

MESURE ET ANALYSE DE PERFORMANCE D'ENCODAGE VIDÉO TEMPS RÉEL

avec Dash et Plotly

Kevin Pouget, Red Hat SPICE Senior Software Engineer 2020-02-18 – Python User Group @ La Turbine, Grenoble

ABOUT ME

...

Kevin Pouget

- 2011-2017: PhD & Post-doc (Grenoble)
 - Improve interactive debugging for multicore programming
- 2017-2019: Virtualization engineer (Grenoble)
 - OpenCL inside virtual machines
 - Qemu/KVM checkpoint restart

• June 2019...: Senior Software Engineer (remote)









ABOUT ME

Kevin Pouget

- 2011-2017: PhD & Post-doc (Grenoble)
 - Improve interactive debugging for multicore programming
 - Python to extend GDB and inspect C code
- 2017-2019: Virtualization engineer (Grenoble)
 - OpenCL inside virtual machines
 - Python for C code generation, benchmarking & performance analyses
 - Qemu/KVM checkpoint restart
 - Python for benchmarking & performance analyses
 - ...
- June 2019...: Senior Software Engineer (remote)



 \Rightarrow Python for driving C low-level apps and collection/analyze of execution data

...in addition to multi-process low-level C programming... !









SPICE Remote Display

Recording Infrastructure

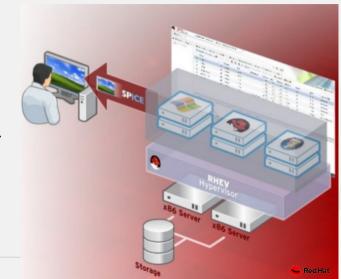
Performance Benchmarking & Visualization





Simple Protocol for Independent Computing Environments

- GUI in remote client
 - ★ client/VM copy-and-paste
 - ★ file drag-and-drop
 - ★ multi-monitor support
 - * window-to-screen resolution
 - * CD-ROM/ISO sharing
 - ★ seamless VM host migration
 - * SSL, smartcard authentication, ...
- Server library inside Qemu-KVM
 - * OS-independent screen capture
 - hypervisor level
 - * MJPEG video encoding



3D, vGPU and Video Streaming

3D inside the VM?

A https://webglsamples.org/dynamic-cubemap/dynamic-cubemap.html

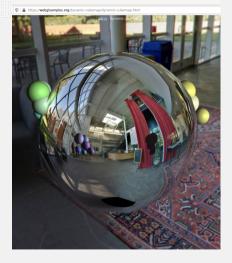


3D, vGPU and Video Streaming

3D inside the VM?

GPU passthrough?

• disconnect from the host, give it to the VM





3D, vGPU and Video Streaming

3D inside the VM?

https://webglsamples.org/dynamic-cubemap/dynamic-cubemap.html



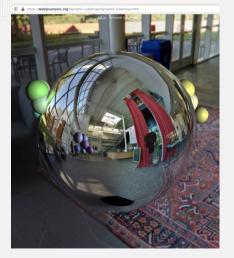
GPU passthrough mediated passthrough

- disconnect from the host, give it to the VM
- cut the GPU in 4, give a vGPU slice to the VM
 - recent Intel/NVidia/... GPUs



3D, vGPU and Video Streaming

3D inside the VM?

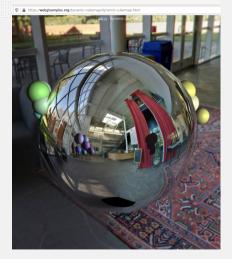


GPU passthrough mediated passthrough

- disconnect from the host, give it to the VM
- cut the GPU in 4, give a vGPU slice to the VM
 - recent Intel/NVidia/... GPUs
- forward 3D commands to the host (VirGL)
 - under investigation for SPICE



3D, vGPU and Video Streaming



3D inside the VM?

GPU passthrough mediated passthrough

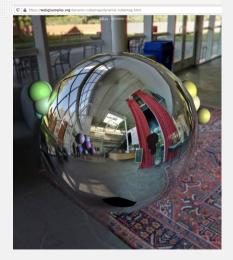
- disconnect from the host, give it to the VM
- cut the GPU in 4, give a vGPU slice to the VM
 - recent Intel/NVidia/... GPUs
- forward 3D commands to the host (VirGL)
 - under investigation for SPICE

but what about the video encoding?

• MJPEG OK for office usage, but not for 3D!



3D, vGPU and Video Streaming



3D inside the VM?

GPU passthrough mediated passthrough

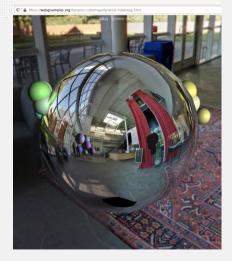
- disconnect from the host, give it to the VM
- cut the GPU in 4, give a vGPU slice to the VM
 - recent Intel/NVidia/... GPUs
- forward 3D commands to the host (VirGL)
 - under investigation for SPICE

but what about the video encoding?

- MJPEG OK for office usage, but not for 3D!
- H264, VP8 more suitable (<u>++ with HW-support</u> :)



3D, vGPU and Video Streaming



3D inside the VM?

GPU passthrough mediated passthrough

- disconnect from the host, give it to the VM
- cut the GPU in 4, give a vGPU slice to the VM
 - recent Intel/NVidia/... GPUs
- forward 3D commands to the host (VirGL)
 - under investigation for SPICE

but what about the video encoding?

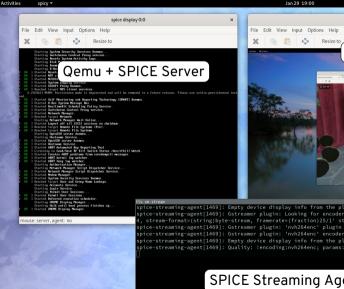
- MJPEG OK for office usage, but not for 3D!
- H264, VP8 more suitable (<u>++ with HW-support</u> :)

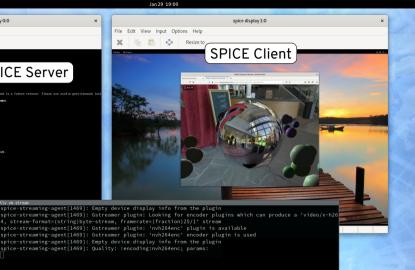
Video encoding in the guest vGPU:

- capture the framebuffer (*ie*, do a screenshot)
- 2 give it to the vGPU for encoding
- ③ send it (1) outside of the VM and (2) to the client



File Edit View Input Options Help

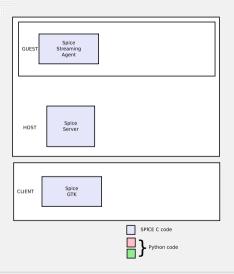




よ (1) () -

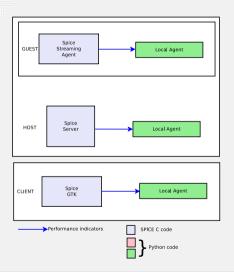
SPICE Streaming Agent





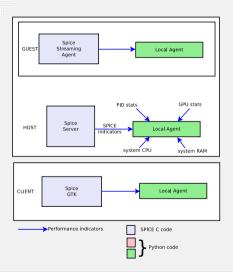
• SPICE is running and streaming





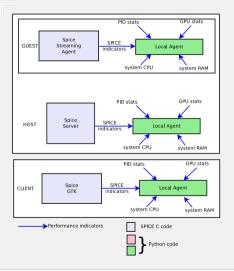
- SPICE is running and streaming
- On each system,
 - deploy a Python local agent



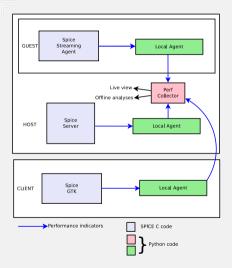


- SPICE is running and streaming
- On each system,
 - deploy a Python local agent
 - 2 collect various performance indicators





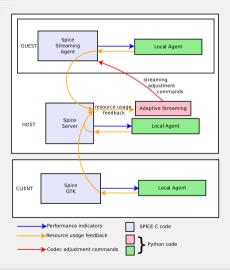
- SPICE is running and streaming
- On each system,
 - deploy a Python local agent
 - 2 collect various performance indicators
 - 3 ... in each of the agents



- SPICE is running and streaming
- On each system,
 - deploy a Python local agent
 - 2 collect various performance indicators
 - 3 ... in each of the agents
- Collect all the indicators
 - for live & interactive view
 - for offline processing

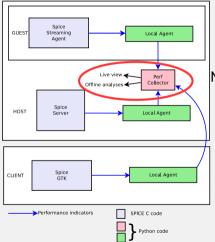


| ttv qemu | ttv ctient | LLV VN SLIEUN | keyingpinea: ~ |
|---|---|---|--|
| <pre>qemu-system-x86_64: warning: Spice: Agent Inter Starting CUPS Scheduler [0K] Reached target Remote File Systems. Starting Permit User Sessions Starting Permit User Sessions Starting Network Nanager Wait Online Starting Network Nanager Wait Online [0K] Started Permit User Sessions. [0K] Started Permit User Sessions. [0K] Starting Not Network Scheduler. [0K] Starting Not Network Scheduler. [0K] Started Command Scheduler. [0K] Started Command Scheduler. [0K] Started Command Scheduler. [0K] Starting Not Network Scheduler. [0K] Started COMED Display Manager [0K] Started COMED Display Manager. [0K] Started COMED Display Manager. Starting Hostname Service Federa 30 (Workstation Edition) Kernel 5.3.15-200.fc30.x86_64 on an x86_64 (tty S0) chamechaude login: []</pre> | Leptop Laptop Stlv Client CLIENT: SPICE_NOGRAB=1 /home/kevin/client GSpice-Hessage: 18:25:34.798: main channel: ope (spicy:22238): Spice-WARNING **: 18:25:41.973: | <pre>'video/x-h264, stream-format=(string)byte-stre am, framerate=(fraction)25/1' stream spice-streaming-agent[1474]: Gstreamer plugin: 'nvh264enc' plugin is available spice-streaming-agent[1474]: Gstreamer plugin: 'nvh264enc' encoder plugin is used</pre> | <pre>> code=nv.plug.h264_displaywebgl=cubemap_reco nd+time=485_tramerate=200_ratecontrol_cbr_max-b itrate=64_gop=0_blocking=0 reset encoder parameters set_enc:nv.plug.h264 framerate=200, nv.ratecon trol=cbr, nv.max-bitrate=64, nv.gop=0, nv.block ing=0 wait2 seconds clear graphs clear quality clear quality append to quality: script: rodec: nv.plug.h264 append to quality: script: record_time: 40s append to quality: script: 40s append to quality: script: 40s append to quality: script: 40s append to quality: script: 40s append to quality: 40s</pre> |
| * Starting the socket for the Perf Collector on | <pre>WARRING: Intel_GPU_Top: failed to load (intel_g 'intel_gpu_top': 'intel_gpu_top' * Preparing the environment * Running! Connecting to 1234 Client: received 5 recorders: frames_info, qual _info</pre> | * Starting the socket for the Perf Collector on WARNING: Intel_GPU Top: failed to load (intel_g o 2) No such file or directory: 'intel_gpu_top' * Preparing the environment * Running! Connecting to 1236 guest: received 6 recorders: frame, streaming_i | <pre>save record into /home/kevin/spice/streaming_st ats/recuits/cubenp/framerate=200 _ratecontrol=cbr_max-bitrate=64_gop=0_blocking= 0.rec write result: /home/kevin/spice/streaming_stats /reults/current/matrix-cov < code=cnv.plug.h2 64_display-webgl=cubensp_record=time=40s webg Licbensp framerate=200 ratecontrol=cbr_max-b itrate=64_gop=0_blocking=0 teradom(display) system=exec: wm display stop https://webglsampl teardom() system=exec: set s on +dpms System=exec: set s</pre> |
| | | | 🤐 IRad Haft |



- SPICE is running and streaming
- On each system,
 - deploy a Python local agent
 - 2 collect various performance indicators
 - 3 ... in each of the agents
- Collect all the indicators
 - for live & interactive view
 - for offline processing
- - collect resource-usage feedback
 - adapt streaming quality accordingly





Next: benchmarking and visualization tool



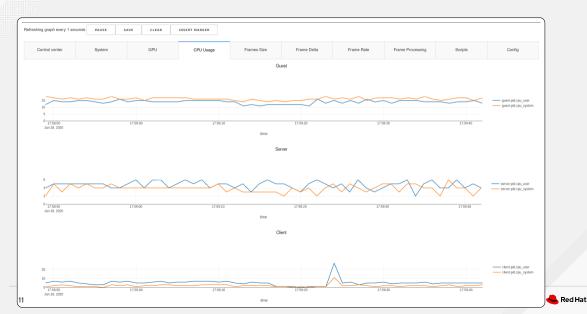
PERFORMANCE BENCHMARKING & VISUALIZATION



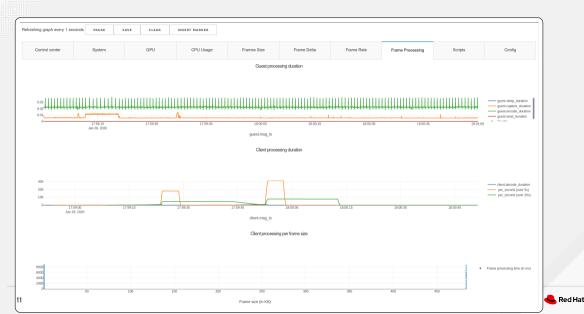
LIVE CONTROL AND VISUALIZATION

| freshing graph every 1 seconds | PAUSE | SAVE | CLEAR | INSERT MA | RKER | | | | | | |
|--|------------------|-----------|---------|-----------|---------------|------------------------------------|---------------------|-------------|---------------------|---------|--------------|
| Control center | System | | GPU | CP | U Usage | Frames Size | Frame Delta | Frame Rate | Frame Processing | Scripts | Config |
| Qua | lity Mess | ages | | | | | | Video E | ncoding | | |
| Enter a quality message | SENDI | CLEAR | REFRESH | | | gst.vp8.vp8enc | gst.vp8.vaar | vivp8enc | gst.vp9.vaapivp9enc | | nv.plug.h264 |
| Refreshi | ng quality every | 0 seconds | | | | | | | | | |
| | | | | 30s | | | | GOI | RESET to: 30 | | |
| uest: encoder:nv.plug.h264 uest: encoder:gst.h264.vaapih24 uest: encoder:gst.vp8.vaapivp86 | | | | | 0 | | 0 | bitra | to <u>'</u> | | |
| uest: lencoding:vaapivp8enc; pa | | | | | Enter a nume | eric value for "bitrate" default | : 0 | | | | |
| | | | | | | | | default-roi | delta-qp_ | | |
| | | | | | Enter a nume | ric value for "default-roi-delta | -qp* default; -10 | | | | |
| | | | | | | | | keyframe | period_ | | |
| | | | | | Enter a nume | eric value for "keyframe-period | f" default: 30 | | | | |
| | | | | | | | | quality | level_ | | |
| | | | | | Enter a nume | ric value for "quality-level" d | efault: 4 | | | | |
| | | | | | | | | rate-co | ntrol_ | | |
| | | | | | Enter a value | for "rate-control" default: cq | p | | | | |
| | | | | | | | | sharpne | s-level_ | | |
| | | | | | Enter a nume | eric value for "sharpness-level | " default: 0 | | | | |
| | | | | | | for "encoder-reload" | | encoder | reload | | |
| | | | | | Enter a value | IOL AUCODEL-LEIOBQ. | | cust | | | |
| | | | | | Enter a value | | | cusi | ani - | | |

LIVE CONTROL AND VISUALIZATION



LIVE CONTROL AND VISUALIZATION



Live view and manual control is nice, but ...



We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60 \text{ FPS}$
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period \Rightarrow *auto*, 10, 60, *none*

We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60$ FPS
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period \Rightarrow *auto*, 10, 60, *none*



We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60 \text{ FPS}$
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period \Rightarrow *auto*, 10, 60, *none*

• over multiple resources

• CPU/GPU on the host/guest/client, frame sizes, framerate, ...





We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60$ FPS
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period \Rightarrow *auto*, 10, 60, *none*
- over multiple resources
 - CPU/GPU on the host/guest/client, frame sizes, framerate, ...

| 4 | \times | 3 | \times | 4 | \times | 3 | = | 144 |
|---|----------|----------|----------|---|----------|----------|---|-----|
|---|----------|----------|----------|---|----------|----------|---|-----|



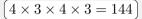


We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60 \text{ FPS}$
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period \Rightarrow *auto*, 10, 60, *none*

• over multiple resources

- CPU/GPU on the host/guest/client, frame sizes, framerate, ...
- with different VM use-cases
 - office work, 3D processing, still desktop, ...



> 8 graphs

 $... \times 3 = 432$



We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60 \text{ FPS}$
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period \Rightarrow *auto*, 10, 60, *none*

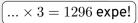
• over multiple resources

- CPU/GPU on the host/guest/client, frame sizes, framerate, ...
- with different VM use-cases
 - office work, 3D processing, still desktop, ...
- with different resources usage
 - VM CPU busy, client GPU busy, slow network, ...

$$4 \times 3 \times 4 \times 3 = 144$$

$$(>8 \text{ graphs})$$

$$\dots \times 3 = 432$$





We want to understand the impact ...

- of encoding parameters
 - bitrate $\Rightarrow 0.5, 1, 10, 100 \text{ MB/s}$
 - framerate $\Rightarrow 20, 40, 60 \text{ FPS}$
 - resolution $\Rightarrow 1K, 2K, 4K$
 - keyframe-period $\Rightarrow \textit{auto}, 10, 60, \textit{none}$

• over multiple resources

- CPU/GPU on the host/guest/client, frame sizes, framerate, ...
- with different VM use-cases
 - office work, 3D processing, still desktop, ...
- with different resources usage
 - VM CPU busy, client GPU busy, slow network, ...

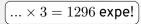
Script all of it! And make it reproducible!

 $4 \times 3 \times 4 \times 3 = 144$

> 8 graphs

 $... \times 3 = 432$

Red Hat



SPICE RECORDING INFRASTRUCTURE

| Refreshing graph every 1 seconds PAUSE SAVE CLEAR II | NSERT MARKER | | | | | |
|--|---|---------------|---------------|---|--------|--|
| Control center System GPU | CPU Usage Frames Size | Scripts | Config | | | |
| Test | t-case scripts TEST RUN | REFRESH CLEAR | RELOAD SCRIPT | s | | |
| Estimated time: 162mm0s Running mathr. (4yr) setuac) system-aece: vine solution 1920x1080 system-aece: setual of d-dpms system-aece: setual of d-dpms Loading previous matrix results Loading previous matrix results Loading previous matrix results Estimated termines setup(d-dpm), Img., 180/, 1920 setup(results), 180/, 720 setup(results), 180/, 720 setup(results), 180/, 720 ummong, 1216 = code-setup, 180/, 180/, 720 | name: main's record, time: 40 code: gat ypd: xapyoptime • bitnati ~ bitcoi 1,2000 • rate control - citr • keyframe parked - 0 [2] 30 • stanetati ~ 20 [4] 10 • stapped_properties: • dialogi - ritos, tay 1,202 (19 scienced) | imple-test | 197 | | matrix | |
| erd_mc:gitxtpl:vage/pelice.git.prg=hates=8000, git.prg=-rate=control=cbc, git.prg=-keyfine_end=d_f.tramerata=20 wall 2 seconds clear graphs clear graphs clear graphs append to galify:scrpt:code; git.ybl.vage/pelice append to galify:scrpt:code_mc:e.40s append to galify:scrpt:code_mc:e.40s append to galify:scrpt:code_mc:git.ybl.ybl.git.200 append to galify:scrpt:code_mc:git.ybl.git.200 append to galify:scrpt:code_mc:git.ybl.git.200 append to galify:scrpt:code_mc:git.ybl.git.200 append to galify:scrpt:code_git.bl.git.200 append to galify:scrpt:scrpt:code_mc:git.bl.git.200 append to galify:scrpt:scrpt:sclpt:sclpt:sclpt.git.scl_200; append to galify:scrpt:scl_200; bl.git.scl_200; scl | SCRIPTS | | | | | |

Red Hat

SPICE RECORDING INFRASTRUCTURE

Scripting and Matrix Benchmarking

```
record_time: 40s
codec: gst.vp8.vaapivp8enc
matrix:
```

```
framerate: 20, 40, 60
gst.prop=bitrate: 8000, 16000, 32000
gst.prop=keyframe-period: 0, 2, 30
```

scripted_properties:

```
display:
    img_lady_1920:
    vlc_wipeout:
    webgl_wipeout: https://phoboslab.org/wipeout/
    webgl_cubemap: https://webglsamples.org/dynamic-cubemap/dynam
resolution: 1280x720, 1920x1080
```



SPICE RECORDING INFRASTRUCTURE

Scripting and Matrix Benchmarking

scripts :

setup:

- vm resolution 1920x1080; server xset s off -dpms **teardown**:

- xset s on +dpms; server xset s on +dpms resolution:

before:

- vm exec xrandr --output DP-2 --mode \$resolution after:

- vm exec xrandr --output DP-2 --mode 1920x1200 display:

before :

- vm display start \$display

after:

- vm display stop \$display



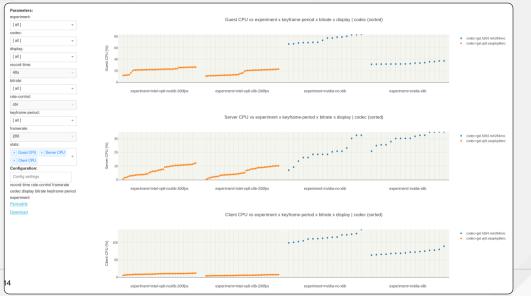
SPICE RECORDING INFRASTRUCTURE

Scripting and Matrix Benchmarking

... and let it run during the coffee break!

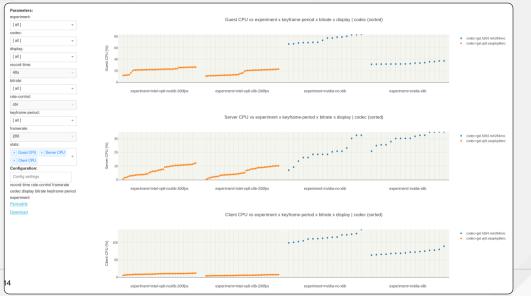


Overview: multiple plots for overview view of CPU usage (Guest/Server/Client))



Red Hat

Overview: multiple plots for overview view of CPU usage (Guest/Server/Client))



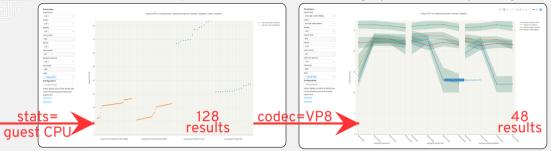
Red Hat

Overview: Guest CPU usage with 4+ variable parameters



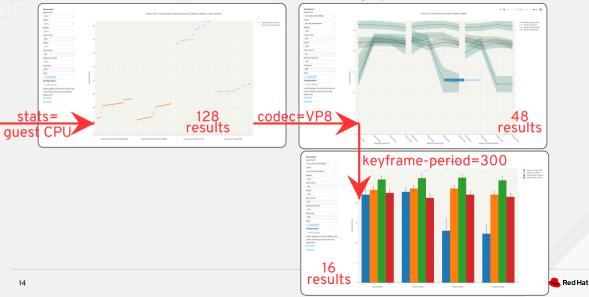


Overview: Guest CPU usage with 3 variable parameters (display, bitrate, keyframe-period)

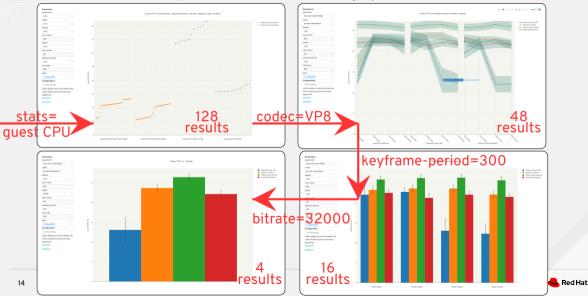




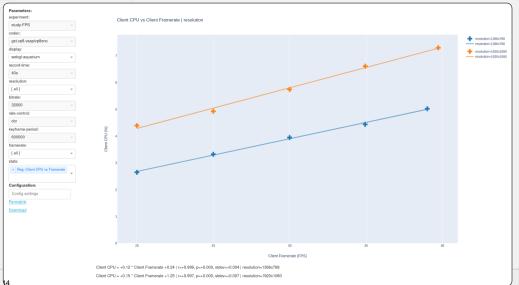
Overview: Guest CPU usage with 2 variable parameters (display, bitrate)



Overview: Guest CPU usage with 1 variable parameter (display)

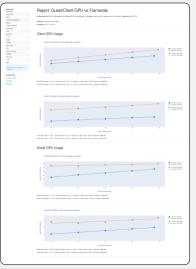


Overview: linear regression (Client CPU vs Framerate)



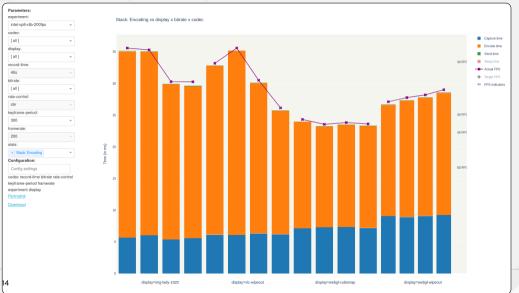
Red Hat

Overview: report mode (GPU vs Framerate on Guest/Client)





Overview: encoding stack on the guest (current work)

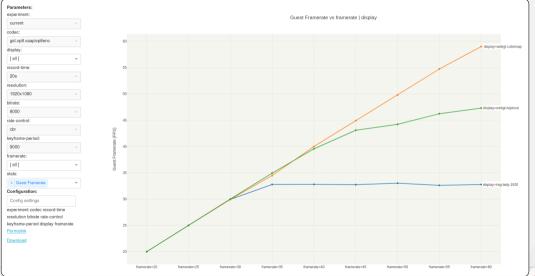


Red Hat

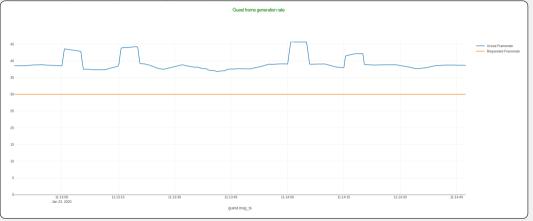
Some interesting cases and open questions...



FPS: why desktop encoding is slow?

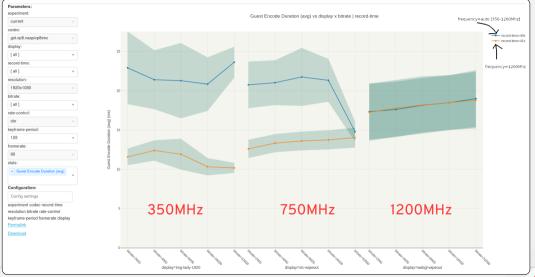


FPS: why desktop encoding is bumpy?

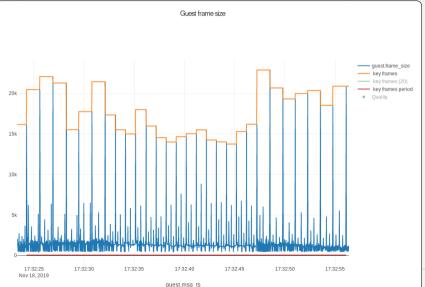




FPS: why? because the GPU tries to save energy!!



keyframe-period=0 \rightarrow "auto". Thanks Intel, but what does it really mean?





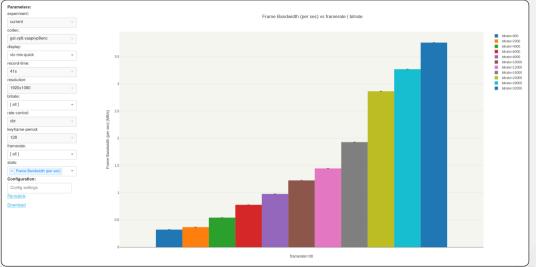
keyframe-period=0 \rightarrow "auto" \Rightarrow 1 keyframe per second!

| Parameters : experiment: current | Keyframe Period vs keyframe | -period x webpage x resolution x rate-control x bitrate fram | erate (sorted) |
|---|--|--|--|
| | | | framerate=20 |
| codec: | | | framerate=30 framerate=45 |
| gst.vp8.vaapivp8enc ~ | 60 • • • • • • • • • • • • • • • • • • • | | • framerate=60 |
| record_time: | | | |
| 30s - | | | |
| webpage: | | | |
| [al] - | 50 | | |
| bitrate: | | | |
| [all] ~ | | | |
| rate-control: | | | |
| [all] ~ | § 40 | | |
| | (ju sau | | |
| [all] ~ | | | |
| framerate: | 2 9 | | |
| [al] ~ | 5 30 · · · · · · · · · · · · · · · · · · | | |
| stats: | 2 | | |
| × Keyframe Period * | | | |
| Aspect: | | | |
| Show text | 20 • • • • • • • • • • • • • • • • • • • | • | |
| experiment codec record_time framerate bitrate rate-control resolution | | | |
| webpage keyframe-period | | | |
| Invalids: | | | |
| SHOW DELETE | 10 | | |
| | | | |
| | | | |
| | keyframe-period=0 | keyframe-period=20 | keyframe-period=40 |



MATRIX VISUALIZER: OPEN QUESTIONS

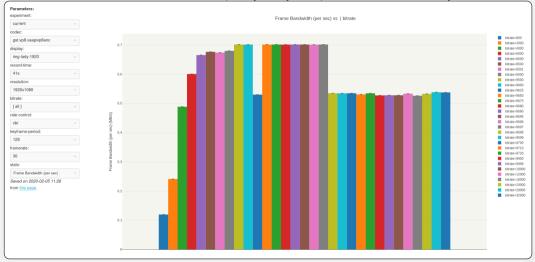
Bandwidth vs bitrate with VLC playback: everything as expected





MATRIX VISUALIZER: OPEN QUESTIONS

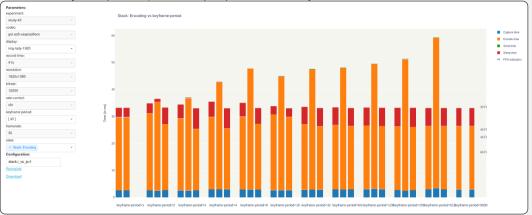
Bandwidth vs bitrate with still desktop: beginning as expected, but what's wrong with the tail?





MATRIX VISUALIZER: OPEN QUESTIONS

Encoding vs keyframe-period: why keyframe encoding take more and more time?







THANK YOU



plus.google.com/+RedHat



linkedin.com/company/red-hat



f

twitter.com/RedHatNews

facebook.com/redhatinc



youtube.com/user/RedHatVideos



Extra: a bit of help with the interface ...



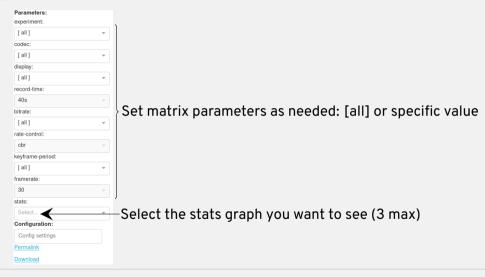
Interface in details

.

| Faranteters. | |
|------------------|---|
| experiment: | |
| [all] | - |
| codec: | |
| [all] | - |
| display: | |
| [all] | ~ |
| record-time: | |
| 40s | ~ |
| bitrate: | |
| [all] | ~ |
| rate-control: | |
| cbr | ~ |
| keyframe-period: | |
| [all] | ~ |
| framerate: | |
| 30 | ~ |
| stats: | |
| Select | |
| Configuration: | |
| Config settings | |
| Permalink | |
| Download | |

Set matrix parameters as needed: [all] or specific value

Interface in details



Interface in details

| experiment: | |
|------------------|---|
| [all] | ~ |
| codec: | |
| [all] | ~ |
| display: | |
| [all] | Ŧ |
| record-time: | |
| 40s | ~ |
| bitrate: | |
| [all] | ~ |
| rate-control: | |
| cbr | ~ |
| keyframe-period: | |
| [all] | ~ |
| framerate: | |
| 30 | ~ |
| stats: | |
| Select | |
| Configuration: | |
| Config settings | |
| Permalink | |
| Download | |

Set matrix parameters as needed: [all] or specific value Click the label to change the property order (subplot, ..., legend)

Select the stats graph you want to see (3 max)

Interface in details

Parameters: experiment: [all] codec: [all] display [all] record-time: 40s bitrate [all] rate-control: chr keyframe-period: [all] framerate 30 state Configuration: Confia settinas

Set matrix parameters as needed: [all] or specific value Click the label to change the property order (subplot, ..., legend)

Select the stats graph you want to see (3 max)

Only works if same tool/dataset available at reload

Interface in details



Set matrix parameters as needed: [all] or specific value Click the label to change the property order (subplot, ..., legend)

Select the stats graph you want to see (3 max)

Only works if same tool/dataset available at reload dill (pickle) object to store/share easily

Interface in details - mainly for dev

| Parameters: | _ | -Click to redraw the draw without reload the page |
|------------------|----|---|
| [all] | Ŧ | |
| codec: | | |
| [all] | ~ | |
| display: | | |
| [all] | - | |
| record-time: | | |
| 40s | ~ | |
| bitrate: | | |
| [all] | - | |
| rate-control: | | |
| cbr | Ŧ | |
| keyframe-period: | | |
| [all] | Ŧ | |
| framerate: | | |
| 30 | w. | |
| stats: | | |
| Select | * | |
| Configuration: | | Extra (an acific acttings |
| Config settings | | Extra/specific settings |
| Permalink | | |
| Download | | |

