ANALYZING LATENCY OF I/O EVENTS

ARCHIT SHARMA

ASSOCIATE PERFORMANCE ENGINEER

BLR | Red Hat India Pvt. Ltd.

THINGS WE'RE GONNA TALK ABOUT

- An I/O use case
- The investigation:
  - Block I/O events
    - native vs. threads in Qemu-KVM
- IOPS performance benchmarking/debugging
  - General approaches
- Tools/utilities we've rolled out:
  - includes benchmarking IOPS
  - postprocessing that data
- Applicability of Latency analysis

redhat.

USE CASE

I/O EVENTS IN QEMU-KVM

- Whether the delay is being produced by filesystem / kvm layer?
- IO engines: How does async compare to sync ?
  - How does a setup with target:threads compare to one with target:native for a kernel version?
- Would I achieve better results if I changed iodepth?
- Block I/O and File I/O

BLOCK I/O EVENTS IN QEMU-KVM

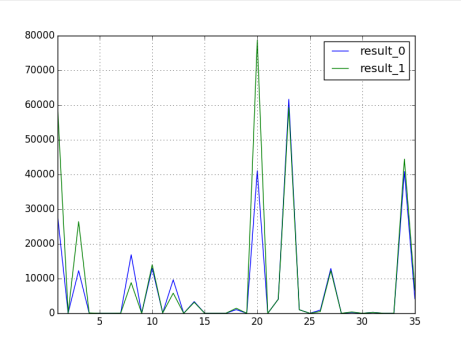
- An investigation of blockIO events: tracing and analyzing them
- Came up with a couple of utilities to help analyze I/O latency..

[Native]

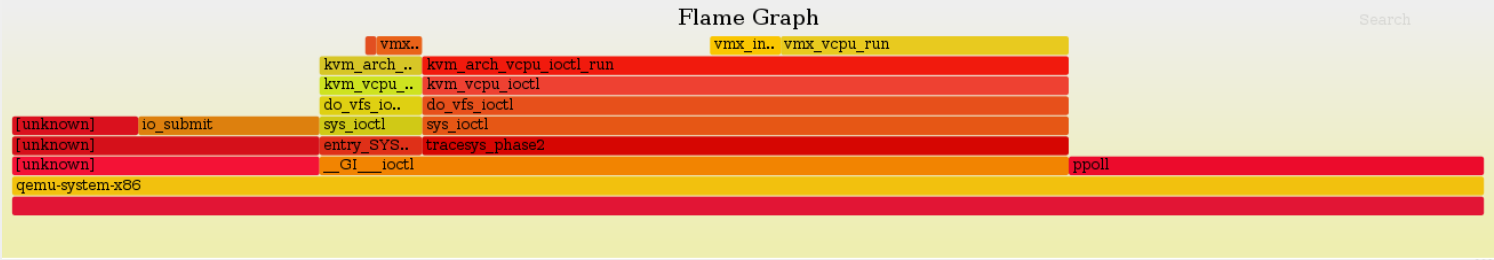
```
kvm_exit -> sys_exit_ppoll -> sys_enter_io_submit -> sys_exit_io_submit ..
.. -> sys_enter_io_getevents -> sys_exit_io_getevents
```

GENERAL APPROACHES

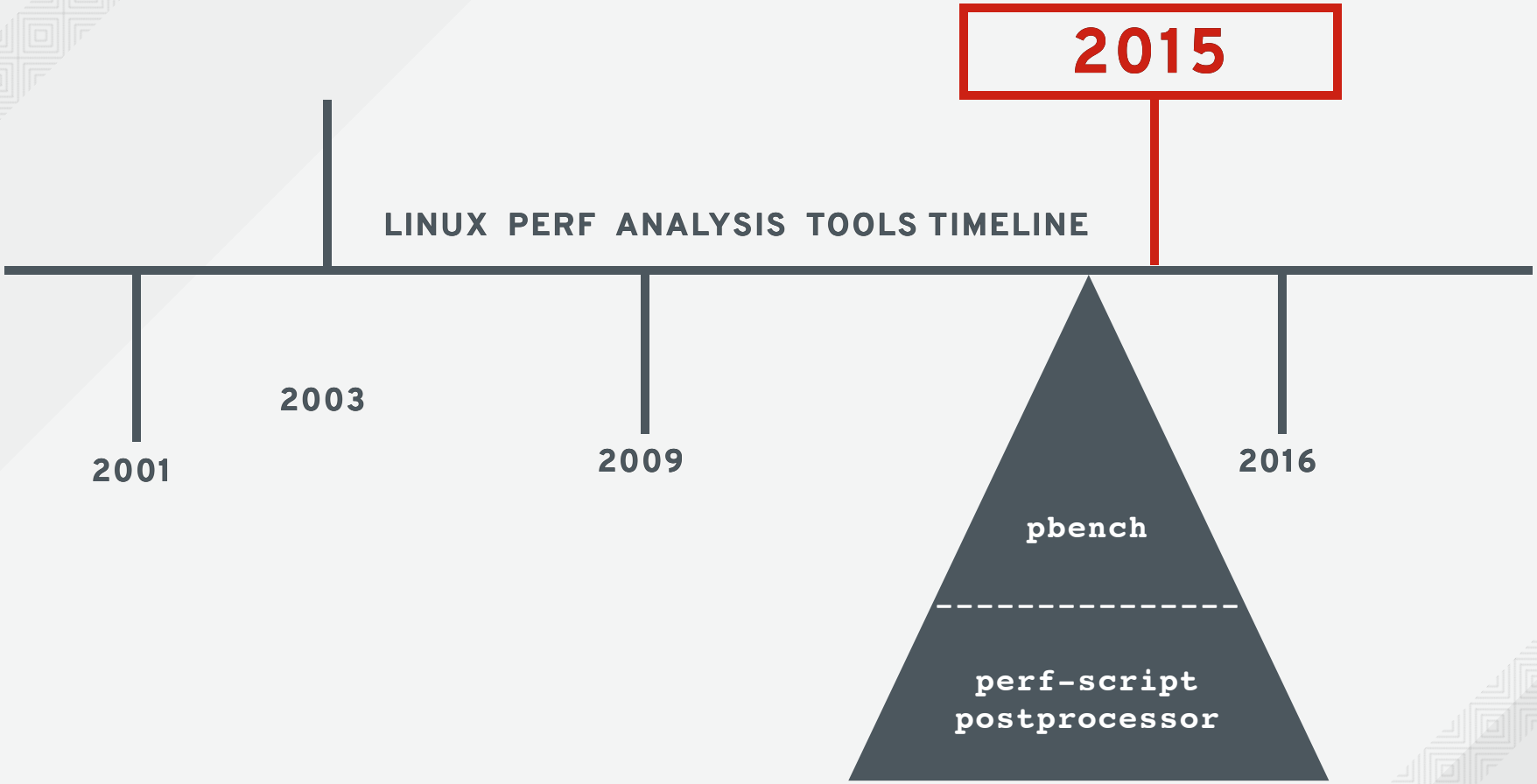
IOPS PERFORMANCE BENCHMARKING/DEBUGGING



- IOPS Benchmarking - FIO
  - Our addon: pbench_fio



- Debugging: Widely used perf-tools
  - Our addon: I/O Event loop latency processor


2015
LINUX PERF ANALYSIS TOOLS TIMELINE
2001
2003
2009
2016
pbench
perf-script
postprocessor

PBENCH

A Benchmarking and Performance Analysis Framework

http://distributed-system-analysis.github.io/pbench/

- Allows commonly used / even custom benchmarking scripts!
- Dynamic visualizations enabling hands-on exploration and deeper insights into potential bottleneck regions
- Easy to use and setup
- Exciting upcoming features..
- Open for contributions!

PBENCH

A Benchmarking and Performance Analysis Framework

http://distributed-system-analysis.github.io/pbench/

1. A collection agent (pbench-agent) -> Handles TLC - Telemetry, Logs and Configurations

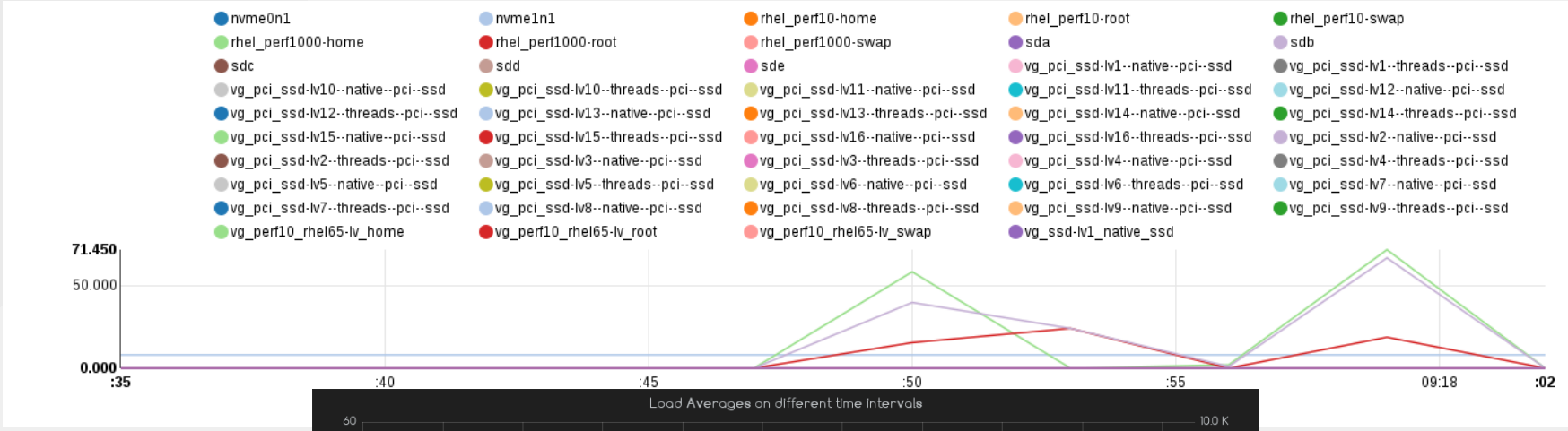
2. Background tasks (bgtasks) -> Archives result tar balls, indexes them, and unpacks them for display.

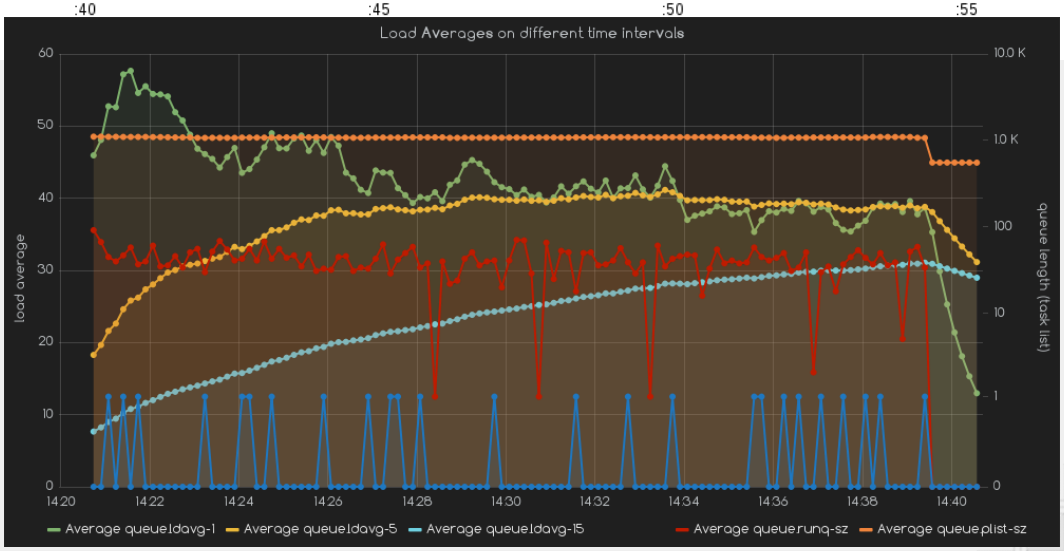
3. Web server -> display various graphs and results

PBENCH

A Benchmarking and Performance Analysis Framework

http://distributed-system-analysis.github.io/pbench/





redhat.

PERF SCRIPT POSTPROCESSOR

A DEBUGGING TOOL

Github: arcolife/perf-script-postprocessor

- Hands-on tracing with flexible approach
  - specify your own event loops!
  - Lots of use cases - disk I/O, network I/O, ..
- A statistical, descriptive and visual approach to latency analysis
- Available on pypi!
  - $ pip install perf-script-postprocessor

PERF SCRIPT POSTPROCESSOR

A DEBUGGING TOOL

(PERF TOOLS) - $ PERF KVM RECORD

GENERATES BINARY DATA FILE

PERF.DATA

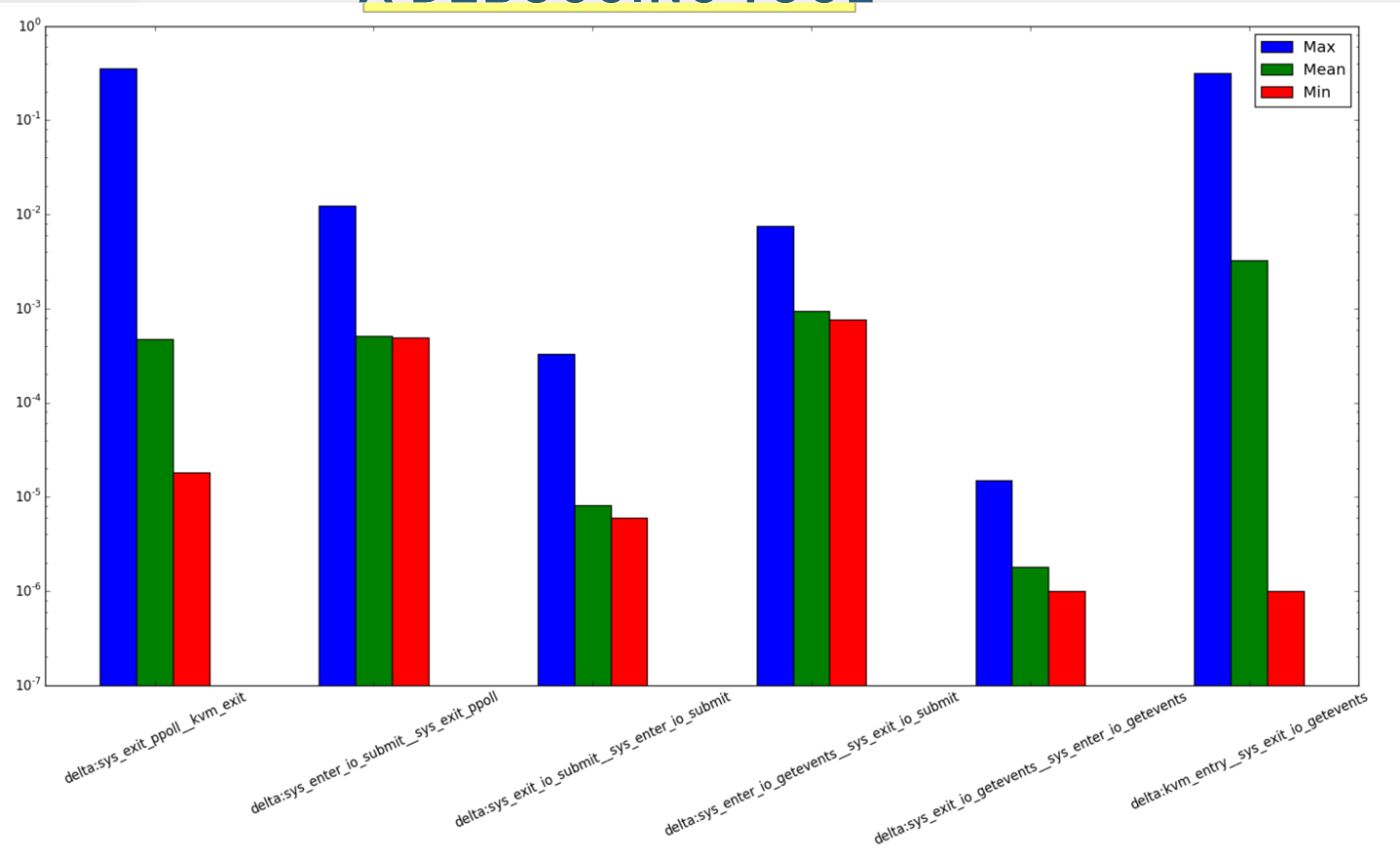
$ PERF_SCRIPT_PROCESSOR

{MEAN, MEDIAN, STD_DEVIATION}
EVENT LOOP LATENCIES

redhat.

PERF SCRIPT POSTPROCESSOR

A DEBUGGING TOOL



redhat

ADDITIONAL UTILS

KVM_IO - BENCH_ITER.SH

Example Results Layout

```
[root@perf results]# ls
1/  2/  3/  4/  5/
perf_record_.txt
perf_kvm_record_.txt
perf_trace_.txt
strace_.txt

[root@perf results]# ls 1/
output_perf_trace
output_strace
perf_record.data
perf_kvm_record.data
results_1_perf_record_
results_1_perf_trace_
results_1_perf_trace_record_
results_1_strace_

[root@perf results]# cat perf_record_.txt
Min: 160756.05
Max: 177846.30
Avg: 170572.8880
Std Dev %: 3.7418
```

ADDITIONAL UTILS

LATENCY_ANALYZER

`Github: arcolife/latency_analyzer`

“ swiss knife for getting started with [native] [file I/O] latency analysis [for Qemu-KVM]

- Chewbacca

“ I love this script!

- Luke Skywalker

“ pfft..Whatever

- Darth Vader

redhat.

WHY ANALYZE LATENCY ?

- Code Optimization
  - eg: OS profiling
- Distributed Computing
  - latency distributions
- Cache tuning
  - distributed cache performance
  - (timed cache access)^N
- Web Performance
  - high latency may involve:
    - Load Balancing
    - Network Latency
    - Web server configuration
- Performance Engineering (throughput & latency)
  - Databases
    - recommended I/O schedulers
    - memory / caching
  - Virtualization
    - Block and File I/O
  - Networking
    - Network I/O
- ..

FOOD FOR THOUGHT?

1. how much time spent on each event, WHILE control is in user/kernel space
2. Sorting out anomalies: IOPS throughput different with strace, perf record .. At the same time, nr values should be long (they're not when using perf record).
3. .. ?

THANKS!!

- Twitter: @arcolife
- Website: http://work.arcolife.in/
- LinkedIn: https://www.linkedin.com/in/arcolife