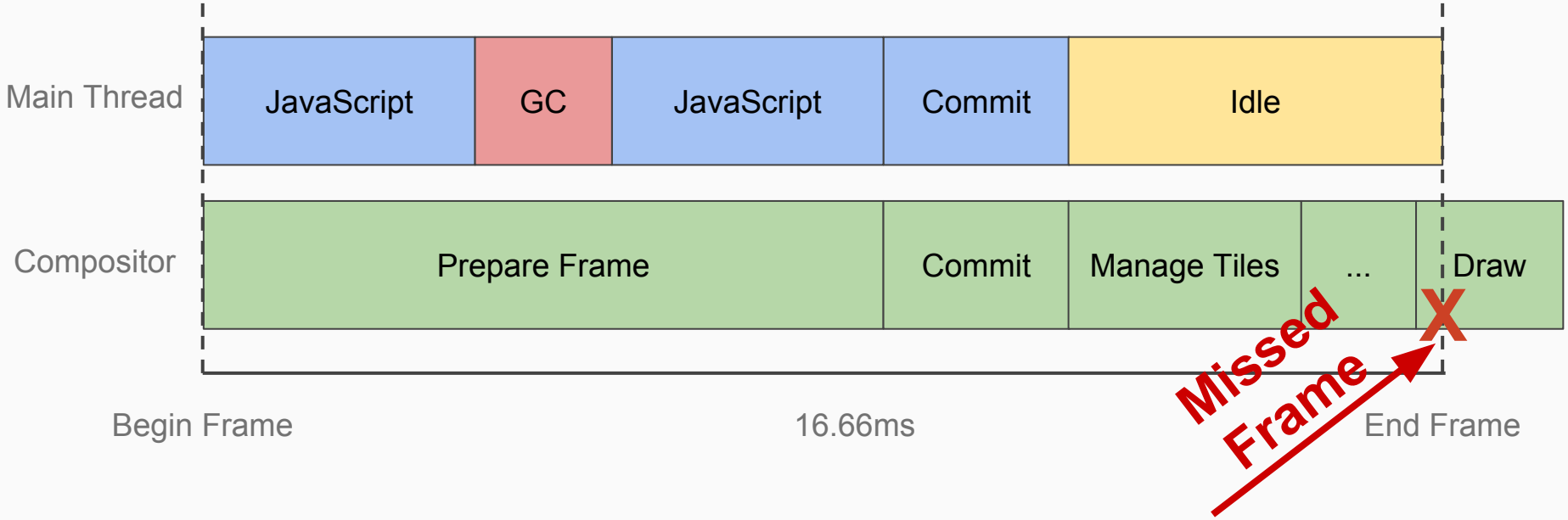


Latency

Life of a Frame

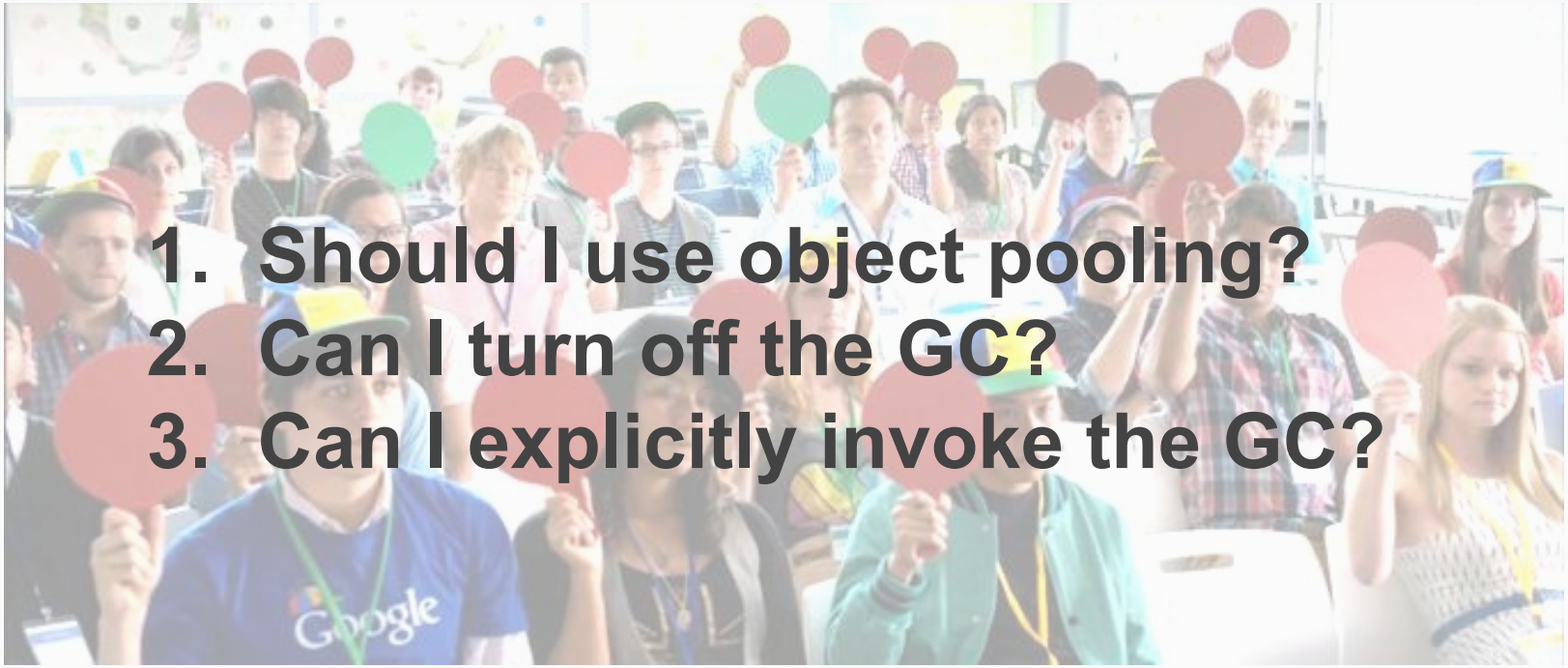


Life of a Frame



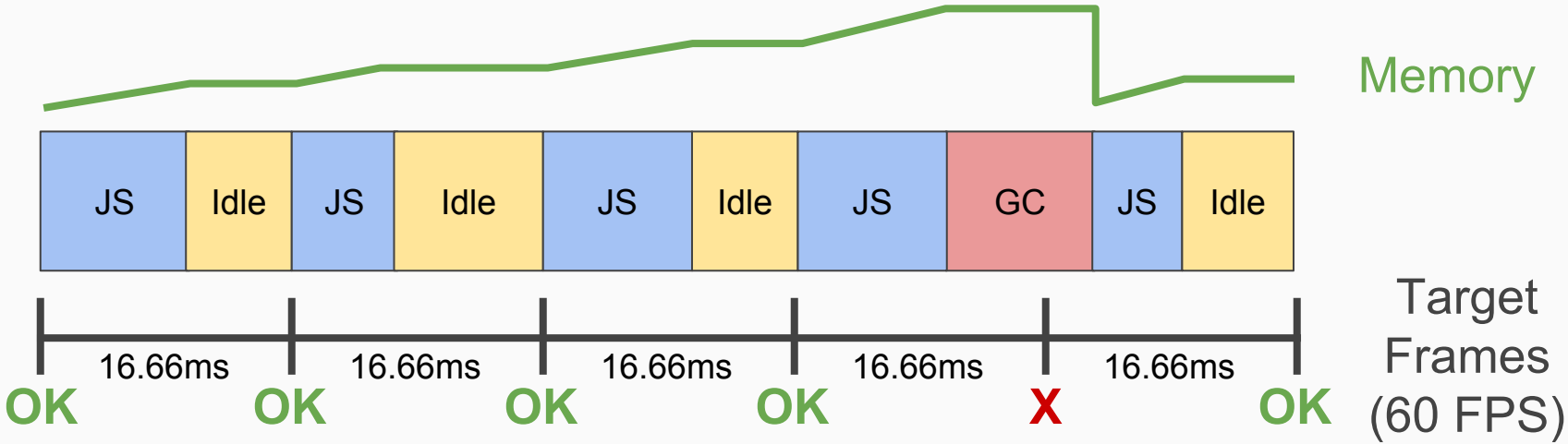
The Three Deadly Sins of Garbage Collection

The Three Deadly Sins of Garbage Collection

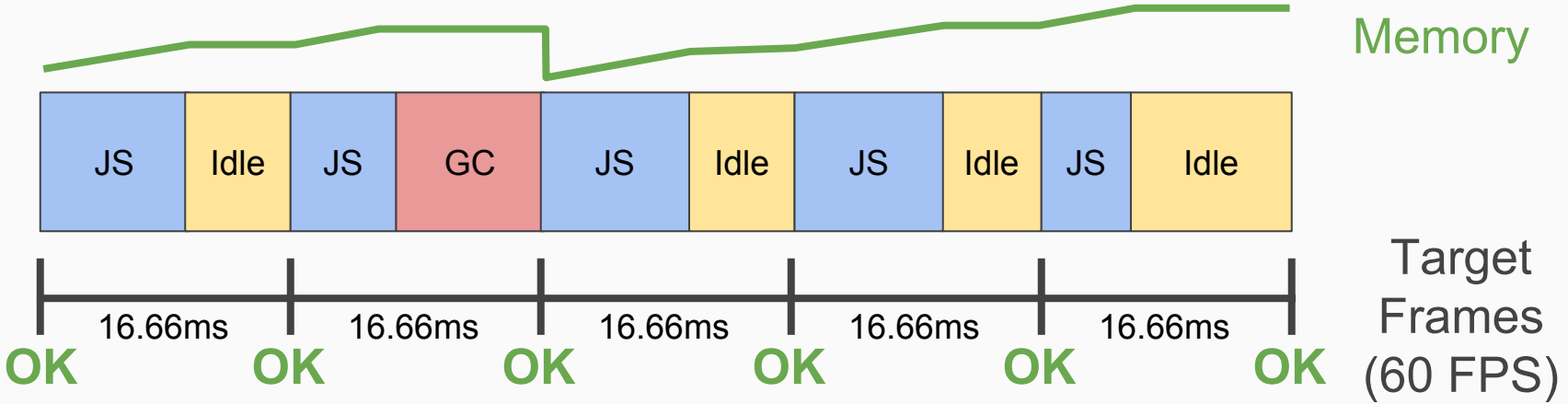


1. Should I use object pooling?
2. Can I turn off the GC?
3. Can I explicitly invoke the GC?

Life of a Frame



Life of a Frame



Idle Time Garbage Collection Scheduling

Step 1

Registering GC Idle Task

- V8 GC checks every *n* allocations or *m* time units if a garbage collection event may happen soon.
- V8 GC registers an idle task for the event in the task scheduler.

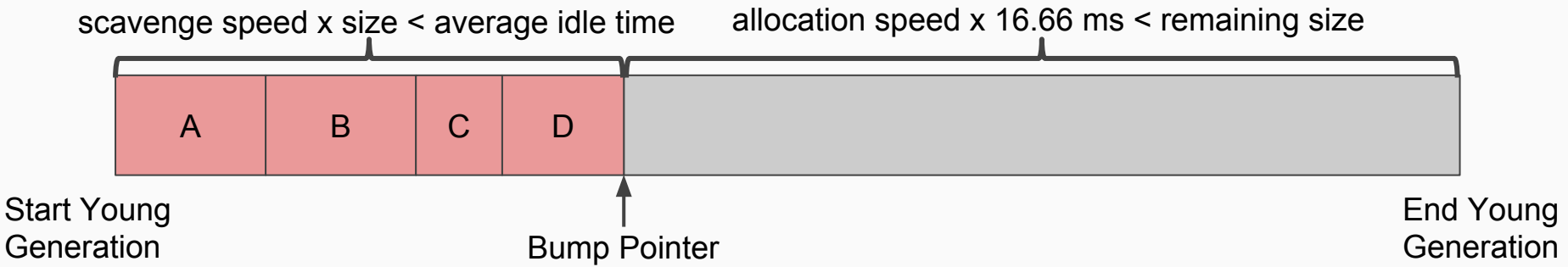
Idle Time Garbage Collection Scheduling

Step 2

Handling GC Idle Task

- The task scheduler will schedule the idle task and invoke the given callback with the available idle time.
- V8 GC will check if the task is still pending and if enough idle time is provided to handle the task.

Scheduling Young Generation GCs



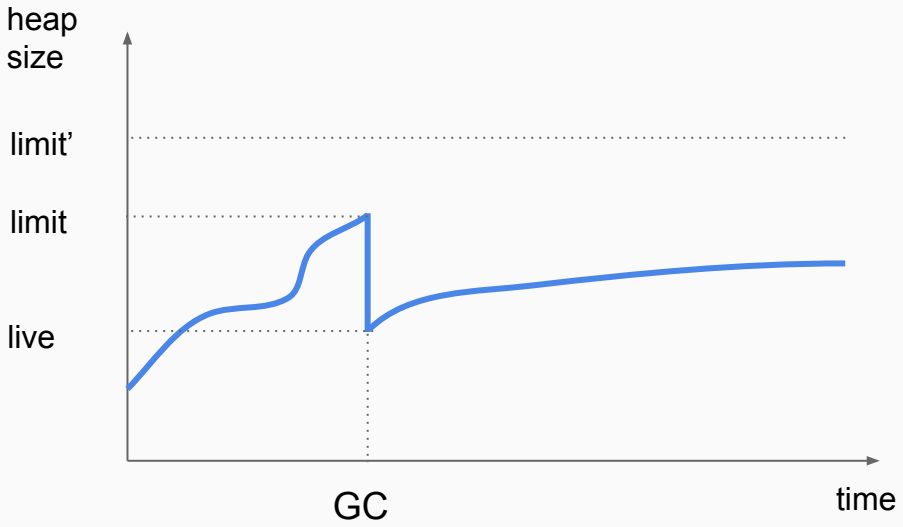
- Allocation interrupt at every 512K of allocations
  - Check if GC will happen within this frame or if there is more time based on average allocation rate and current heap size?
  - Compute how much idle time we have on average and how long young generation garbage collections take on average to proactively schedule garbage collections.

Scheduling Full GCs

- Incremental marking is started close to the old generation limit
- Incremental marking steps
  - Average marking speed in bytes/sec
  - Current idle task deadline
- Finalization of full garbage collection
  - Performed when marking is almost done
  - Average full garbage collection finalization time
  - Current idle task deadline

Memory

Memory Reducing GCs for Inactive Tabs



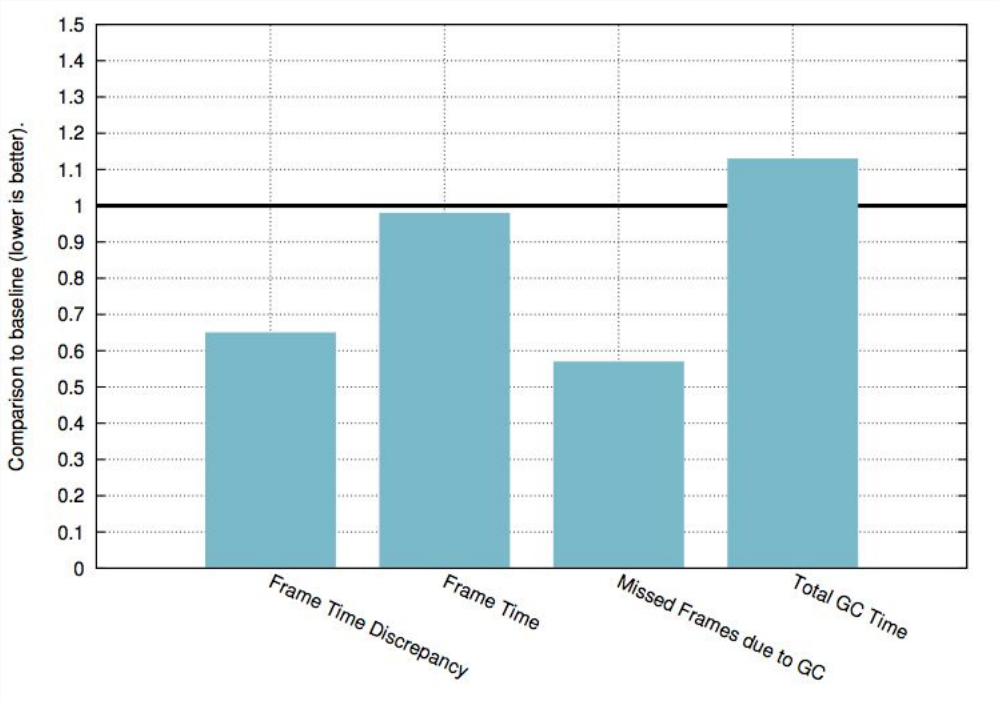
- Started only when JavaScript allocation rate and the rate of JavaScript invocations become low
- Watchdog checks state every 8 seconds
- Register memory reducing idle task
- Idle time up to 50ms is provided by the scheduler

Experiments

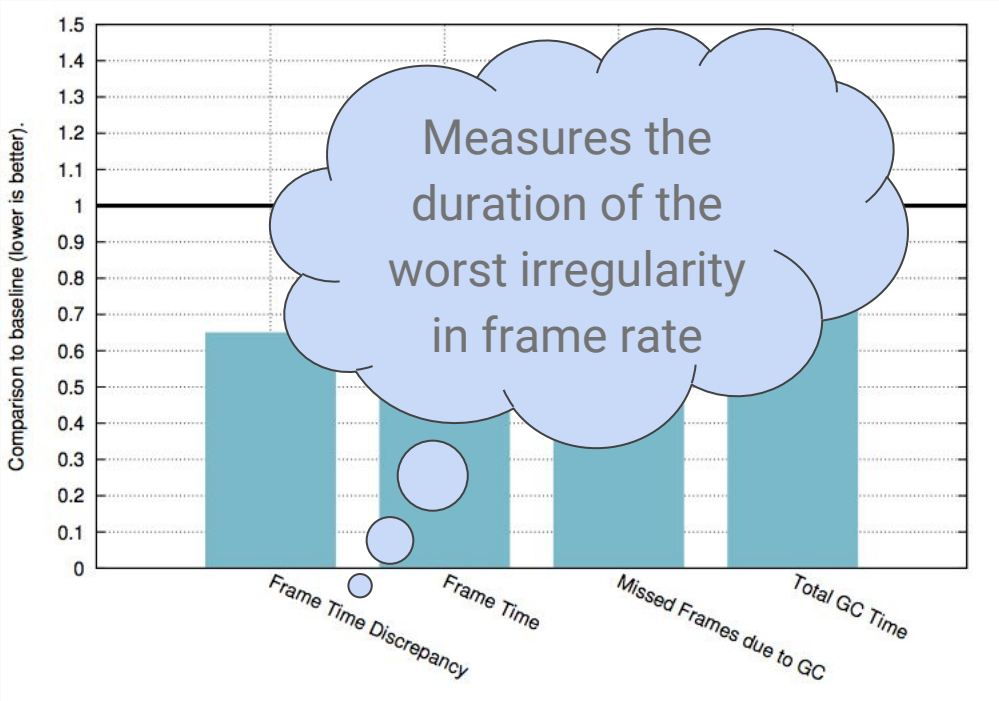
Experiments

- Chrome version 48.0.2564.109 (February 2016)
- Platforms
  - Linux workstation with two Intel Xeon E5-2680 V2 deca-core 2.80 GHz CPUs and 64GB of main memory
  - Nexus 6P Android smartphone with 3GB of main memory and a BIG.little configuration of a Quad-core 1.55 GHz Cortex A53 and a Quad-core 2.0 GHz Cortex-A57.
- Chrome's Telemetry performance benchmarking framework to evaluate recorded samples of real webpages
- Baseline Chrome with *--disable-v8-idle-tasks*
- Each benchmark run was repeated 20 times

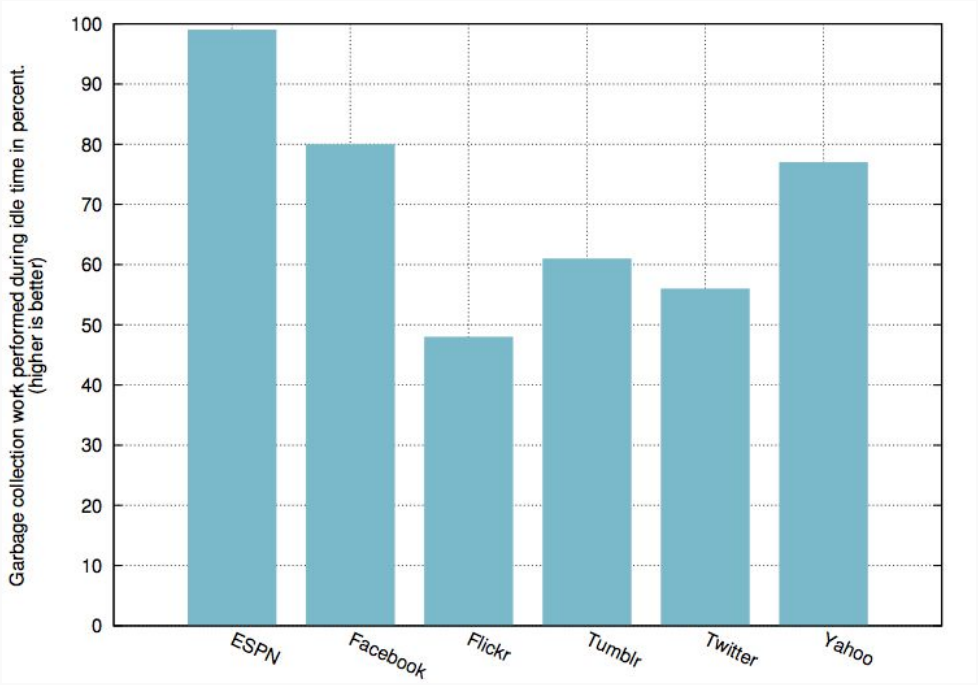
oortonline.gl



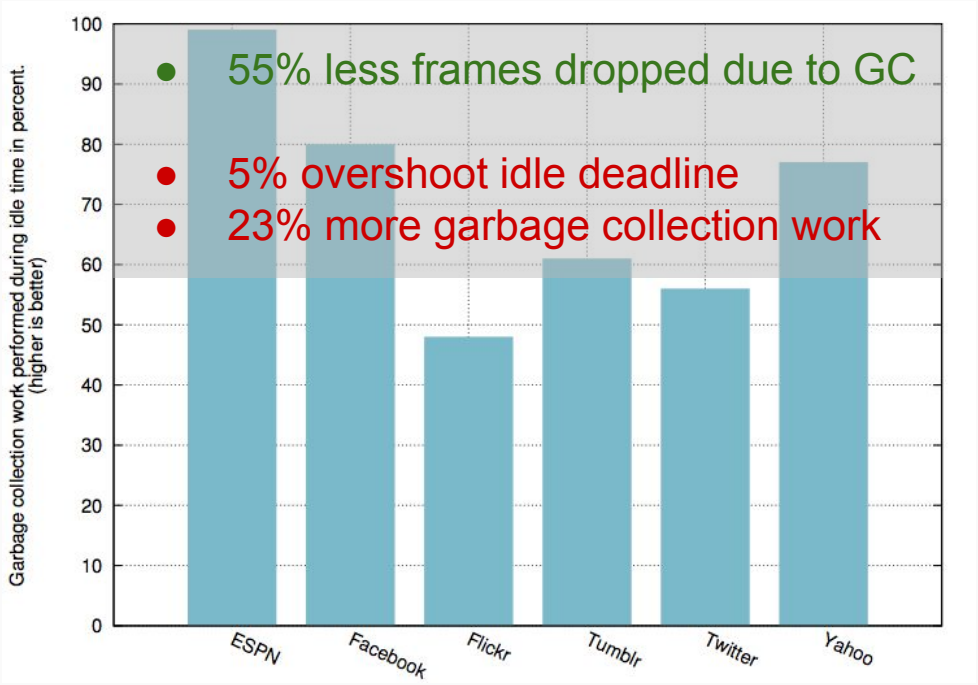
oortonline.gl



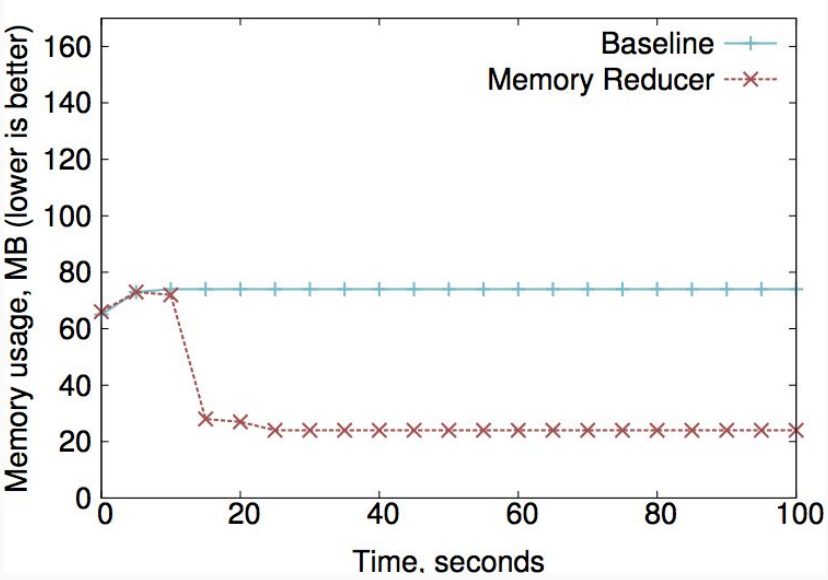
Infinite Scrolling Webpages



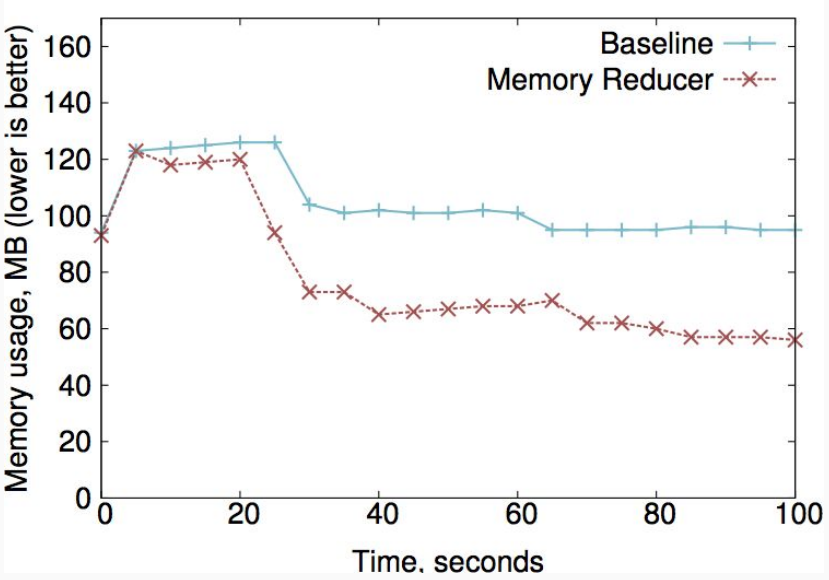
Infinite Scrolling Webpages



Memory Reducer: Idle Webpages



Google Search



Google Docs

Related Work

- Idle time garbage collection
  - Periodic: Metronome Bacon et. al POPL'03
  - Slack-based: Henriksson et. al LCTES'03
  - Hybrid: Metronome-TS Auerbach et. al EMSOFT'08
  - Kalibera et. al TOCS'11 found that hybrid systems are superior
- Concurrent, parallel, and incremental garbage collection
  - Can be combined with idle time garbage collection scheduling
  - Costly memory compaction phases can be hidden during idle time without introducing memory or barrier overhead

Conclusions

- Idle time garbage collection scheduling may
  - improve responsiveness and
  - decrease memory consumption of webpages
  - without introducing additional garbage collection implementation complexity.
- Shipped and enabled in Chrome by default
- New metric *frame time discrepancy* to better quantify user experience
- Applicability:
  - Other virtual machines that are aware of screen rendering
  - Servers that are not constantly under 100% CPU load, e.g. node.js

Thank you!
Questions?

+Hannes Payer
hpayer@google.com
http://research.google.com/pubs/HannesPayer.html

Google