AUTOMOTIVE LINUX SUMMIT EUROPE

Debugging Tools and Techniques for Virtualized Automotive Systems

Hiroyuki Ishii, *Panasonic Automotive Systems Co., Ltd.*
AGL System Architect Team & Virtualization Expert

#EMBEDDEDOSSUMMIT

Agenda

- Introduction
- Debugging Tools for Virtualized Systems
- Example: Analyzing vhost-net
- Summary

Agenda

- Introduction
- Debugging Tools for Virtualized Systems
- Example: Analyzing vhost-net
- Summary

Who Am I

- Hiroyuki Ishii
- >15 years of experience in software development for IVI and CDC products at Panasonic Automotive Systems
- Expertise: Linux kernel, performance engineering, virtualization, cloud-native technologies
- Joined AGL project since 2021 as:
  Steering Committee, System Architect, Virtualization Expert etc.

Background Trends in Automotive Software

Panasonic AUTOMOTIVE

Various requirements arising related to Software-Defined Vehicles

- High-performance computing
- Hardware scalability
- Workload distribution
- Mixed criticality

Growing demand and initiatives for virtualization as a key technology

- AGL actively worked on Virtual Machine (VM) based demo integration
- SOAFEE, Eclipce SDV and Android Automotive are also actively contributing

Challenges: Complexity of Virtualized Systems

Panasonic AUTOMOTIVE

- Complex interactions between host and guest components
- Increased system footprint, complicated integration
- Limited capabilities/visibility within guest environments

↓

- Debugging/performance engineering become increasingly difficult
- Complications in understanding system behavior

↓

Needs for specialized tools and techniques for virtualized systems

Agenda

- Introduction
- Debugging Tools for Virtualized Systems
- Practice: Analyzing vhost-net
- Summary

Tools Available for Virtualized Systems

Panasonic AUTOMOTIVE

- perf
  - perf stat, perf kvm stat|record
- trace-cmd
  - trace-cmd record -e kvm
- bcc
  - virtiostat, kvmexit, …
- flamegraph.pl
- debuginfod

perf

- A powerful profiling/tracing tool for Linux system
- Frontend for perf_events, ftrace, kprobes etc.
- Developed within the Linux kernel source tree
  - Stable and reliable
  - Has strong dependency on kernel version
- Some hardware events are platform dependent

`trace-cmd`

- Frontend tool for ftrace
  - Developed by Steven Rostedt
- Provides features similar to perf
  - Simpler and more user-frendly (IMO)
- Recent updates designed for embedded/virtualized use cases
  - `trace-cmd agent`
  - `trace-cmd listen`
- https://github.com/rostedt/trace-cmd

bcc

- A BPF-based tracing tool for performance analysis and debugging
- Offers rich set of features:
  - Python support
  - Various practical example scripts included
  - Simplifies creation of custom tools built upon it
- Drawbacks:
  - Relatively large dependency (LLVM, Python)
  - Limited support for the ARM platform
- https://github.com/iovisor/bcc

flamegraph.pl

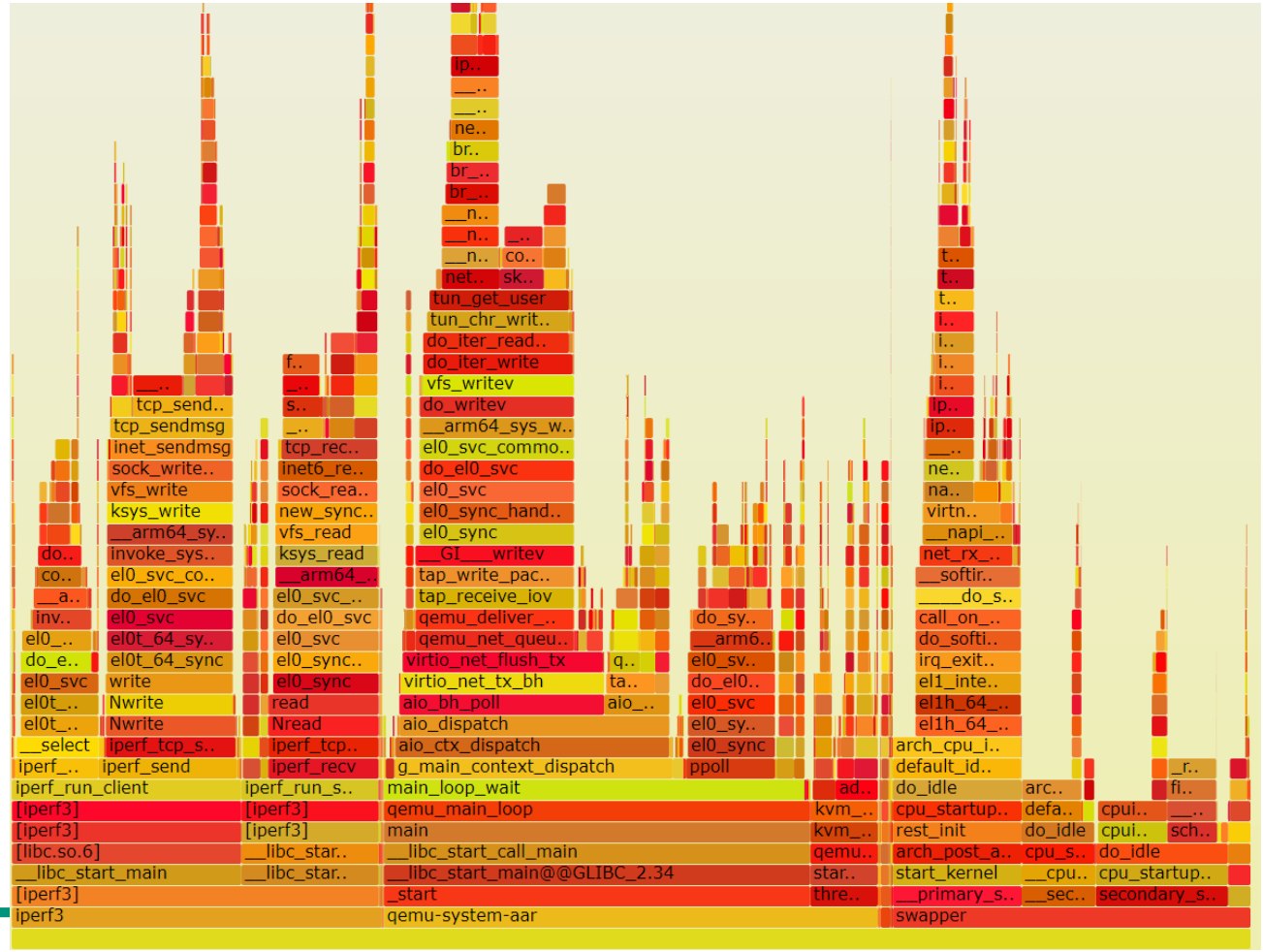
Panasonic AUTOMOTIVE

- A visualization tool for performance data samples
- Helps understanding workload overview and identifying bottlenecks
- https://github.com/brendangregg/FlameGraph

```
qemu-system-aar  2822 [005] 105761.404072:    8063588   cycles:
        ffff8000102cdb44 get_page_from_freelist+0x174 ([kernel.kallsyms])
        ffff8000102cfa2c __alloc_pages_nodemask+0x16c ([kernel.kallsyms])
        ffff8000102ed824 alloc_pages_current+0x94 ([kernel.kallsyms])
        ffff8000102c9970 __get_free_pages+0x20 ([kernel.kallsyms])
        ffff800010348020 __pollwait+0x60 ([kernel.kallsyms])
        ffff8000103981d8 eventfd_poll+0x68 ([kernel.kallsyms])
        ffff800010349868 do_sys_poll+0x268 ([kernel.kallsyms])
        ffff800010349e5c __arm64_sys_ppoll+0xac ([kernel.kallsyms])
        ffff800010028fa4 el0_svc_common.constprop.0+0x84 ([kernel.kallsyms])
        ffff800010029104 do_el0_svc+0x34 ([kernel.kallsyms])
        ffff8000111cdb40 el0_svc+0x20 ([kernel.kallsyms])
        ffff8000111ce0f4 el0_sync_handler+0xa4 ([kernel.kallsyms])
        ffff800010012700 el0_sync+0x180 ([kernel.kallsyms])
            ffffa2dfe298 ppoll+0x98 (/usr/lib/libc.so.6)
            aaaad028fdfc main_loop_wait+0x14c (/usr/bin/qemu-system-aarch64)
            aaaad00151a8 qemu_main_loop+0x144 (/usr/bin/qemu-system-aarch64)
            aaaacfaded34 main+0x14 (/usr/bin/qemu-system-aarch64)
            ffffa2d4b230 __libc_start_call_main+0x60 (/usr/lib/libc.so.6)
            ffffa2d4b30c __libc_start_main@@GLIBC_2.34+0x9c (/usr/lib/libc.so.6)
            aaaacfae4cf0 _start+0x30 (/usr/bin/qemu-system-aarch64)

qemu-system-aar  2829 [007] 105761.407769:    2762147   cycles:
        ffff80001025e380 clear_rseq_cs+0x20 ([kernel.kallsyms])
        ffff80001001e74c do_notify_resume+0x14c ([kernel.kallsyms])
        ffff8000100131dc work_pending+0xc ([kernel.kallsyms])
            ffffa2dffe90 __GI___ioctl+0x10 (/usr/lib/libc.so.6)
            aaaad00ff06c kvm_cpu_exec+0xac (/usr/bin/qemu-system-aarch64)
            aaaad0100630 kvm_vcpu_thread_fn+0xb0 (/usr/bin/qemu-system-aarch64)
            aaaad0272404 qemu_thread_start+0xa0 (/usr/bin/qemu-system-aarch64)
            ffffa2da02a8 start_thread+0x2c8 (/usr/lib/libc.so.6)
            ffffa2e07c1c thread_start+0xc (/usr/lib/libc.so.6)
...
```

(Continues over 100k lines)



debuginfod

- A host daemon for handling/managing debuginfo files
  - Part of the elfutils project (git://sourceware.org/git/elfutils.git)
  - Compatible with numerous popular debugging tools (gdb, perf, trace-cmd, bcc, and many others)
- Beneficial for embedded systems
  - Allows for offloading debuginfo files from the target system
  - Resolves path conflicts of debuginfo within virtualized environments
- Integrated in Yocto build system (oe-debuginfod)

Agenda

- Introduction
- Debugging Tools for Virtualized Systems
- Example: Analyzing vhost-net
- Summary

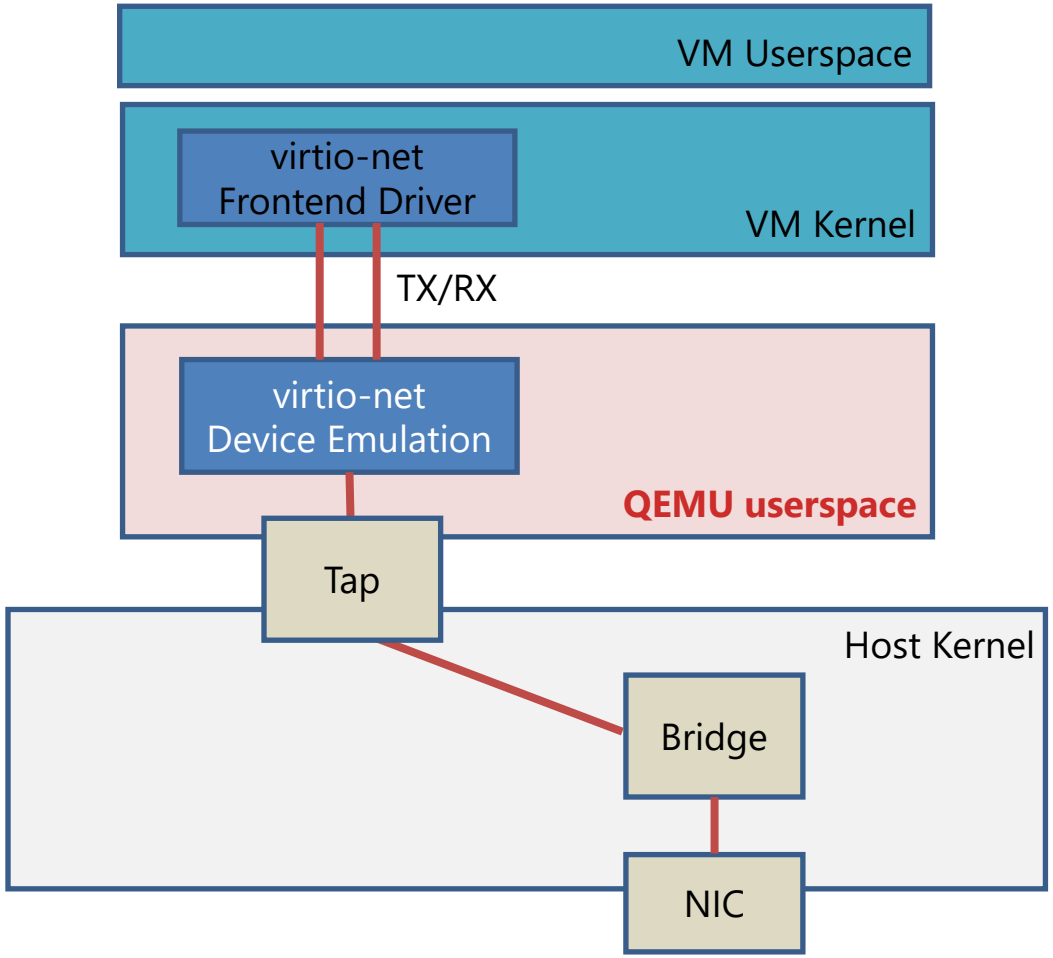
Goal of This Section

- Gaining knowledge of debugging tools' usage and techniques in a virtualized environment, through a practical example
- Investigating the behavior of **vhost-net** (acceleration for virtio-net), by comparing it with **standard virtio-net**
  - How and why it can improve performance?
  - Any potential side effects?

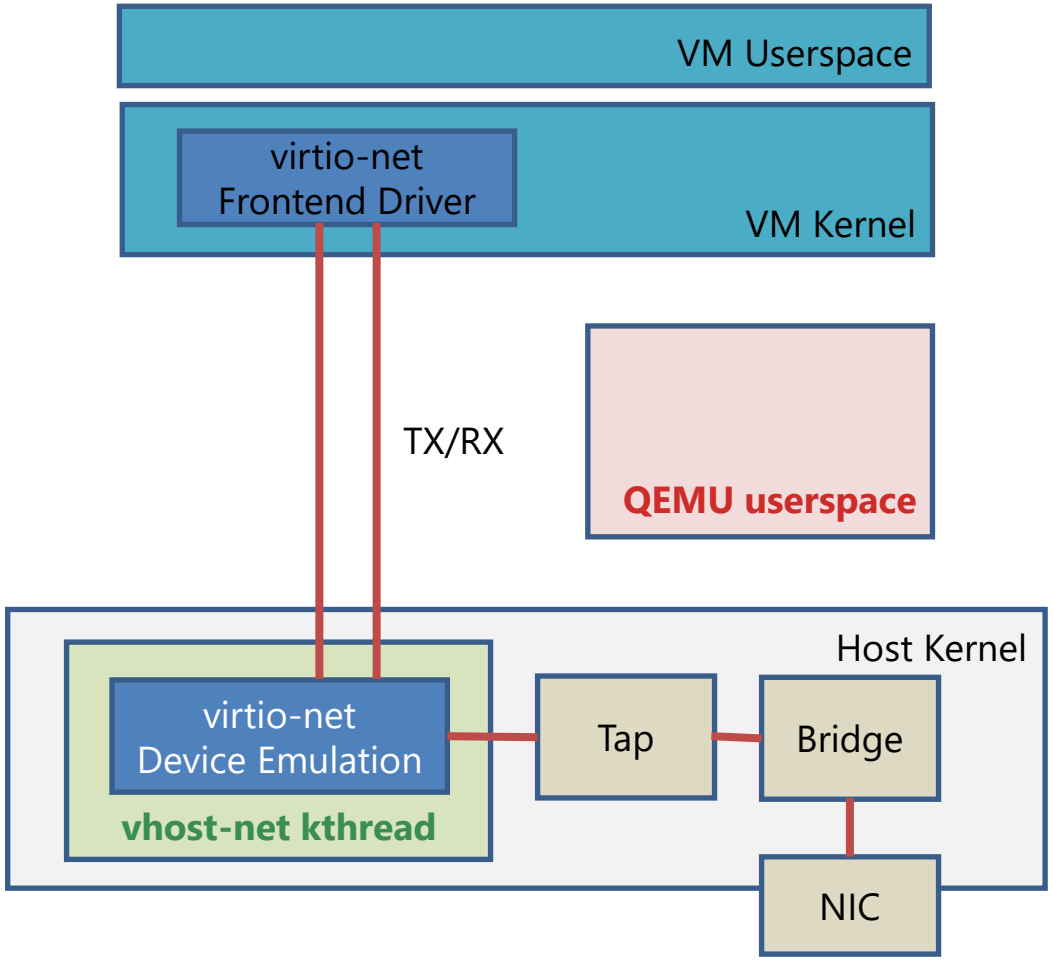
Standard virtio-net vs vhost-net

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:



When vhost-net is enabled, virtio-net device emulation is shifted from QEMU userspace to the vhost-net kernel thread

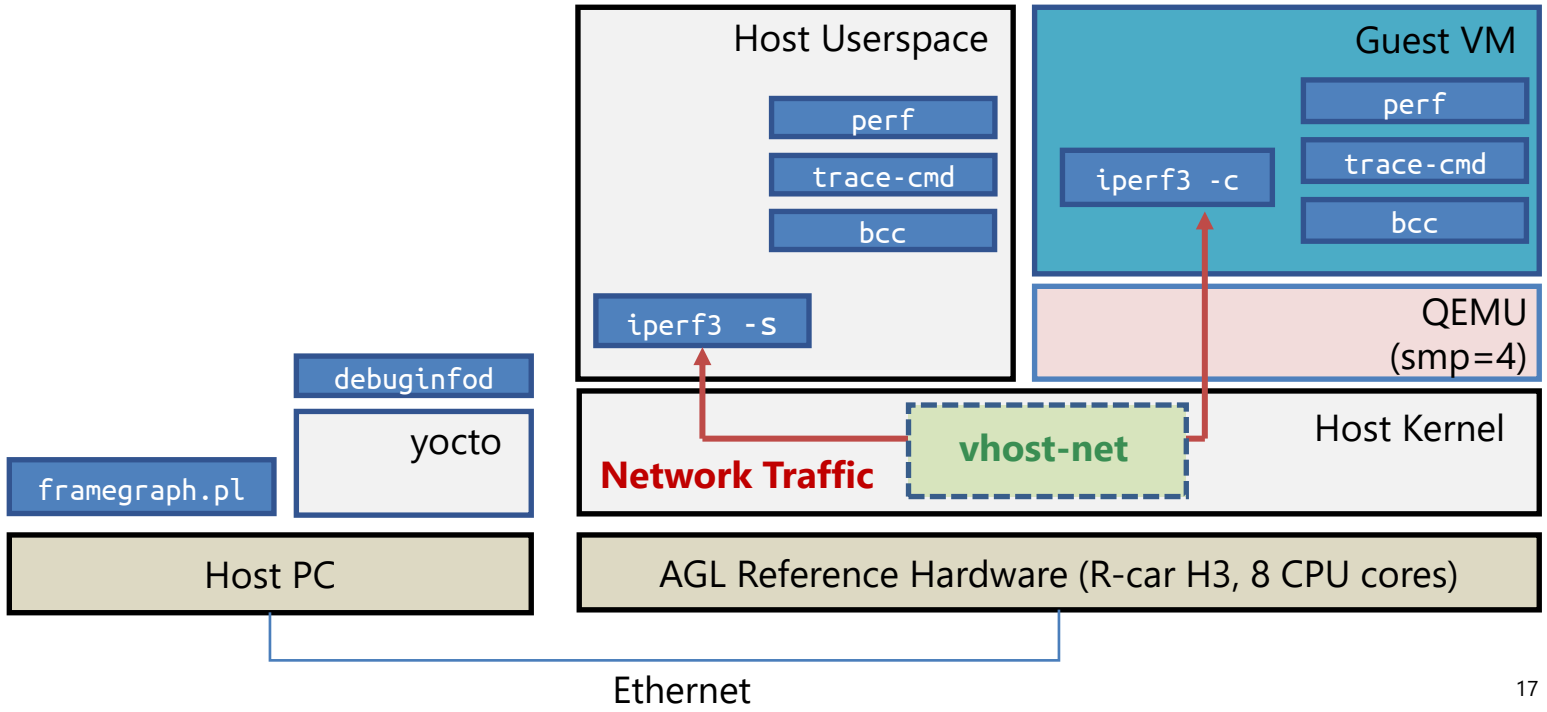
Preconditions

Panasonic AUTOMOTIVE

- Base image: agl-kvm-demo-platform built from AGL-UCB master branch
- Additional installations: vhost-net, iperf3, ssh and various debugging tools
  - For more details, please refer to the setup procedures in the appendix
- Running on AGL Reference Hardware (R-car H3)
- Hosting a single guest VM by QEMU using smp=4 (4 vCPUs)
- Generating network traffic by iperf3 between host and guest
- Halting all workloads other than iperf3 prior to the investigation
- Comparing metrics/traces between vhost-net and standard virtio by utilizing `perf`, `trace-cmd`, and `bcc`

AGL Reference Hardware

Photo from: Getting Close to Real Automotive Product with AGL Reference Hardware - Naoya Tanimura, Panasonic Corporation



Measuring Network Bitrate: iperf

Panasonic AUTOMOTIVE

Standard virtio:

```
root@guest:~# iperf3 -c 172.16.10.1
Connecting to host 172.16.10.1, port 5201
[  5] local 172.16.10.2 port 42458 connected to 172.16.10.1 port 5201
[ ID] Interval           Transfer     Bitrate         Retr  Cwnd
[  5]   0.00-1.00   sec   175 MBytes  1.47 Gbits/sec    0   1.53 MBytes
[  5]   1.00-2.00   sec   191 MBytes  1.61 Gbits/sec    0   1.53 MBytes
[  5]   2.00-3.00   sec   190 MBytes  1.59 Gbits/sec    0   1.61 MBytes
[  5]   3.00-4.00   sec   192 MBytes  1.61 Gbits/sec    0   1.69 MBytes
[  5]   4.00-5.00   sec   192 MBytes  1.61 Gbits/sec    0   1.80 MBytes
[  5]   5.00-6.00   sec   210 MBytes  1.76 Gbits/sec    0   1.80 MBytes
[  5]   6.00-7.00   sec   199 MBytes  1.67 Gbits/sec    0   1.80 MBytes
[  5]   7.00-8.00   sec   202 MBytes  1.70 Gbits/sec    0   1.89 MBytes
[  5]   8.00-9.00   sec   212 MBytes  1.78 Gbits/sec    0   1.99 MBytes
[  5]   9.00-10.00  sec   215 MBytes  1.80 Gbits/sec    0   2.10 MBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bitrate         Retr
[  5]   0.00-10.00  sec  1.93 GBytes  1.66 Gbits/sec    0             sender
[  5]   0.00-10.01  sec  1.93 GBytes  1.66 Gbits/sec                  receiver
```

vhost-net:

```
root@guest:~# iperf3 -c 172.16.10.1
Connecting to host 172.16.10.1, port 5201
[  5] local 172.16.10.2 port 55840 connected to 172.16.10.1 port 5201
[ ID] Interval           Transfer     Bitrate         Retr  Cwnd
[  5]   0.00-1.00   sec   391 MBytes  3.28 Gbits/sec    0   1.64 MBytes
[  5]   1.00-2.00   sec   382 MBytes  3.21 Gbits/sec    0   1.64 MBytes
[  5]   2.00-3.00   sec   385 MBytes  3.23 Gbits/sec    0   1.74 MBytes
[  5]   3.00-4.00   sec   385 MBytes  3.23 Gbits/sec    0   1.74 MBytes
[  5]   4.00-5.00   sec   382 MBytes  3.21 Gbits/sec    0   1.74 MBytes
[  5]   5.00-6.00   sec   386 MBytes  3.23 Gbits/sec    0   1.74 MBytes
[  5]   6.00-7.00   sec   401 MBytes  3.37 Gbits/sec    0   2.06 MBytes
[  5]   7.00-8.00   sec   391 MBytes  3.29 Gbits/sec    0   2.06 MBytes
[  5]   8.00-9.00   sec   386 MBytes  3.24 Gbits/sec    0   2.06 MBytes
[  5]   9.00-10.00  sec   392 MBytes  3.29 Gbits/sec    0   2.06 MBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bitrate         Retr
[  5]   0.00-10.00  sec  3.79 GBytes  3.26 Gbits/sec    0             sender
[  5]   0.00-10.01  sec  3.79 GBytes  3.26 Gbits/sec                  receiver
```

vhost-net is about 2x faster than standard virtio-net

KVM Event Stats: `perf stat -e 'kvm:*'`

Panasonic AUTOMOTIVE

Standard virtio:

```
root@host:~# perf stat -e 'kvm:*' -a -- ssh guest iperf3 -c 172.16.10.1 -n 100M
              3656      kvm:vgic_update_irq_pending
              1145      kvm:kvm_wfx_arm64
              5453      kvm:kvm_arm_setup_debug
              5453      kvm:kvm_arm_clear_debug
             10906      kvm:kvm_arm_set_dreg32
              2436      kvm:kvm_handle_sys_reg
              2436      kvm:kvm_sys_access
              5453      kvm:kvm_entry
              5453      kvm:kvm_exit
              1613      kvm:kvm_guest_fault
              1181      kvm:kvm_irq_line
              2475      kvm:kvm_timer_update_irq
              7511      kvm:kvm_get_timer_map
              2328      kvm:kvm_timer_save_state
              2328      kvm:kvm_timer_restore_state
              2328      kvm:kvm_timer_emulate
              1180      kvm:kvm_userspace_exit
              1143      kvm:kvm_vcpu_wakeup
              2201      kvm:kvm_mmio
               159      kvm:kvm_halt_poll_ns
```

Limiting number of packets

vhost-net:

```
root@host:~# perf stat -e 'kvm:*' -a -- ssh guest iperf3 -c 172.16.10.1 -n 100M
              5377      kvm:vgic_update_irq_pending
               611      kvm:kvm_wfx_arm64
              7348      kvm:kvm_arm_setup_debug
              7348      kvm:kvm_arm_clear_debug
             14696      kvm:kvm_arm_set_dreg32
              2429      kvm:kvm_handle_sys_reg
              2429      kvm:kvm_sys_access
              7348      kvm:kvm_entry
              7348      kvm:kvm_exit
              3510      kvm:kvm_guest_fault
              2548      kvm:kvm_irq_line
              2829      kvm:kvm_timer_update_irq
             12315      kvm:kvm_get_timer_map
              2678      kvm:kvm_timer_save_state
              2678      kvm:kvm_timer_restore_state
              2678      kvm:kvm_timer_emulate
              2188      kvm:kvm_userspace_exit
               611      kvm:kvm_vcpu_wakeup
              4604      kvm:kvm_mmio
               128      kvm:kvm_halt_poll_ns
```

Significant reduction in WFXx (CPU Standby) and wakeup occurrences on guest vCPUs

KVM VM-Exit Stats: `perf kvm stat`

Panasonic AUTOMOTIVE

Standard virtio:

```
root@host:~# perf kvm stat record -a -- ssh guest iperf3 -c 172.16.10.1 -n 500M && perf kvm report
Analyze events for all VMs, all VCPUs:

             VM-EXIT    Samples  Samples%     Time%    Min Time    Max Time         Avg time

            DABT_LOW       4578    40.12%    26.53%      6.60us   1961.41us    104.75us ( +-   1.95% )
                 WFx       3546    31.08%    46.04%      8.04us   2113.93us    234.70us ( +-   2.56% )
               SYS64       2430    21.30%     2.18%      5.64us   1621.33us     16.18us ( +-   7.93% )
                 IRQ        856     7.50%    25.25%     11.16us   2793.49us    533.31us ( +-   3.70% )
```

vhost-net:

```
root@host:~# perf kvm stat record -a -- ssh guest iperf3 -c 172.16.10.1 -n 500M && perf kvm report
Analyze events for all VMs, all VCPUs:

             VM-EXIT    Samples  Samples%     Time%    Min Time    Max Time         Avg time

            DABT_LOW      23050    66.71%    66.59%      6.24us   3449.53us     59.70us ( +-   1.13% )
                 IRQ       6750    19.54%    19.44%      8.28us   3951.14us     59.50us ( +-   4.89% )
               SYS64       2433     7.04%     2.36%      5.76us   3352.33us     20.03us ( +-  10.95% )
                 WFx       2318     6.71%    11.62%     12.72us   1807.21us    103.55us ( +-   2.08% )
```

- While we've observed a decrease in WFx again, VM-Exit themselves have increased (mainly due to Data abort and IRQ events)
- Common belief: a higher number of VM-Exit events leads to poorer performance
  - But this may not be always true, WFx could be more costly

Monitoring I/O Operations on Virtio: virtiostat (bcc)

Panasonic AUTOMOTIVE

Standard virtio:

```
root@guest:~# python3 /usr/share/bcc/tools/virtiostat
[...]
        Driver   Device    VQ Name  In SGs Out SGs          In BW         Out BW
 b'virtio_net' b'virtio0' b'input.0'    8161       0       12733952              0
 b'virtio_net' b'virtio0' b'output.0'       0    8556              0      376253940
[...]
```

vhost-net:

```
root@guest:~# python3 /usr/share/bcc/tools/virtiostat
[...]
        Driver   Device    VQ Name  In SGs Out SGs          In BW         Out BW
 b'virtio_net' b'virtio0' b'input.0'    7385       0       11523072              0
 b'virtio_net' b'virtio0' b'output.0'       0    8572              0      375796680
[...]
```

- Traffic handled by virtio-net is almost consistent
  - Behavior of the guest side should not be changed

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

Before using `perf kvm record`, it's necessary to mount the guest filesystem from the host, as shown below:

```
root@host:~# mntpoint="/tmp/guestmount/$(pgrep -f qemu-system-aarch64)"
root@host:~# mkdir -p "$mntpoint"
root@host:~# sshfs -o allow_other,direct_io guest:/"$mntpoint"
```

Afterward, we can pass the mountpoint using `--guestmount` option:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount record ...
```

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount record -F99 -a -- ssh guest iperf3 -c 172.16.10.1 -b 100M -t 1
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio --show-cpu-utilization

# Overhead       sys       usr  guest sys  guest usr  Command          Shared Object                  Symbol
# ........  ........  ........  .........  .........  ...............  .............................  ................
#
    20.11%     0.00%     0.00%     20.11%      0.00%  :1035            [guest.kernel.kallsyms.1026]  [g] receive_buf
    13.47%    13.47%     0.00%      0.00%      0.00%  swapper          [kernel.kallsyms]             [k] cpuidle_enter_state
     9.56%     0.00%     9.56%      0.00%      0.00%  qemu-system-aar  qemu-system-aarch64           [.] replay_mutex_unlock
     4.97%     0.00%     0.00%      4.97%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] cpu_do_idle
     4.01%     4.01%     0.00%      0.00%      0.00%  swapper          [kernel.kallsyms]             [k] _raw_spin_unlock_irq
     3.19%     0.00%     3.19%      0.00%      0.00%  ssh              ssh.openssh                   [.] square
     2.57%     0.00%     0.00%      2.57%      0.00%  :1035            [guest.kernel.kallsyms.1026]  [g] unwind_next
     2.00%     0.00%     0.00%      0.00%      2.00%  :1035            [unknown]                     [u] 0x000000557d660af8
     1.84%     0.00%     0.00%      1.84%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] update_load_avg
     1.82%     0.00%     1.82%      0.00%      0.00%  qemu-system-aar  qemu-system-aarch64           [.] object_dynamic_cast_assert
     1.82%     0.00%     1.82%      0.00%      0.00%  qemu-system-aar  libc.so.6                     [.] clock_gettime@@GLIBC_2.17
     1.78%     0.00%     0.00%      1.78%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] unwind_next
     1.68%     1.68%     0.00%      0.00%      0.00%  qemu-system-aar  [kernel.kallsyms]             [k] _raw_spin_unlock_irq
     1.54%     0.00%     0.00%      1.54%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] __schedule
```

guest kernel function named receive_buf() consumes 20.11% of CPU

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount record -F99 -a -- ssh guest iperf3 -c 172.16.10.1 -b 100M -t 1
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio --show-cpu-utilization

# Overhead       sys       usr  guest sys  guest usr  Command          Shared Object                  Symbol
# ........  ........  ........  .........  .........  ...........  ...............................  ..............................
#
    20.11%     0.00%     0.00%     20.11%      0.00%  :1035                                        receive_buf
    13.47%    13.47%     0.00%      0.00%      0.00%  swapper                                      cpuidle_enter_state
     9.56%     0.00%     9.56%      0.00%      0.00%  qemu-system-aar  qemu-system-aarch64         [.] replay_mutex_unlock
     4.97%     0.00%     0.00%      4.97%      0.00%  :1037            [guest.kernel.kallsyms.1026] [g] cpu_do_idle
     4.01%     4.01%     0.00%      0.00%      0.00%  swapper          [kernel.kallsyms]            [k] _raw_spin_unlock_irq
     3.19%     0.00%     3.19%      0.00%      0.00%  ssh              ssh.openssh                  [.] square
     2.57%     0.00%     0.00%      2.57%      0.00%  :1035            [guest.kernel.kallsyms.1026] [g] unwind_next
     2.00%     0.00%     0.00%      0.00%      2.00%  :1035            [unknown]                    [u] 0x000000557d660af8
     1.84%     0.00%     0.00%      1.84%      0.00%  :1037            [guest.kernel.kallsyms.1026] [g] update_load_avg
     1.82%     0.00%     1.82%      0.00%      0.00%  qemu-system-aar  qemu-system-aarch64         [.] object_dynamic_cast_assert
     1.82%     0.00%     1.82%      0.00%      0.00%  qemu-system-aar  libc.so.6                    [.] clock_gettime@@GLIBC_2.17
     1.78%     0.00%     0.00%      1.78%      0.00%  :1037            [guest.kernel.kallsyms.1026] [g] unwind_next
     1.68%     1.68%     0.00%      0.00%      0.00%  qemu-system-aar  [kernel.kallsyms]            [k] _raw_spin_unlock_irq
     1.54%     0.00%     0.00%      1.54%      0.00%  :1037            [guest.kernel.kallsyms.1026] [g] __schedule
```

:tid represents a child thread of another process (in this case, vCPU threads of qemu)

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount record -F99 -a -- ssh guest iperf3 -c 172.16.10.1 -b 100M -t 1
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio --show-cpu-utilization

# Overhead       sys       usr  guest sys  guest usr  Command          Shared Object                 Symbol
# ........  ........  ........  .........  .........  ...............  ............................  ....................................
#
    20.11%     0.00%     0.00%     20.11%      0.00%  :1035            [guest.kernel.kallsyms.1026]  [g] receive_buf
    13.47%    13.47%     0.00%      0.00%      0.00%  swapper          [kernel.kallsyms]             [k] cpuidle_enter_state
     9.56%     0.00%     9.56%      0.00%      0.00%  qemu-system-aar  qemu-system-aarch64           [.] replay_mutex_unlock
     4.97%     0.00%     0.00%      4.97%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] cpu_do_idle
     4.01%     4.01%     0.00%      0.00%      0.00%  swapper          [kernel.kallsyms]             [k] _raw_spin_unlock_irq
     3.19%     0.00%     3.19%      0.00%      0.00%  ssh              ssh.openssh                   [.] square
     2.57%     0.00%     0.00%      2.57%      0.00%  :1035            [guest.kernel.kallsyms.1026]  [g] unwind_next
     2.00%     0.00%     0.00%      0.00%      2.00%  :1035            [unknown]                     [u] 0x000000557d660af8
     1.84%     0.00%     0.00%      1.84%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] update_load_avg
     1.82%     0.00%     1.82%      0.00%      0.00%  qemu-system-aar  qemu-system-aarch64           [.] object_dynamic_cast_assert
     1.82%     0.00%     1.82%      0.00%      0.00%  qemu-system-aar  libc.so.6                     [.] clock_gettime@@GLIBC_2.17
     1.78%     0.00%     0.00%      1.78%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] unwind_next
     1.68%     1.68%     0.00%      0.00%      0.00%  qemu-system-aar  [kernel.kallsyms]             [k] _raw_spin_unlock_irq
     1.54%     0.00%     0.00%      1.54%      0.00%  :1037            [guest.kernel.kallsyms.1026]  [g] __schedule
```

- Can analyze high-load functions across both host and guest systems
- However, this function-level of detail might be too much in this analysis

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

- Summarize the results using `perf kvm report --sort` option

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    30.72%     0.00%     0.00%     28.72%      2.00%  :1035
    25.31%     4.38%    20.94%      0.00%      0.00%  qemu-system-aar
    18.23%     0.00%     0.00%     18.23%      0.00%  :1037
    18.13%    18.13%     0.00%      0.00%      0.00%  swapper
     3.98%     0.00%     3.98%      0.00%      0.00%  ssh
     1.35%     0.00%     0.00%      1.35%      0.00%  :1038
     1.29%     0.85%     0.44%      0.00%      0.00%  iperf3
     0.97%     0.00%     0.00%      0.97%      0.00%  :1039
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/6
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

vhost-net:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    40.37%     0.00%     0.00%     40.37%      0.00%  :1146
    15.19%    15.19%     0.00%      0.00%      0.00%  swapper
    13.08%    13.08%     0.00%      0.00%      0.00%  qemu-system-aar
    11.72%     0.00%     0.00%     11.72%      0.00%  :1147
     6.87%     0.00%     6.87%      0.00%      0.00%  ssh
     5.45%     0.00%     0.00%      5.45%      0.00%  :1149
     4.28%     4.28%     0.00%      0.00%      0.00%  iperf3
     2.78%     2.78%     0.00%      0.00%      0.00%  vhost-1134
     0.21%     0.00%     0.21%      0.00%      0.00%  sshd
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/0
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

- Summarize the results using `perf kvm report --sort` option

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    30.72%     0.00%     0.00%     28.72%      2.00%  :1035
    25.31%     4.38%    20.94%      0.00%      0.00%  qemu-system-aar
    18.23%     0.00%     0.00%     18.23%      0.00%  :1037
    18.13%    18.13%     0.00%      0.00%      0.00%  swapper
     3.98%     0.00%     3.98%      0.00%      0.00%  ssh
     1.35%     0.00%     0.00%      1.35%      0.00%  :1038
     1.29%     0.85%     0.44%      0.00%      0.00%  iperf3
     0.97%     0.00%     0.00%      0.97%      0.00%  :1039
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/6
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

These overheads includes idle functions

vhost-net:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    40.37%     0.00%     0.00%     40.37%      0.00%  :1146
    15.19%    15.19%     0.00%      0.00%      0.00%  swapper
    13.08%    13.08%     0.00%      0.00%      0.00%  qemu-system-aar
    11.72%     0.00%     0.00%     11.72%      0.00%  :1147
     6.87%     0.00%     6.87%      0.00%      0.00%  ssh
     5.45%     0.00%     0.00%      5.45%      0.00%  :1149
     4.28%     4.28%     0.00%      0.00%      0.00%  iperf3
     2.78%     2.78%     0.00%      0.00%      0.00%  vhost-1134
     0.21%     0.00%     0.21%      0.00%      0.00%  sshd
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/0
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

- Summarize the results using `perf kvm report --sort` option

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    30.72%     0.00%     0.00%     28.72%      2.00%  :1035
    25.31%     4.38%    20.94%      0.00%      0.00%  qemu-system-aar
    18.23%     0.00%     0.00%     18.23%      0.00%  :1037
    18.13%    18.13%     0.00%      0.00%      0.00%  swapper
     3.98%     0.00%     3.98%      0.00%      0.00%  ssh
     1.35%     0.00%     0.00%      1.35%      0.00%  :1038
     1.29%     0.85%     0.44%      0.00%      0.00%  iperf3
     0.97%     0.00%     0.00%      0.97%      0.00%  :1039
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/6
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
-show-cpu-utilization | grep idle | sed -e 's!\[.*\]!!'
    13.47%    13.47%     0.00%      0.00%      0.00%  swapper  cpuidle_enter_state
     4.97%     0.00%     0.00%      4.97%      0.00%  :1037    cpu_do_idle
     0.64%     0.64%     0.00%      0.00%      0.00%  swapper  tick_nohz_idle_exit
     0.53%     0.00%     0.00%      0.53%      0.00%  :1035    cpu_do_idle
```

vhost-net:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    40.37%     0.00%     0.00%     40.37%      0.00%  :1146
    15.19%    15.19%     0.00%      0.00%      0.00%  swapper
    13.08%    13.08%     0.00%      0.00%      0.00%  qemu-system-aar
    11.72%     0.00%     0.00%     11.72%      0.00%  :1147
     6.87%     0.00%     6.87%      0.00%      0.00%  ssh
     5.45%     0.00%     0.00%      5.45%      0.00%  :1149
     4.28%     4.28%     0.00%      0.00%      0.00%  iperf3
     2.78%     2.78%     0.00%      0.00%      0.00%  vhost-1134
     0.21%     0.00%     0.21%      0.00%      0.00%  sshd
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/0
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
-show-cpu-utilization | grep idle | sed -e 's!\[.*\]!!'
    30.01%     0.00%     0.00%     30.01%      0.00%  :1146    cpu_do_idle
    13.05%    13.05%     0.00%      0.00%      0.00%  swapper  cpuidle_enter_state
     6.55%     0.00%     0.00%      6.55%      0.00%  :1147    cpu_do_idle
     3.59%     0.00%     0.00%      3.59%      0.00%  :1149    cpu_do_idle
```

identify the idle functions and manually exclude them from each processes

Performance Sampling Across Host and Guest: `perf kvm record`

Panasonic AUTOMOTIVE

- Summarize the results using `perf kvm report --sort` option

Standard virtio:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    30.72%     0.00%     0.00%     28.72%      2.00%  :1035
    25.31%     4.38%    20.94%      0.00%      0.00%  qemu-system-aar
    18.23%     0.00%     0.00%     18.23%      0.00%  :1037
    18.13%    18.13%     0.00%      0.00%      0.00%  swapper
     3.98%     0.00%     3.98%      0.00%      0.00%  ssh
     1.35%     0.00%     0.00%      1.35%      0.00%  :1038
     1.29%     0.85%     0.44%      0.00%      0.00%  iperf3
     0.97%     0.00%     0.00%      0.97%      0.00%  :1039
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/6
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

```
root@host:~# perf kvm --host --guest --gues
-show-cpu-utilization | grep idle | sed -e
    13.47%    13.47%     0.00%      0.00%
     4.97%     0.00%     0.00%      4.97%
     0.64%     0.64%     0.00%      0.00%
     0.53%     0.00%     0.00%      0.53%      0.00%  :1035    cpu_do_idle
```

```
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
    30.19%     0.00%     0.00%     28.19%      2.00%  :1035
    25.31%     4.38%    20.94%      0.00%      0.00%  qemu-system-aar
    13.26%     0.00%     0.00%     13.26%      0.00%  :1037
```

Results after exclusion:
CPU consumption has significantly decreased in vhost-net

vhost-net:

```
root@host:~# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
show-cpu-utilization --sort=comm
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
#
    40.37%     0.00%     0.00%     40.37%      0.00%  :1146
    15.19%    15.19%     0.00%      0.00%      0.00%  swapper
    13.08%    13.08%     0.00%      0.00%      0.00%  qemu-system-aar
    11.72%     0.00%     0.00%     11.72%      0.00%  :1147
     6.87%     0.00%     6.87%      0.00%      0.00%  ssh
     5.45%     0.00%     0.00%      5.45%      0.00%  :1149
     4.28%     4.28%     0.00%      0.00%      0.00%  iperf3
     2.78%     2.78%     0.00%      0.00%      0.00%  vhost-1134
     0.21%     0.00%     0.21%      0.00%      0.00%  sshd
     0.01%     0.01%     0.01%      0.00%      0.00%  perf
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/0
     0.01%     0.01%     0.00%      0.00%      0.00%  migration/5
```

```
# perf kvm --host --guest --guestmount=/tmp/guestmount report --stdio -
tilization | grep idle | sed -e 's!\[.*\]!!'
     0.00%     0.00%    30.01%      0.00%  :1146    cpu_do_idle
    13.05%     0.00%     0.00%      0.00%  swapper  cpuidle_enter_state
     0.00%     0.00%     6.55%      0.00%  :1147    cpu_do_idle
     3.59%     0.00%     0.00%      3.59%      0.00%  :1149    cpu_do_idle
```

```
# Overhead       sys       usr  guest sys  guest usr  Command
# ........  ........  ........  .........  .........  ...............
    10.36%     0.00%     0.00%     10.36%      0.00%  :1146
    13.08%    13.08%     0.00%      0.00%      0.00%  qemu-system-aar
     5.17%     0.00%     0.00%      5.17%      0.00%  :1147
     1.86%     0.00%     0.00%      1.86%      0.00%  :1149
```

Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Basic usage:

1. Perform a `perf record` with stack-trace (`-g`) and convert the data into text format using `perf script`

```
# perf record -a -g -F99 -- sleep 1    # or, any command you want to profile
# perf script > script.txt
```

2. Transfer the output to your host PC, then convert it to a SVG file using `stackcollapse-perf.pl` and `framegraph.pl`

```
$ cat script.txt | stackcollapse-perf.pl | flamegraph.pl > output.svg
```

3. Open the SVG file in a web browser to view the results.

Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Tips & Tricks for virtualized systems:

1. Record samples on both host and guest systems at once

Record on host side

Record on guest side

```
root@host:~# perf record -a -g -F99 -o /tmp/perf.data.host -- ssh guest perf record -a -g -F99 -o /tmp/perf.data.guest \
iperf3 -c 172.16.10.1 -b 1G
```

Workload

2. Merge the two outputs into a single SVG file

```
$ sed -i -e 's!\(^[^ \t]\+\)!<guest>\1!' script.guest  # Add prefix to guest output
$ cat script.host script.guest | stackcollapse-perf.pl | flamegraph.pl > output.svg
```

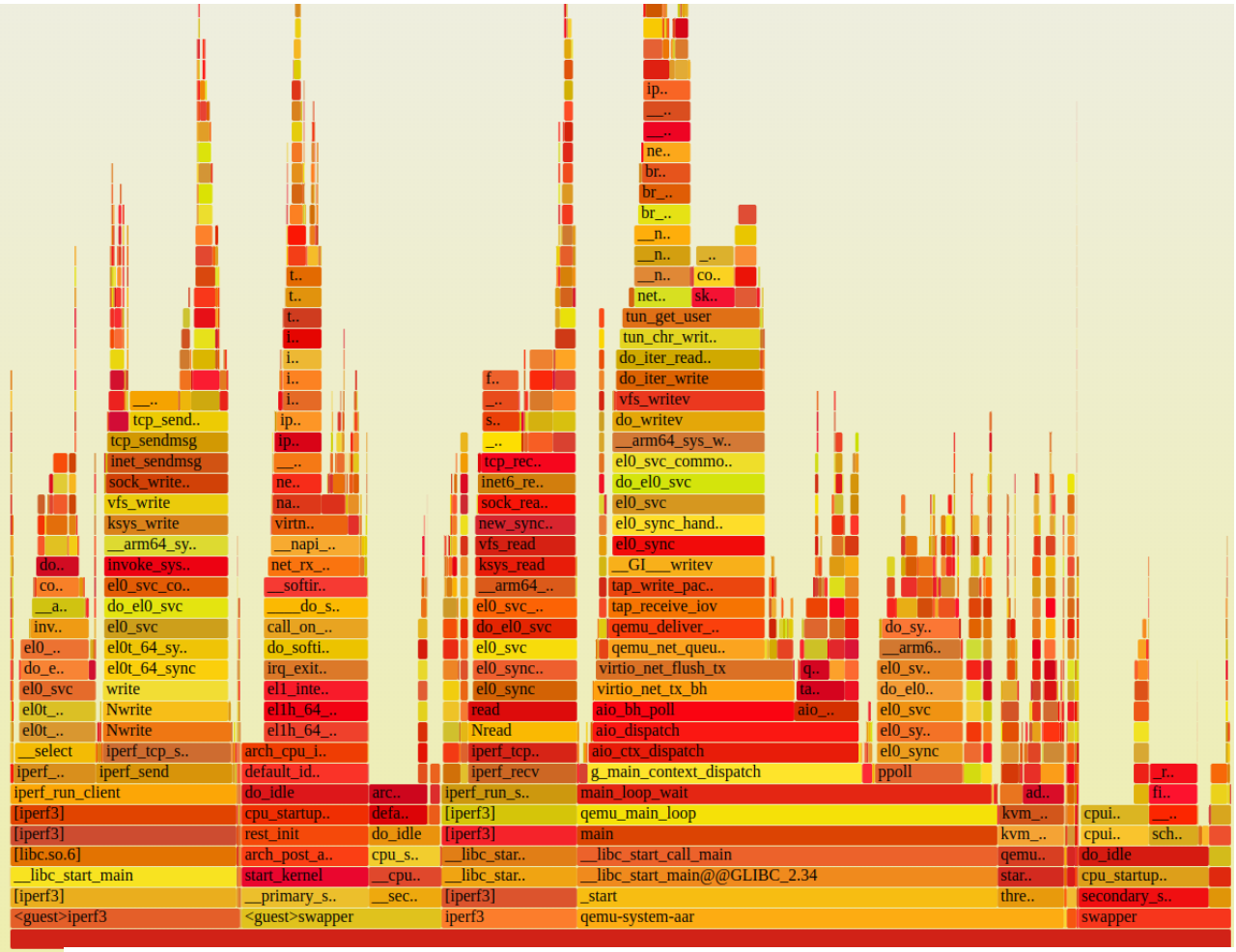
*Note that these procedures are designed for convenience and may not be strictly correct.

There may be overlapping workloads or inconsistent scaling between the guest and host systems.*

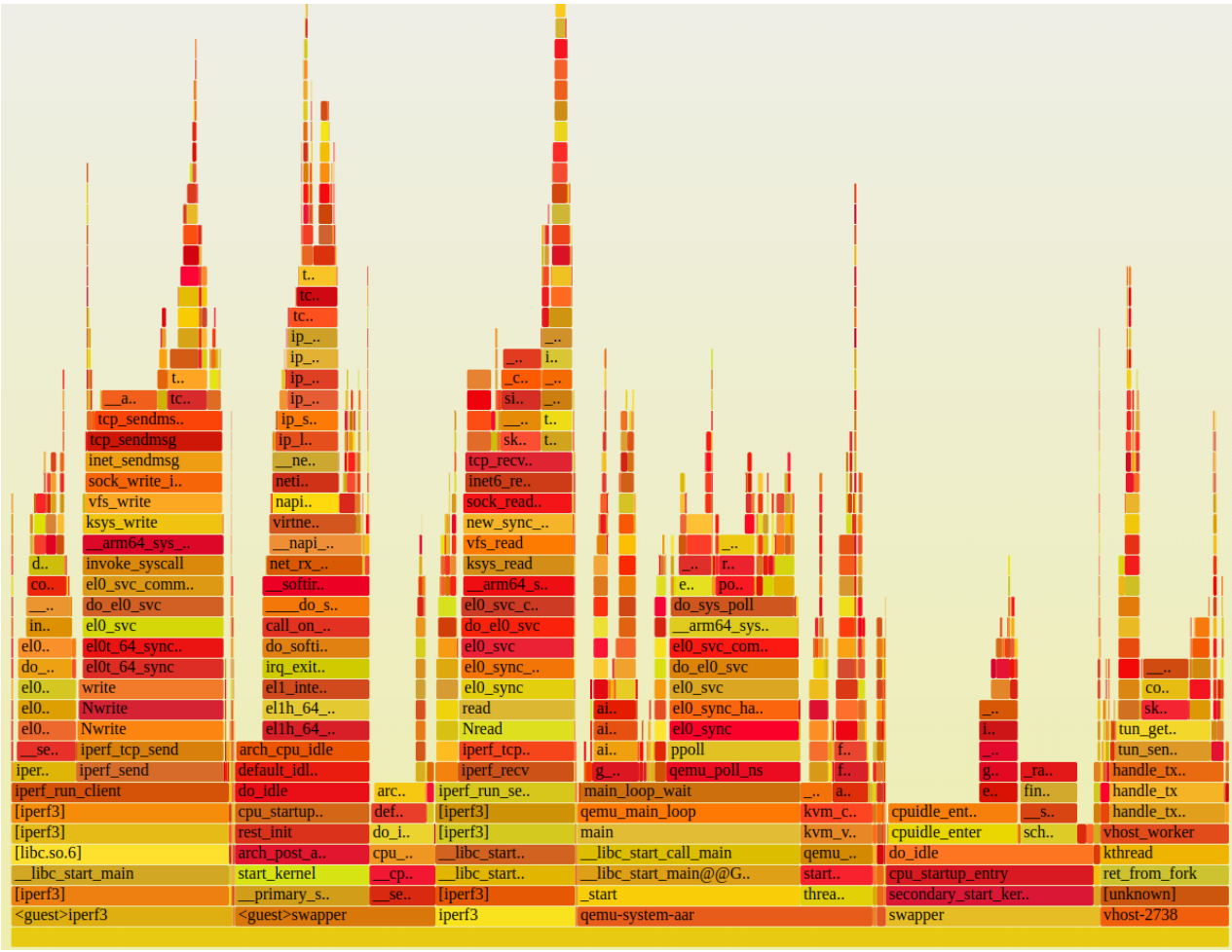
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:

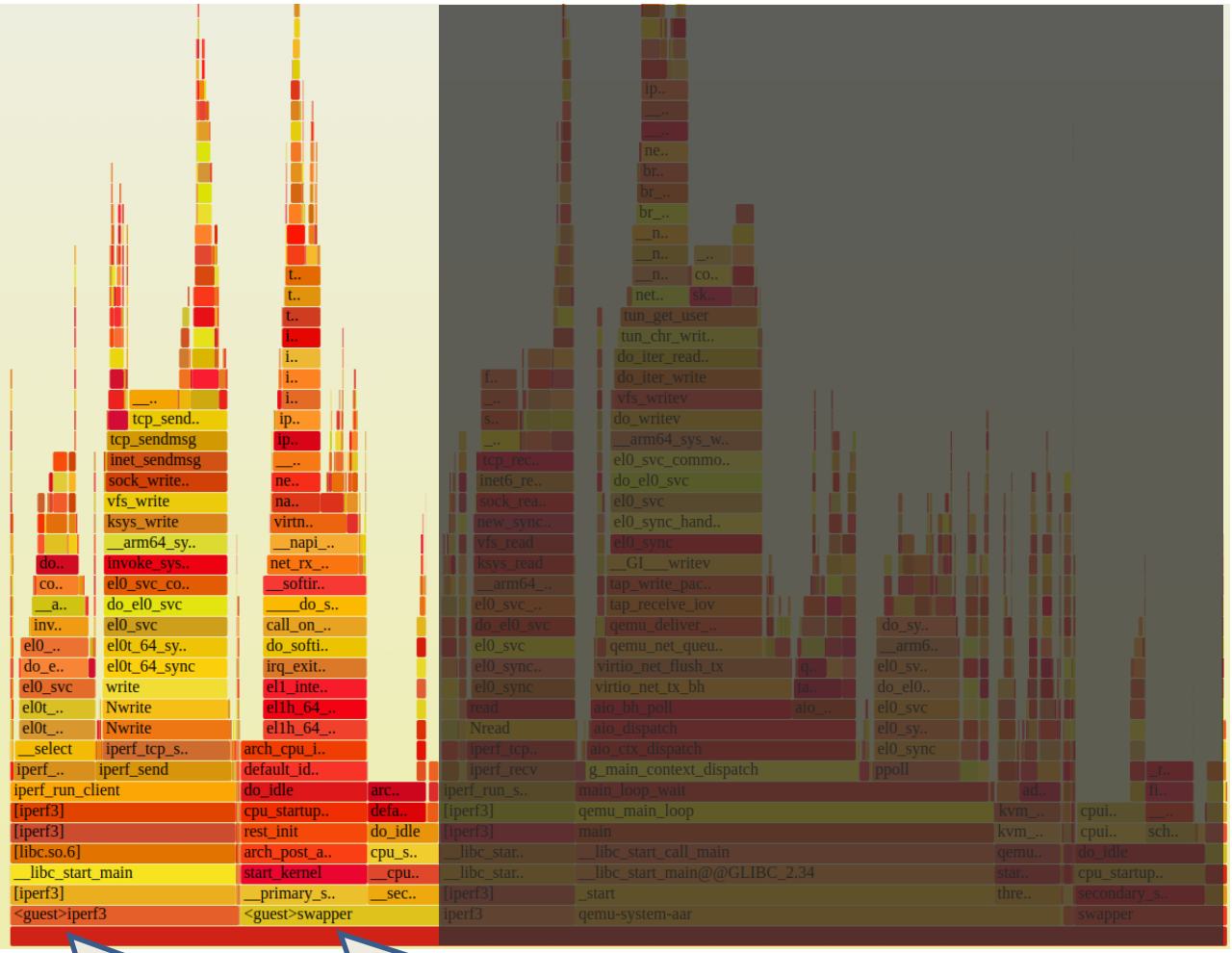


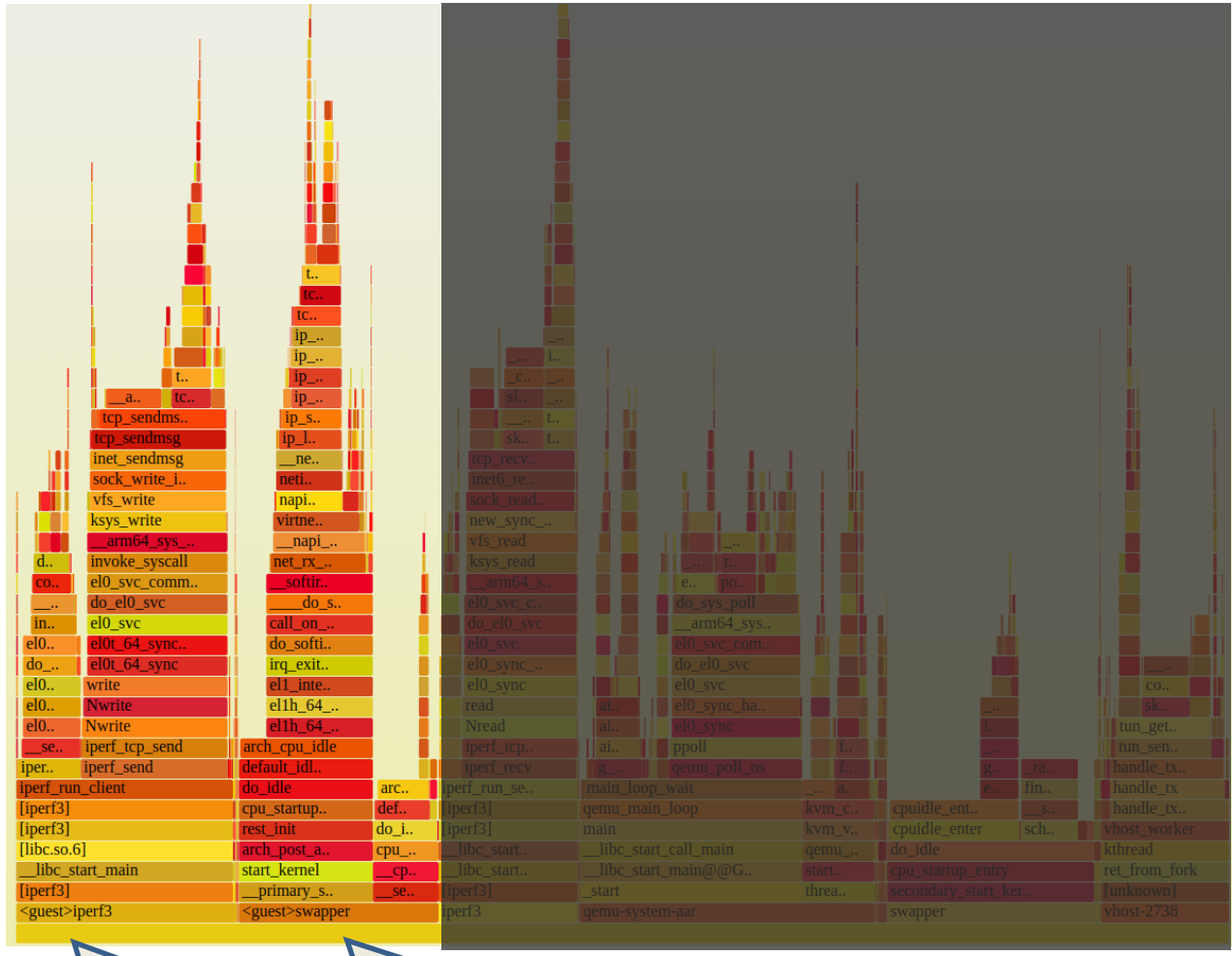
- Analyzing system-wide stack traces (=function call hierarchy)
  - Horizontal axis represents the total duration of the function across all samples
  - Stacks are displayed from bottom to top, with the bottommost element indicating the process name

Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:





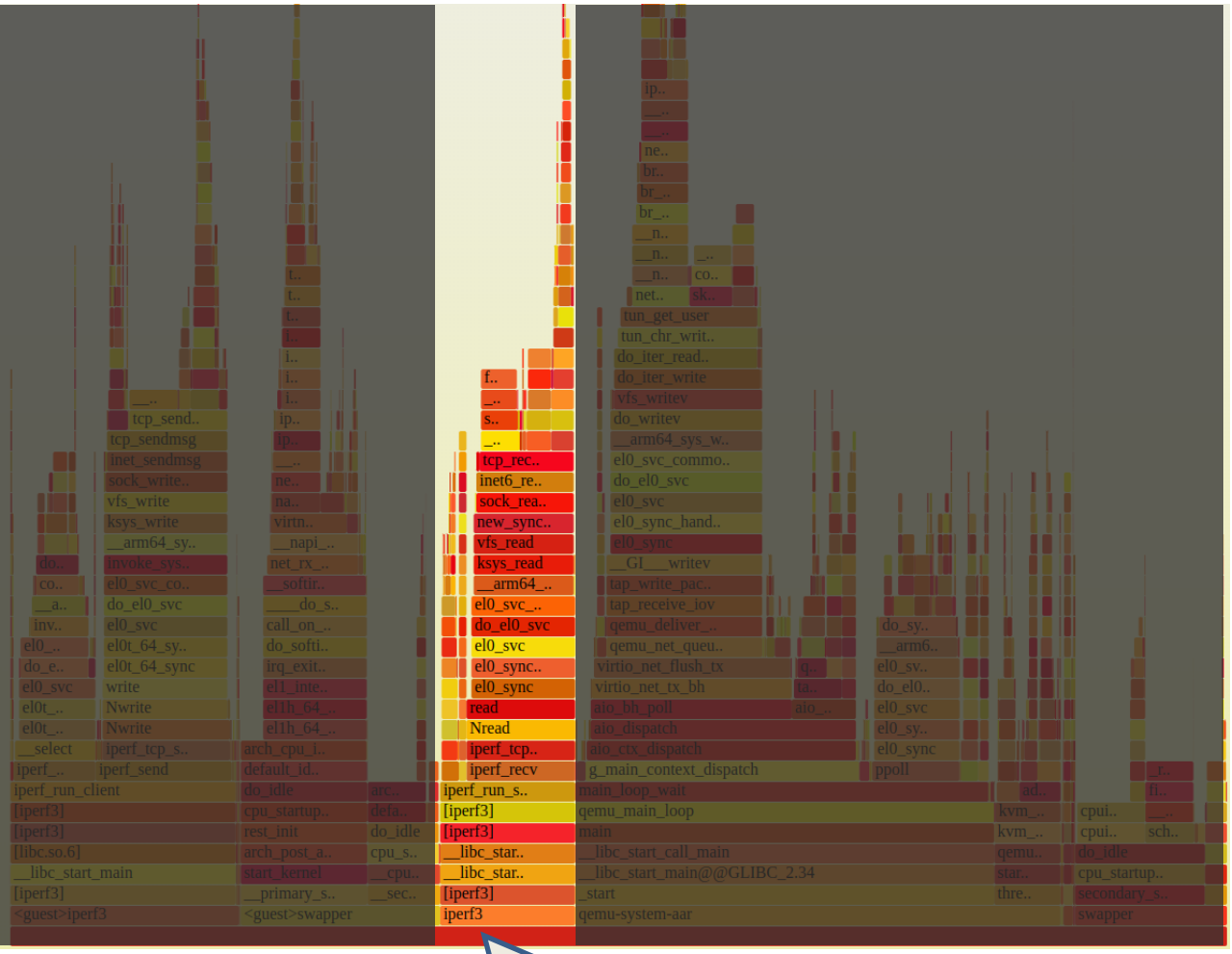
vhost-net:

Guest side: Similar workload between standard virtio and vhost-net

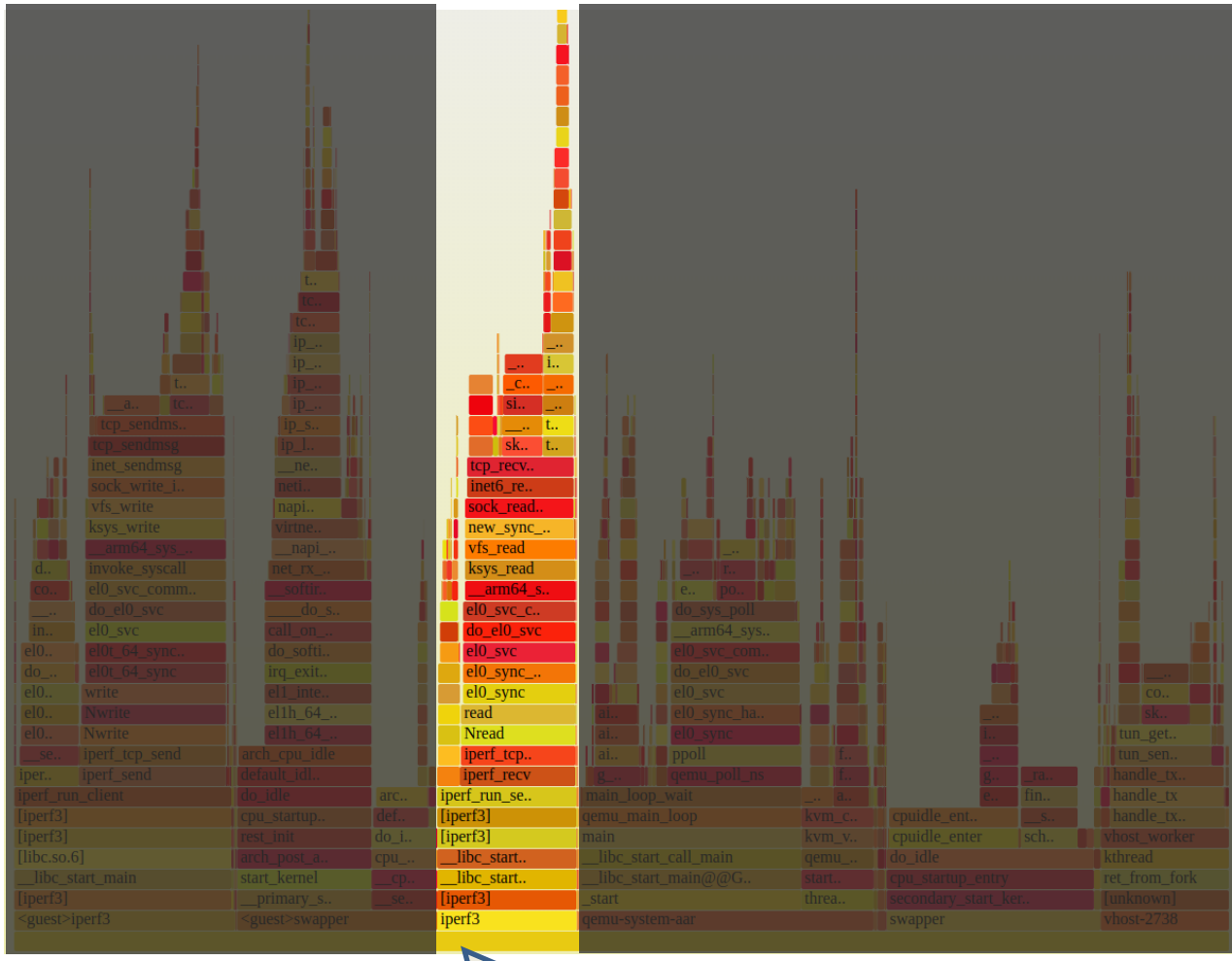
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:

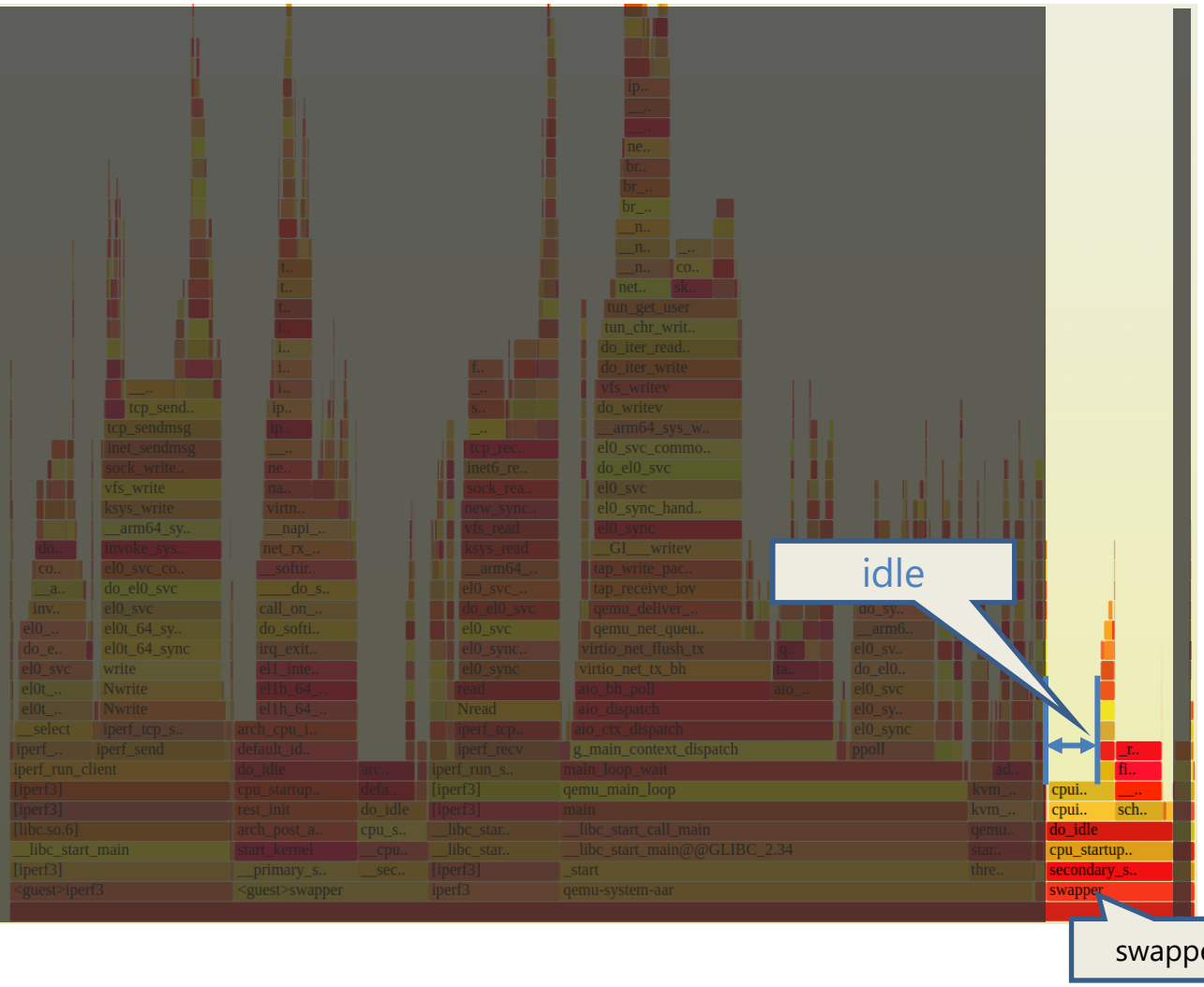


iperf3 on host: Similar workload between standard virtio and vhost-net

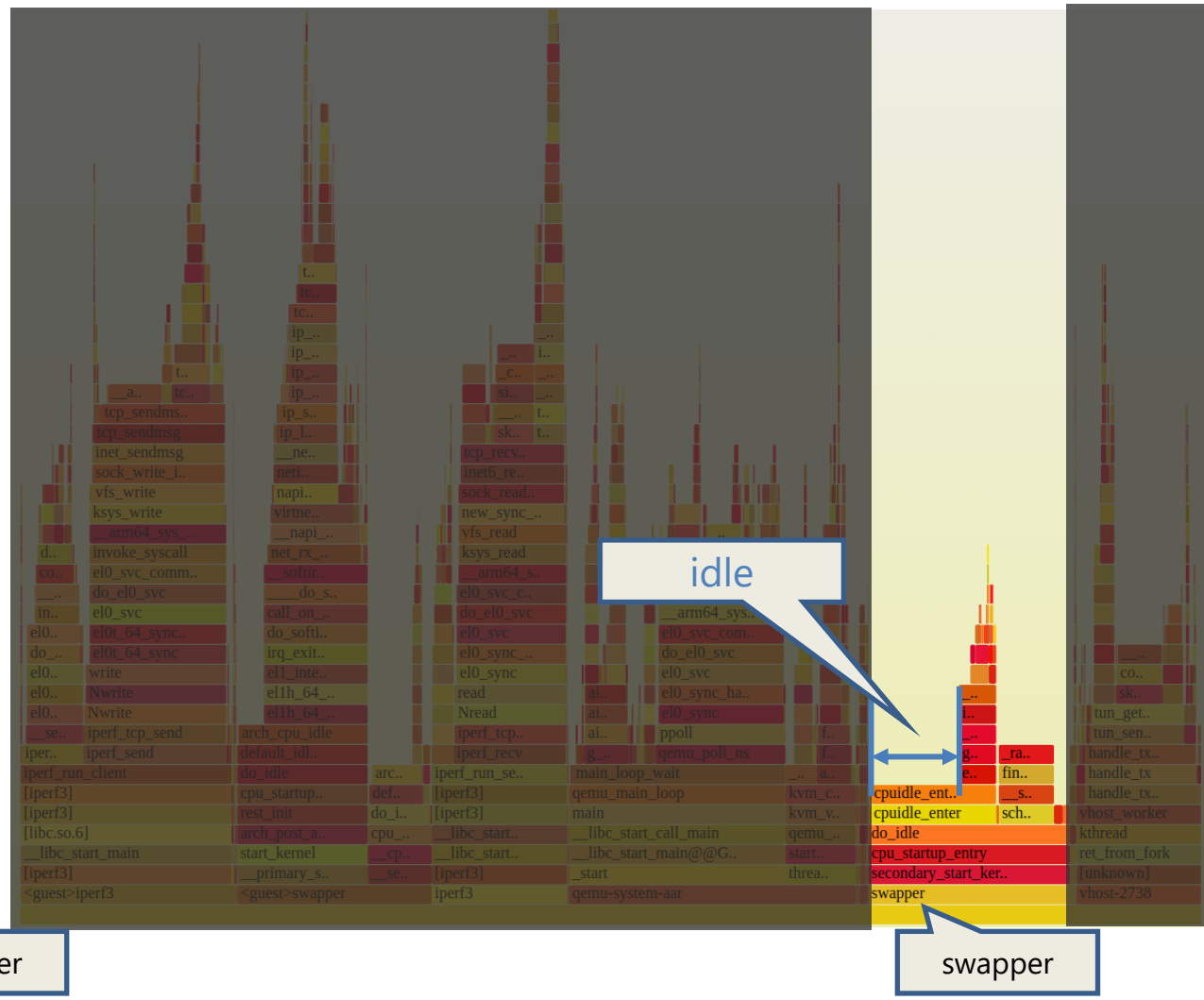
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:

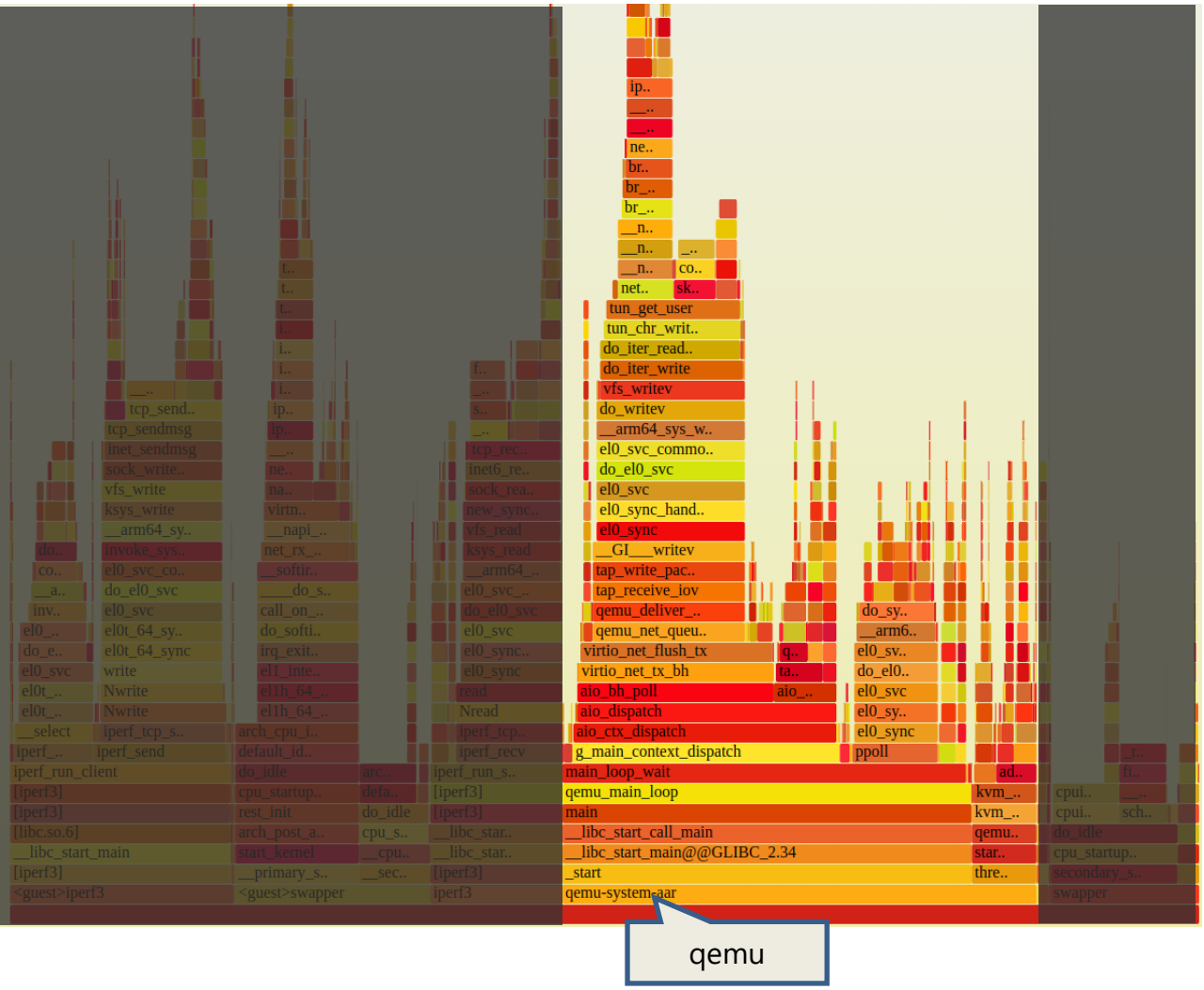


Swapper on host: Somewhat similar, but idle-time is slightly longer in vhost-net

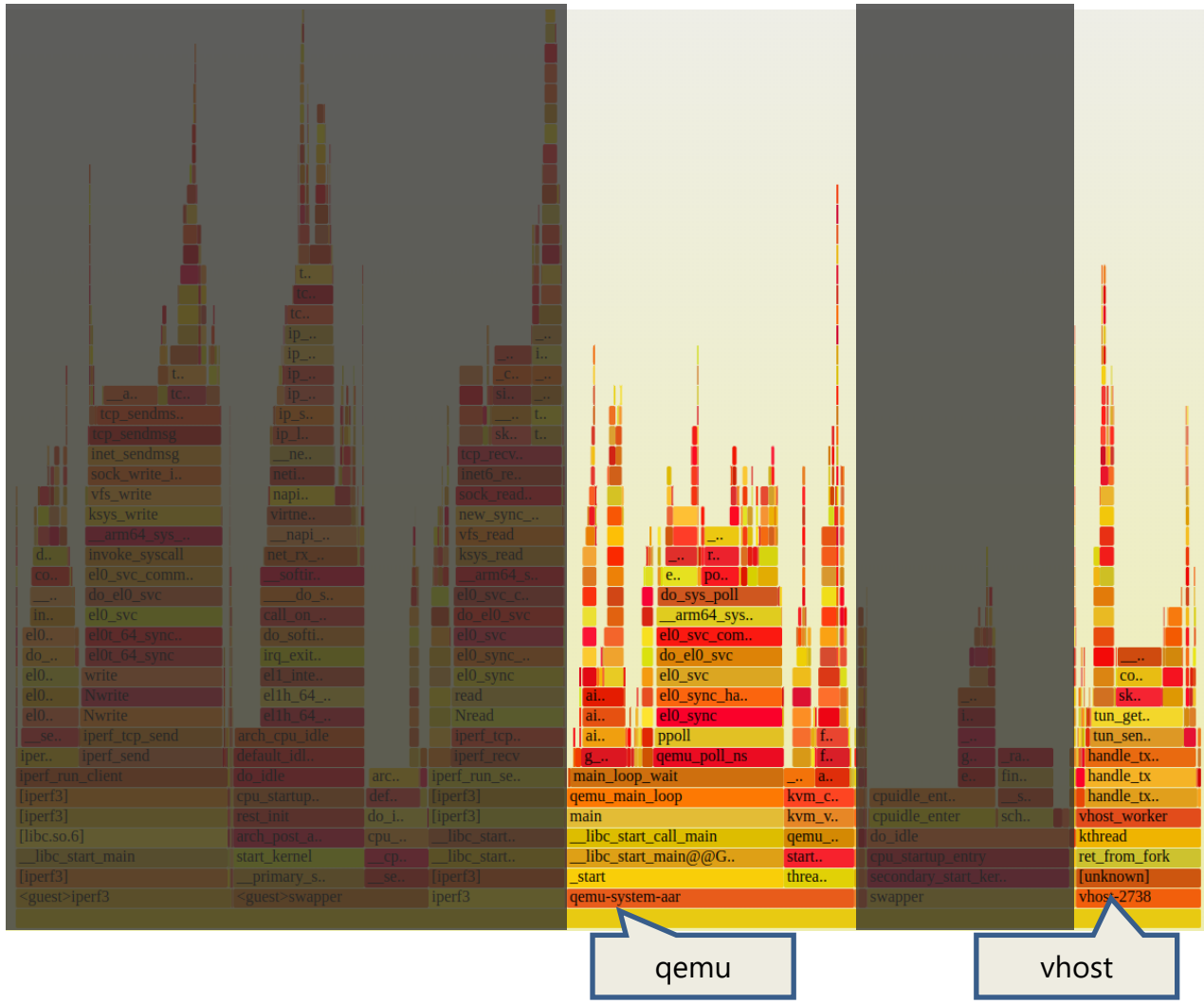
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:

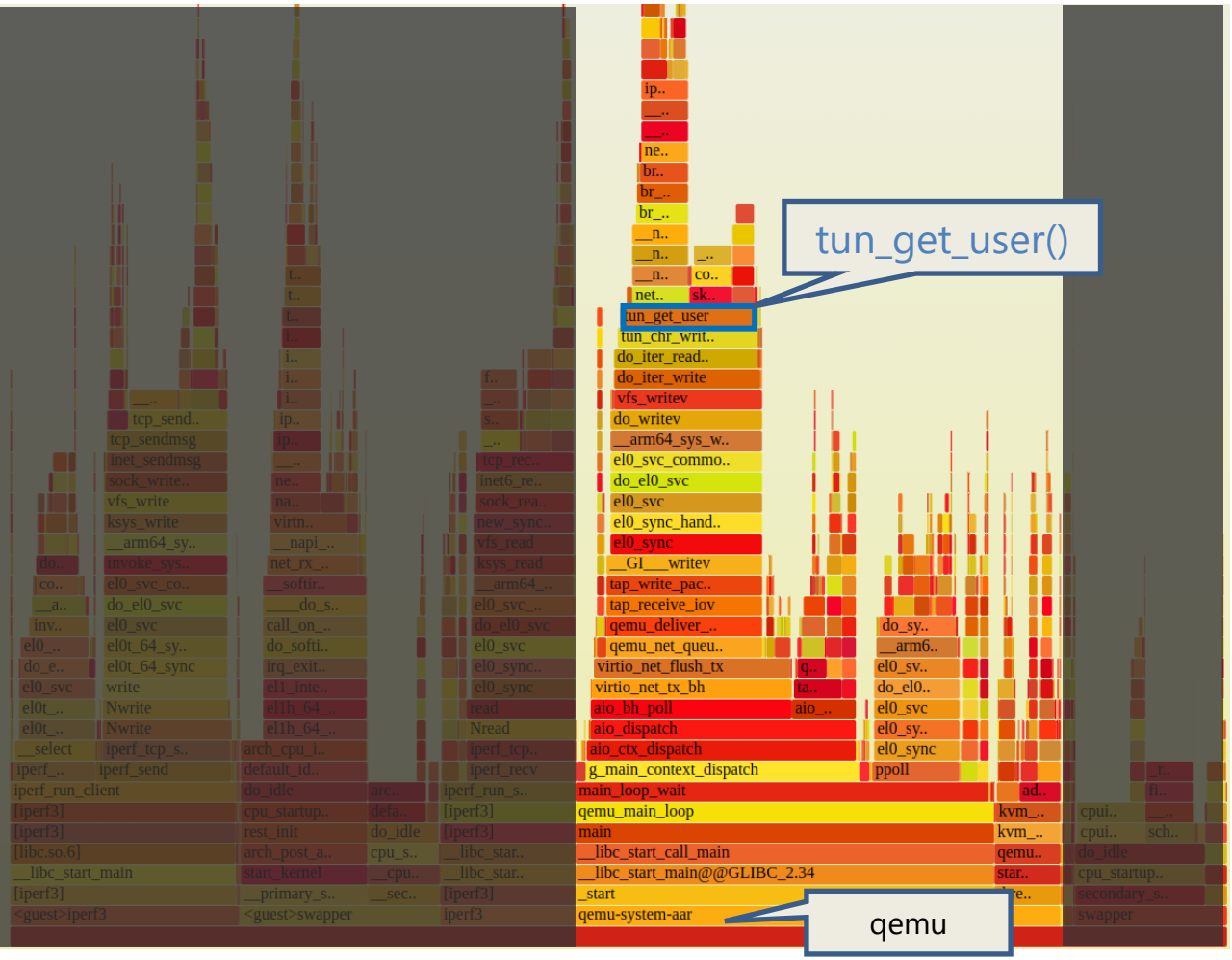


Primary difference lies in qemu and vhost kthread

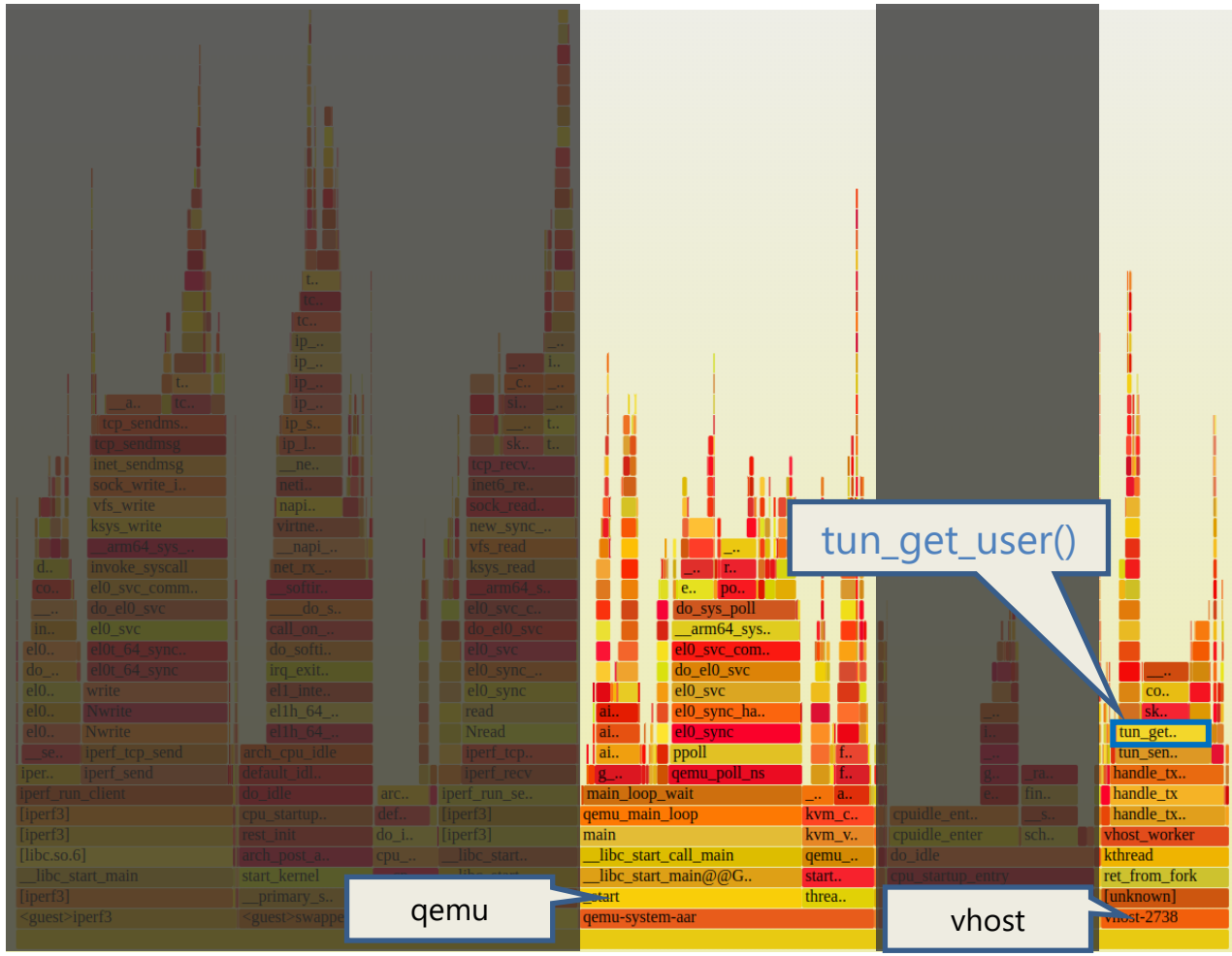
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:

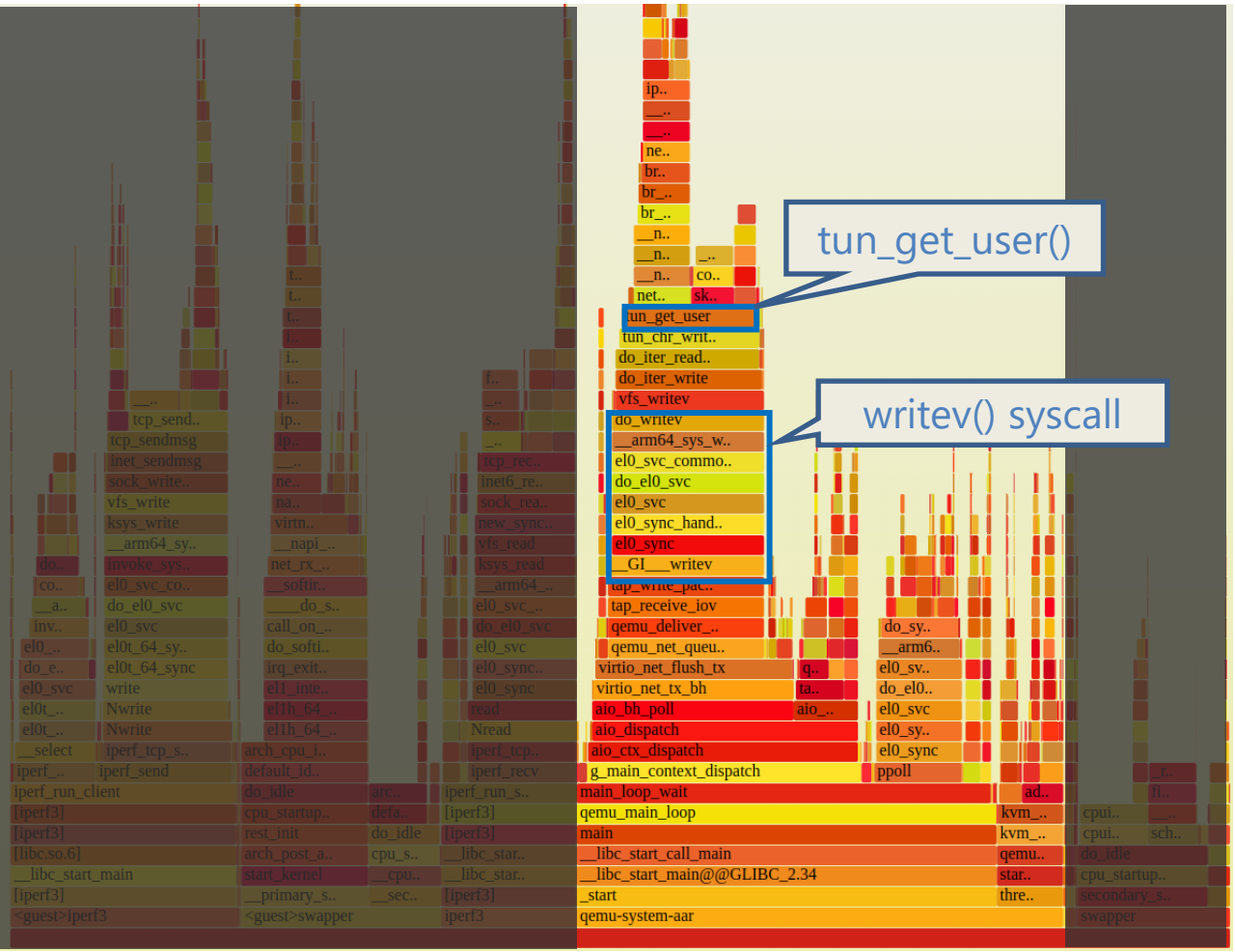


- Same function load (tun_get_user) appears in different contexts (qemu vs vhost)
  - Visualize the shift of network device emulation

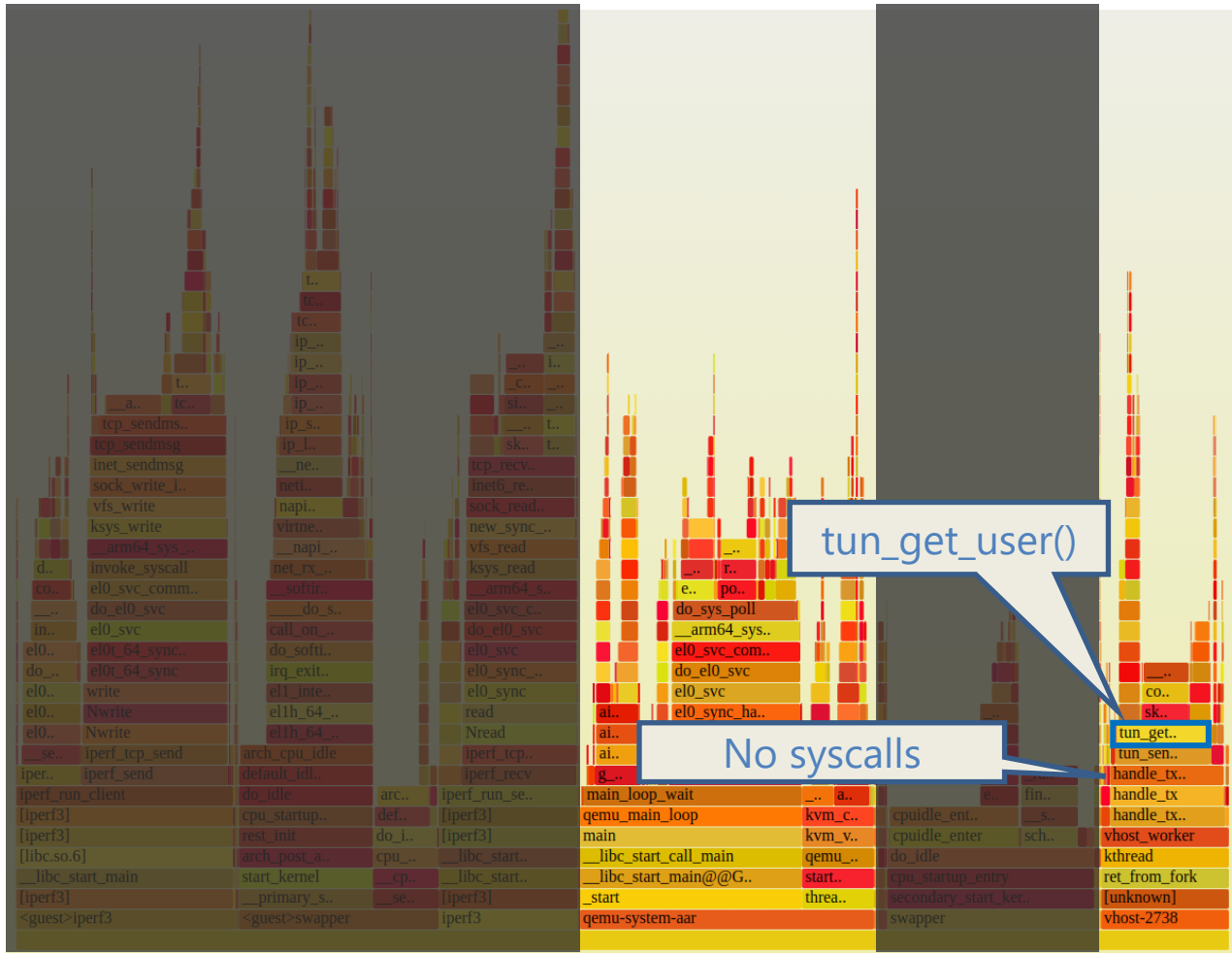
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:

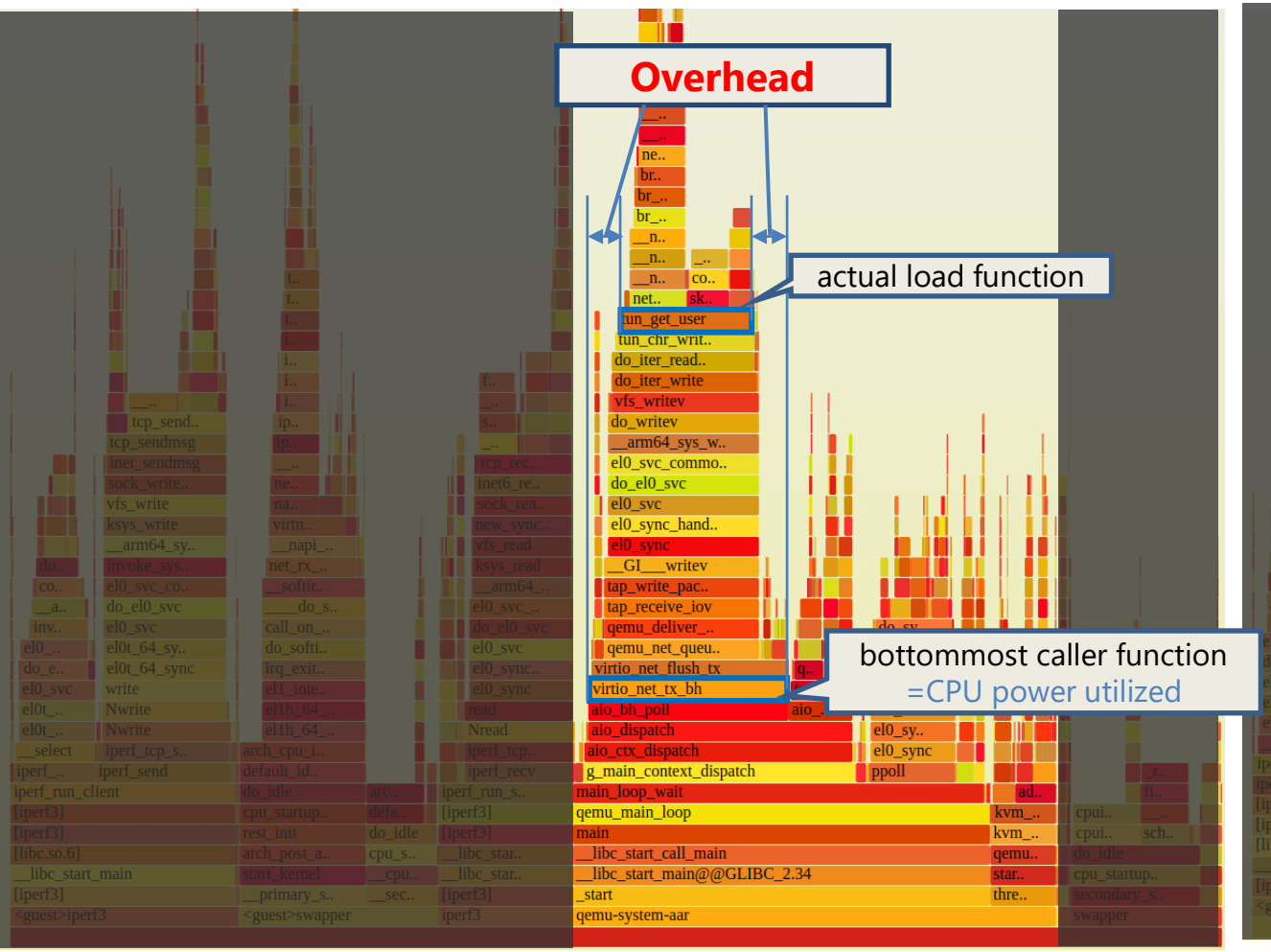


- Standard virtio: As qemu is a user process, it must utilize system calls (writev) to invoke kernel functions (tun_get_user)
- vhost-net: There is no system calls because vhost is a kernel thread

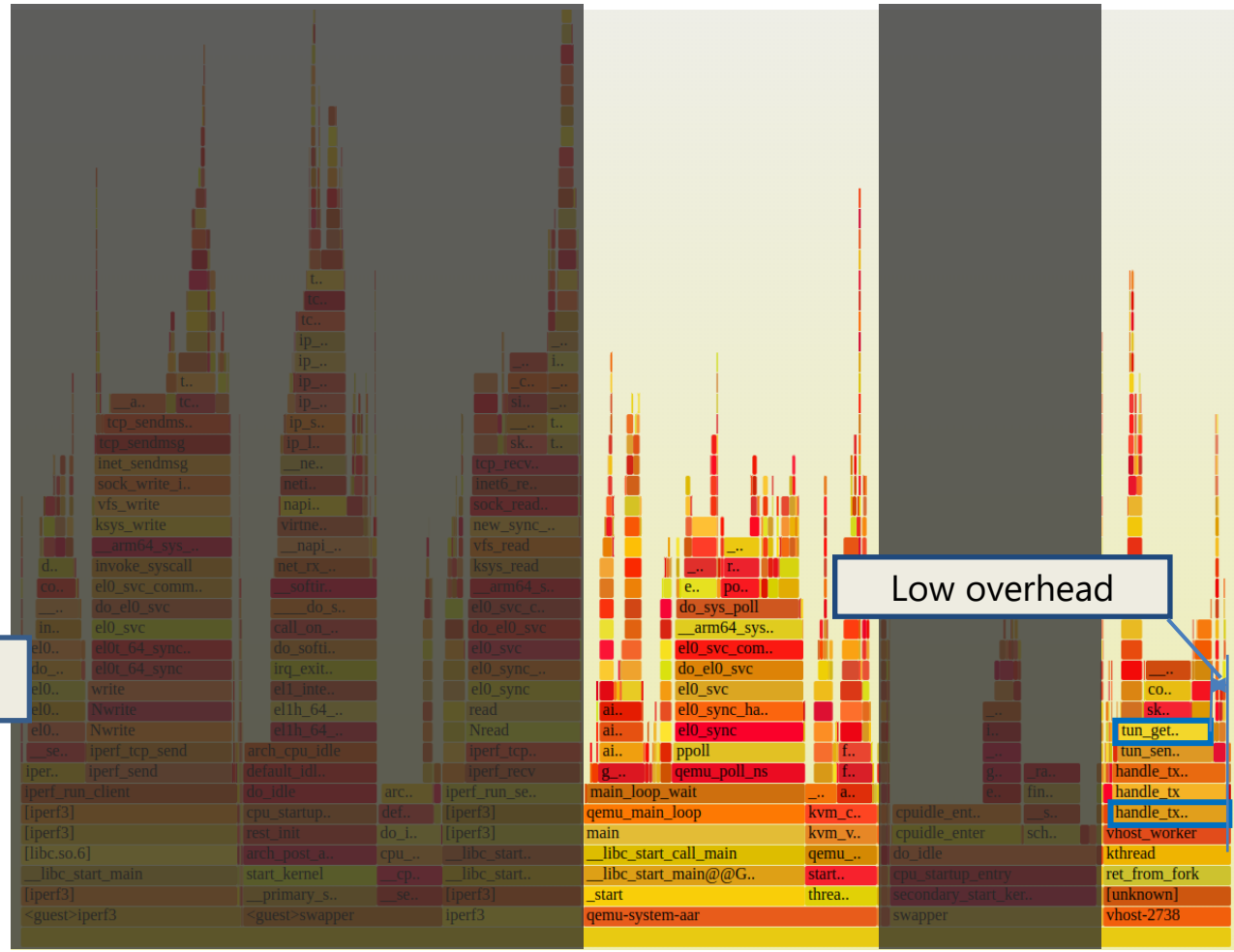
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



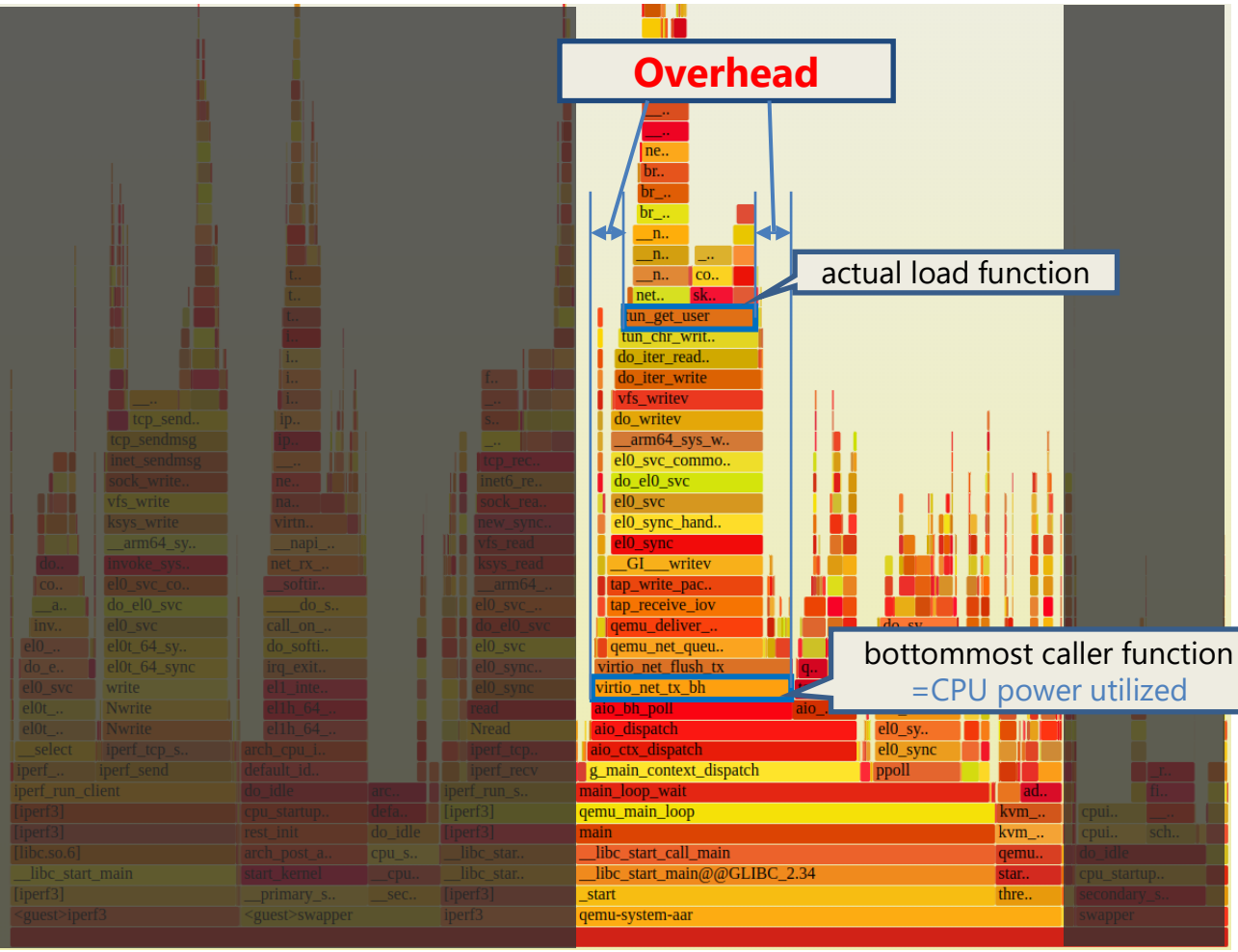
vhost-net:



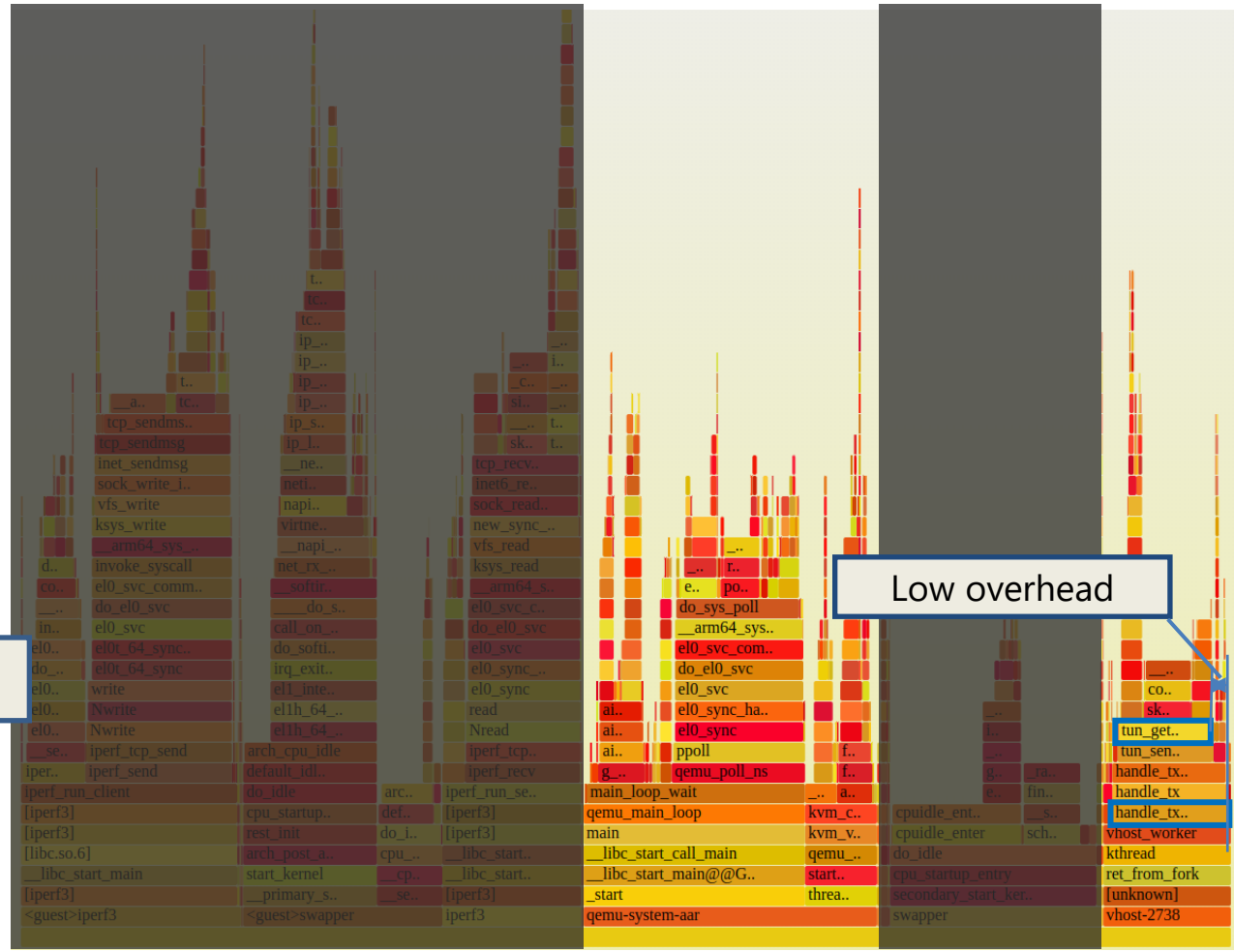
Workload Visualization: `framegraph.pl`

Panasonic AUTOMOTIVE

Standard virtio:



vhost-net:



qemu generates significant overhead during network packet transfer, primarily due to use of system calls

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Panasonic AUTOMOTIVE

- To effectively analyze context switches between host and guest, I recommend tracing both 'kvm' and 'sched' events at once (`trace-cmd -e kvm -e sched`)
- Additionally, kprobes can be utilized to trace almost all kernel functions

Example: Creating and tracing a kprobe event for tun_get_user:

```
# Ensure the target function is traceable
root@host:~# grep tun_get_user /sys/kernel/debug/tracing/available_filter_functions
tun_get_user

# Create a kprobe event for tun_get_user
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events

# Verify that the kprobe event was successfully created
root@host:~# ls /sys/kernel/debug/tracing/events/kprobes/kp_tun_get_user
enable  filter  format  id  trigger

# Trace it with trace-cmd
root@host:~# trace-cmd record -e 'kprobes:kp_tun_get_user' -- sleep 1
```

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Panasonic AUTOMOTIVE

Standard virtio:

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:    CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  75                    <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  75                                 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.92                            CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  759                 923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

CPU6: guest vCPU

CPU7: qemu main (I/O thread)

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

mmio event initiated by guest prompts host to transfer the next packets

```
root@host:~# echo "p:kp_tun_get_user tun_get_user
root@host:~# trace-cmd record -e kvm -e sched -e                                    72.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:                 <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:          ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:                 mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:             comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:             qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:                PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:             swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:                 <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:            guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:        VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:       comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:             qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:        VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:       CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:        VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:          (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:             Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:             comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:             qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:             swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:        VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:        arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:         vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:          wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:                 mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc0
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:    CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

qemu I/O thread awakens

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtim                                                  untime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:    CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

guest vCPU enters sleep state using WFx, awaiting completion of network transfer

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VC
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

qemu I/O thread invokes tun_get_user()

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:    CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

qemu handles IRQ event for the guest

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:    CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0,
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVA
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_in          (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

guest vCPU gets woken up from sleep state

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990: sched_switch:         qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996: kvm_timer_save_state:    CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

guest invokes next mmio event

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Standard virtio:

Panasonic AUTOMOTIVE

```
root@host:~# echo "p:kp_tun_get_user tun_get_user" >> /sys/kernel/debug/tracing/kprobe_events
root@host:~# trace-cmd record -e kvm -e sched -e 'kprobes:kp_tun_get_user' -- ssh guest iperf3 -c 172.16.10.1 -n 100M && trace-cmd report

 qemu-system-aar-1361  [006]  7593.923907: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024
 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923908: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1361  [006]  7593.923910: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1361  [006]  7593.923919: sched_waking:         comm=qemu-system-aar pid=1353 prio=120 target_cpu=007
 qemu-system-aar-1361  [006]  7593.923926: sched_wakeup:         qemu-system-aar:1353 [120] success=1 CPU:007
 qemu-system-aar-1361  [006]  7593.923934: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [007]  7593.923936: sched_switch:         swapper/7:0 [120] R ==> qemu-system-aar:1353 [120]
 qemu-system-aar-1361  [006]  7593.923971: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0001 (WFx), PC: 0xffffffc008be07e8
 qemu-system-aar-1361  [006]  7593.923972: kvm_wfx_arm64:        guest executed wf>c< at: 0x0000000000000000
 qemu-system-aar-1361  [006]  7593.923977: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923986: sched_stat_runtime:   comm=qemu-system-aar pid=1361 runtime=587763 [ns] vruntime=111836102550 [ns]
 qemu-system-aar-1361  [006]  7593.923990:                       qemu-system-aar:1361 [120] S ==> swapper/6:0 [120]
 qemu-system-aar-1361  [006]  7593.923994:                       VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.923996:                          CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.923997: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1353  [007]  7593.924129: kp_tun_get_user:      (ffff800010b335d0)
 qemu-system-aar-1353  [007]  7593.924377: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1353  [007]  7593.924379: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1353  [007]  7593.924386: sched_waking:         comm=qemu-system-aar pid=1361 prio=120 target_cpu=006
 qemu-system-aar-1353  [007]  7593.924392: sched_wakeup:         qemu-system-aar:1361 [120] success=1 CPU:006
          <idle>-0     [006]  7593.924402: sched_switch:         swapper/6:0 [120] R ==> qemu-system-aar:1361 [120]
 qemu-system-aar-1361  [006]  7593.924417: kvm_get_timer_map:    VCPU: 0, dv: 1, dp: -1, ep: 0
 qemu-system-aar-1361  [006]  7593.924419: kvm_timer_update_irq: VCPU: 0, IRQ 27, level 0
 qemu-system-aar-1361  [006]  7593.924420: vgic_update_irq_pending: VCPU: 0, IRQ 27, level: 0
 qemu-system-aar-1361  [006]  7593.924431: kvm_timer_restore_state: CTL: 0x000001 CVAL:        0xf6643c1 arch_timer_ctx_index: 1
 qemu-system-aar-1361  [006]  7593.924432: kvm_timer_emulate:    arch_timer_ctx_index: 0 (should_fire: 0)
 qemu-system-aar-1361  [006]  7593.924440: kvm_halt_poll_ns:     vcpu 0: halt_poll_ns 10000 (grow 0)
 qemu-system-aar-1361  [006]  7593.924441: kvm_vcpu_wakeup:      wait time 462722 ns, polling valid
…
 qemu-system-aar-1361  [006]  7593.924478: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

+411us

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

vhost-net:

```
 qemu-system-aar-1795  [005] 83328.209495: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209496: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209498: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1795  [005] 83328.209512: sched_waking:         comm=vhost-1781 pid=1793 prio=120 target_cpu=001
          <idle>-0     [001] 83328.209522: sched_wakeup:         vhost-1781:1793 [120] success=1 CPU:001
 qemu-system-aar-1795  [005] 83328.209525: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [001] 83328.209525: sched_switch:         swapper/1:0 [120] R ==> vhost-1781:1793 [120]
      vhost-1781-1793  [001] 83328.209533: kp_tun_get_user:      (ffff800010b34550)
      vhost-1781-1793  [001] 83328.209568: sched_waking:         comm=qemu-system-aar pid=1781 prio=120 target_cpu=004
      vhost-1781-1793  [001] 83328.209572: sched_stat_runtime:   comm=vhost-1781 pid=1793 runtime=48480 [ns] vruntime=265274173669 [ns]
      vhost-1781-1793  [001] 83328.209574: sched_switch:         vhost-1781:1793 [120] S ==> swapper/1:0 [120]
          <idle>-0     [004] 83328.209577: sched_wakeup:         qemu-system-aar:1781 [120] success=1 CPU:004
          <idle>-0     [004] 83328.209584: sched_switch:         swapper/4:0 [120] R ==> qemu-system-aar:1781 [120]
 qemu-system-aar-1781  [004] 83328.209649: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1781  [004] 83328.209651: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1795  [005] 83328.209666: kvm_exit:             <CANT FIND FIELD exit_reason>IRQ: HSR_EC: 0x0000 (UNKNOWN), PC: 0xffffffc0080f3298
 qemu-system-aar-1795  [005] 83328.209672: kvm_entry:            PC: 0xffffffc0080f3298
...
 qemu-system-aar-1795  [005] 83328.209685: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

vhost-net:

```
 qemu-system-aar-1795  [005] 83328.209495: kvm_exit:                 4 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209496: kvm_guest_fault:          0675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209498: kvm_mmio:                 mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1795  [005] 83328.209512: sched_waking:             comm=vhost-1781 pid=1793 prio=120 target_cpu=001
          <idle>-0     [001] 83328.209513: sched_wakeup:             vhost-1781:1793 [120] success=1 CPU:001
 qemu-system-aar-1795  [005] 83328.209514: kvm_entry:
          <idle>-0     [001] 83328.209525: sched_switch:
      vhost-1781-1793  [001] 83328.209533: kp_tun_get_user:          (ffff800010b34550)
      vhost-1781-1793  [001] 83328.209568: sched_waking:             comm=qemu-system-aar pid=1781 prio=120 target_cpu=004
      vhost-1781-1793  [001] 83328.209572: sched_stat_runtime:       comm=vhost-1781 pid=1793 runtime=48480 [ns] vruntime=265274173669 [ns]
      vhost-1781-1793  [001] 83328.209574: sched_switch:             vhost-1781:1793 [120] ==> swapper/1:0 [120]
          <idle>-0     [004] 83328.209583: sched_wakeup:             qemu-system-aar:1781:1781 [120] success=1 CPU:004
          <idle>-0     [004] 83328.209584: sched_switch:             swapper/4:0 [120] R ==> qemu-system-aar:1781 [120]
 qemu-system-aar-1781  [004] 83328.209649: kvm_irq_line:             Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1781  [004] 83328.209651: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1795  [005] 83328.209666: kvm_exit:                 <CANT FIND FIELD exit_reason>, PC: 0xffffffc0080f3298
 qemu-system-aar-1795  [005] 83328.209672: kvm_entry:                PC: 0xffffffc0080f3298
...
 qemu-system-aar-1795  [005] 83328.209685: kvm_mmio:                 mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

CPU5: guest vCPU

guest initiates mmio event to prompt vhost to transfer the next packets

vhost awakens

CPU1: vhost

vhost promptly initiates tun_get_user()

vhost wakes up qemu

CPU4: qemu I/O

qemu handles IRQ event for the guest

vCPU doesn't invoke WFx

guest initiates next mmio event

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

vhost-net:

```
 qemu-system-aar-1795  [005] 83328.209495: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209496: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209498: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1795  [005] 83328.209512: sched_waking:         comm=vhost-1781 pid=1793 prio=120 target_cpu=001
         <idle>-0      [001] 83328.209522: sched_wakeup:         vhost-1781:1793 [120] success=1 CPU:001
 qemu-system-aar-1795  [005] 83328.209525: kvm_entry:            PC: 0xffffffc00872cd78
         <idle>-0      [001] 83328.209525: sched_switch:         swapper/1:0 [120] R ==> vhost-1781:1793 [120]
     vhost-1781-1793   [001] 83328.209533: kp_tun_get_user:      (ffff800010b34550)
     vhost-1781-1793   [001] 83328.209568: sched_waking:         comm=qemu-system-aar pid=1781 prio=120 target_cpu=004
     vhost-1781-1793   [001] 83328.209572:                       comm=vhost-1781 pid=1793 runtime=48480 [ns] vruntime=265274173669 [ns]
     vhost-1781-1793   [001] 83328.209574:                       vhost-1781:1793 [120] S ==> swapper/1:0 [120]
         <idle>-0      [004] 83328.209577:                       qemu-system-aar:1781 [120] success=1 CPU:004
         <idle>-0      [004] 83328.209584: sched_switch:         swapper/4:0 [120] R ==> qemu-system-aar:1781 [120]
 qemu-system-aar-1781  [004] 83328.209649: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1781  [004] 83328.209651: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1795  [005] 83328.209666: kvm_exit:             <CANT FIND FIELD exit_reason>IRQ: HSR_EC: 0x0000 (UNKNOWN), PC: 0xffffffc0080f3298
 qemu-system-aar-1795  [005] 83328.209672: kvm_entry:            PC: 0xffffffc0080f3298
...
 qemu-system-aar-1795  [005] 83328.209685: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

+187us

Tracing Interruction Between Host and Guest: `trace-cmd -e kvm record`

Panasonic AUTOMOTIVE

vhost-net:

```
 qemu-system-aar-1795  [005] 83328.209495: kvm_exit:             <CANT FIND FIELD exit_reason>TRAP: HSR_EC: 0x0024 (DABT_LOW), PC: 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209496: kvm_guest_fault:      ipa 0xa003000, hsr 0x93810046, hxfar 0xffffffc009675a50, pc 0xffffffc00872cd74
 qemu-system-aar-1795  [005] 83328.209498: kvm_mmio:             mmio write len 4 gpa 0xa003a50 val 0x1
 qemu-system-aar-1795  [005] 83328.209512: sched_waking:         comm=vhost-1781 pid=1793 prio=120 target_cpu=001
          <idle>-0     [001] 83328.209522: sched_wakeup:         vhost-1781:1793 [120] success=1 CPU:001
 qemu-system-aar-1795  [005] 83328.209525: kvm_entry:            PC: 0xffffffc00872cd78
          <idle>-0     [001] 83328.209525: sched_switch:         swapper/1:0 [120] R ==> vhost-1781:1793 [120]
      vhost-1781-1793  [001] 83328.209533: kp_tun_get_user:      (ffff800010b34550)
      vhost-1781-1793  [001] 83328.209568: sched_waking:         comm=qemu-system-aar pid=1781 prio=120 target_cpu=004
      vhost-1781-1793  [001] 83328.209572:                       comm=vhost-1781 pid=1793 runtime=48480 [ns] vruntime=265274173669 [ns]
      vhost-1781-1793  [001] 83328.209574:                       vhost-1781:1793 [120] S ==> swapper/1:0 [120]
          <idle>-0     [004] 83328.209577:                       qemu-system-aar:1781 [120] success=1 CPU:004
          <idle>-0     [004] 83328.209584: sched_switch:         swapper/4:0 [120] R ==> qemu-system-aar:1781 [120]
 qemu-system-aar-1781  [004] 83328.209649: kvm_irq_line:         Inject VGIC SPI interrupt (1), vcpu->idx: 0, num: 77, level: 1
 qemu-system-aar-1781  [004] 83328.209651: vgic_update_irq_pending: VCPU: 0, IRQ 77, level: 1
 qemu-system-aar-1795  [005] 83328.209666: kvm_exit:             <CANT FIND FIELD exit_reason>IRQ: HSR_EC: 0x0000 (UNKNOWN), PC: 0xffffffc0080f3298
 qemu-system-aar-1795  [005] 83328.209672: kvm_entry:            PC: 0xffffffc0080f3298
...
 qemu-system-aar-1795  [005] 83328.209685: kvm_mmio:             mmio unsatisfied-read len 4 gpa 0xa003a60 val 0x0
```

+187us

- Observed over 2 times time efficiency in vhost-net, as seen in the sequence trace
  - Highly consistent with bitrate measurements obtained using iperf3
- Essential reasons for this improvement would likely include:
  - Removal of qemu userspace (system call) from the critical path
  - Minimization of vCPU sleep & wakeup occurrences

Agenda

- Introduction
- Debugging Tools for Virtualized Systems
- Practice: Analyzing vhost-net
- Summary

Summary

Panasonic AUTOMOTIVE

- Confirmed the value of vhost-net by the bahevior analysis using specialized tools for virtualized environments
  - Twice as fast as standard virtio-net
  - No significant side effects, especially in terms of CPU utilization
- Various methodologies and tools are covered in this talk, they can be applied to a wide range of development scenarios for virtualized systems
  - Analyzing performance improvements/regressions
  - Debugging/walking through virtualized systems

Next Steps

- Contribute materials to AGL
  - Debugging tools
  - vhost-net support
- Explore additional tools
  - Port bcc scripts to aarch64 (e.g., kvmexit)
  - Develop custom tools designed for virtualized systems using bcc
  - Experiment with trace-cmd agent/listen
- Investigation on containerized and cloud environments

Panasonic
AUTOMOTIVE

Thank You!

Appendix

Setting Up Tools on AGL: Installation Steps of Tools

Panasonic
AUTOMOTIVE

```
cat << EOF >> conf/local.conf

# perf, trace-cmd
IMAGE_FEATURES =+ "debug-tweaks tools-debug tools-profile“

# bcc
IMAGE_INSTALL:append = " bcc"

# debuginfod
PACKAGECONFIG:append:pn-elfutils-native = "debuginfod libdebuginfod"
DISTRO_FEATURES:append = " debuginfod“

# tweak for debuginfod: we need to mask following recipe as it disrupts
# gdb's packageconfig and leads to a build error
BBMASK += "meta-qt5/recipes-devtools/gdb“

# ssh/sshfs
IMAGE_FEATURES:append = " ssh-server-dropbear"
IMAGE_INSTALL:append = " sshfs-fuse"

# iperf3
IMAGE_INSTALL:append = " iperf3“
EOF
```

Setting Up Tools on AGL: Adding Kernel Configuration Fragments

- Enable kernel configs for tracing

```
$ cat << EOF >> conf/local.conf
IMAGE_INSTALL:append = " kernel-module-kheaders“
FILESEXTRAPATHS:prepend:pn-linux-renesas := "/path/to/put/kconfig:"
FILESEXTRAPATHS:prepend:pn-linux-yocto := "/path/to/put/kconfig:"
SRC_URI:append:pn-linux-renesas = " file://trace.cfg"
SRC_URI:append:pn-linux-yocto = " file://trace.cfg"
EOF

$ cat /path/to/put/kconfig/trace.cfg
trace.cfg
CONFIG_PERF_EVENTS=y
CONFIG_KPROBES=y
CONFIG_FTRACE=y
CONFIG_DYNAMIC_FTRACE=y
CONFIG_FUNCTION_TRACER=y
CONFIG_FUNCTION_GRAPH_TRACER=y
CONFIG_IRQSOFF_TRACER=y
CONFIG_PREEMPT_TRACER=y
CONFIG_SCHED_TRACER=y
CONFIG_FTRACE_SYSCALLS=y
CONFIG_TRACER_SNAPSHOT=y
CONFIG_PSTORE_FTRACE=y
CONFIG_BPF=y
CONFIG_BPF_SYSCALL=y
CONFIG_BPF_JIT=y
CONFIG_BPF_EVENTS=y
CONFIG_TASKSTATS=y
CONFIG_SCHEDSTATS=y
CONFIG_STACKTRACE=y
CONFIG_IKHEADERS=m
```

Upgrading Tool Recipes

Panasonic AUTOMOTIVE

- Install the latest versions of `trace-cmd` and `bcc`
- Resolve issues related to building `bcc` against older kernel versions

```
# Add new layer
$ source agl-init-build-env
$ bitbake-layers create-layer ../meta-custom-tools
$ bitbake-layers add-layer ../meta-custom-tools
$ mkdir -p ../meta-custom-tools/recipes-devtools ../meta-custom-tools/recipes-kernel

# Clone latest recipes of the tools
$ git clone https://git.yoctoproject.org/poky /tmp/poky && git -C /tmp/poky checkout 34004afb65
$ git clone https://github.com/openembedded/meta-openembedded /tmp/meta-openembedded && \
  git -C /tmp/meta-openembedded checkout def4759e9
$ git clone https://github.com/kraj/meta-clang /tmp/meta-clang && git -C /tmp/meta-clang checkout 384dc8f

# Copy the updated recipes to the new layer
$ cp -r /tmp/meta-openembedded/meta-oe/recipes-kernel/trace-cmd ../meta-custom-tools/recipes-kernel
$ cp -r /tmp/meta-openembedded/meta-oe/recipes-kernel/libtracefs ../meta-custom-tools/recipes-kernel
$ cp -r /tmp/poky/meta/recipes-kernel/libtraceevent ../meta-custom-tools/recipes-kernel

$ cp -r /tmp/meta-clang/dynamic-layers/openembedded-layer/recipes-devtools/bcc ../meta-custom-tools/recipes-devtools
$ cp -r /tmp/meta-openembedded/meta-oe/recipes-kernel/libbpf ../meta-custom-tools/recipes-kernel
```

Miscellaneous Setup

Panasonic AUTOMOTIVE

- Setting up vhost-net

On the host PC (before building):

```
$ cat << EOF >> conf/local.conf
PACKAGECONFIG:append:pn-qemu = " vhost“

IMAGE_INSTALL:append = " kernel-module-vhost-net“
FILESEXTRAPATHS:prepend:pn-linux-renesas := "/path/to/put/kconfig:"
SRC_URI:append:pn-linux-renesas = " file://vhost.cfg"
EOF

$ cat /path/to/put/kconfig/vhost.cfg
CONFIG_VHOST_NET=m
```

On target host system:

```
root@host:~# qemu-system-aarch64 \
    -enable-kvm \
    -cpu host \
    -netdev tap,helper="/usr/libexec/qemu-bridge-helper -
br=vmnet0",id=net0,vhost=on \
[...]
```

- Setting up debuginfod

On the host PC:

```
$ source /path/to/agl-init-build-env
$ oe-debuginfod
```

On target (both host and guest):

```
root@host:~# export DEBUGINFOD_URLS=http://<ip-of-host-pc>:8002
```

```
root@guest:~# export DEBUGINFOD_URLS=http://<ip-of-host-pc>:8002
```