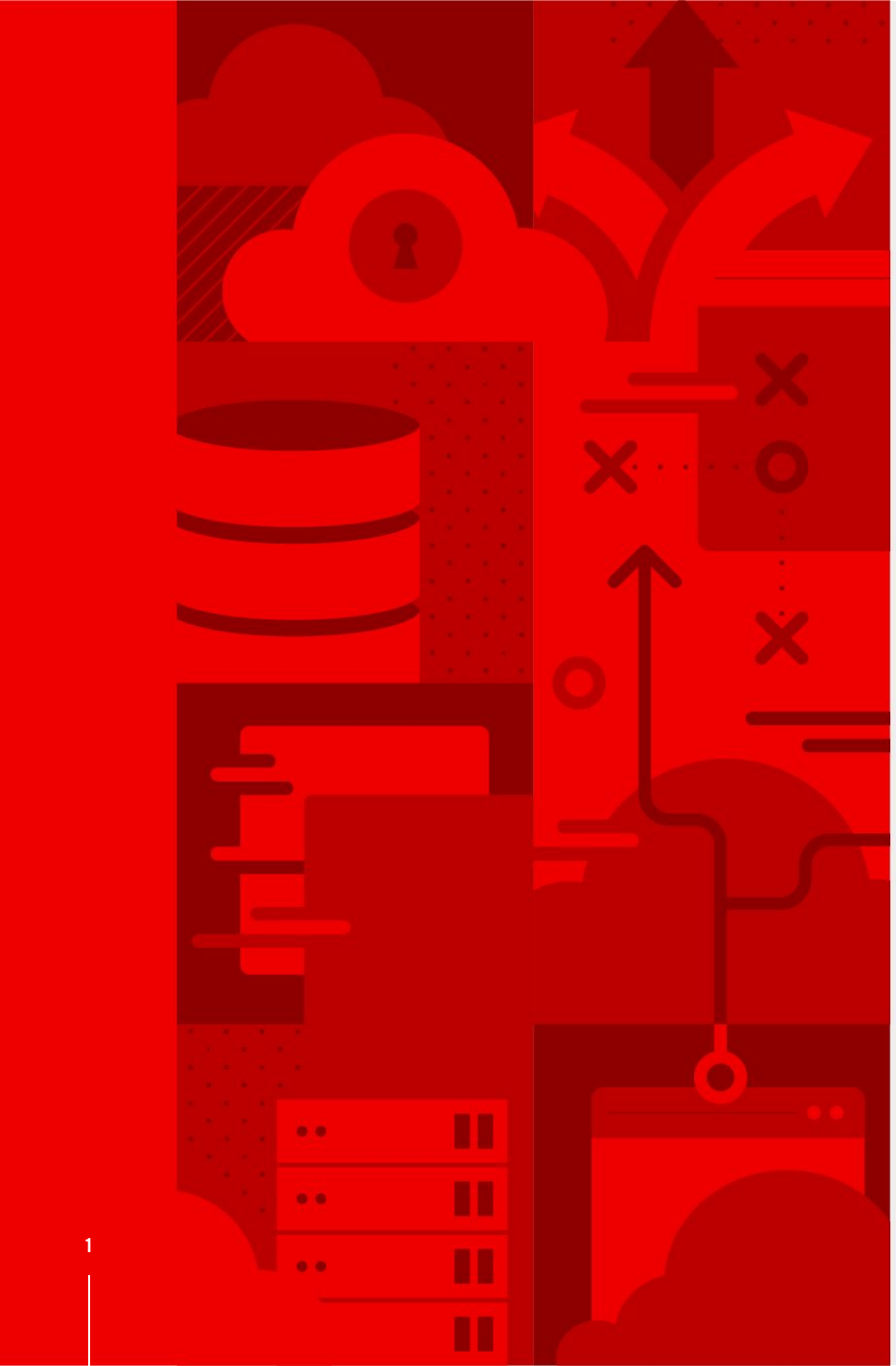


CPU isolation vs jailbreaking IPIs

Valentin Schneider <vschneid@redhat.com>

LPC 2022

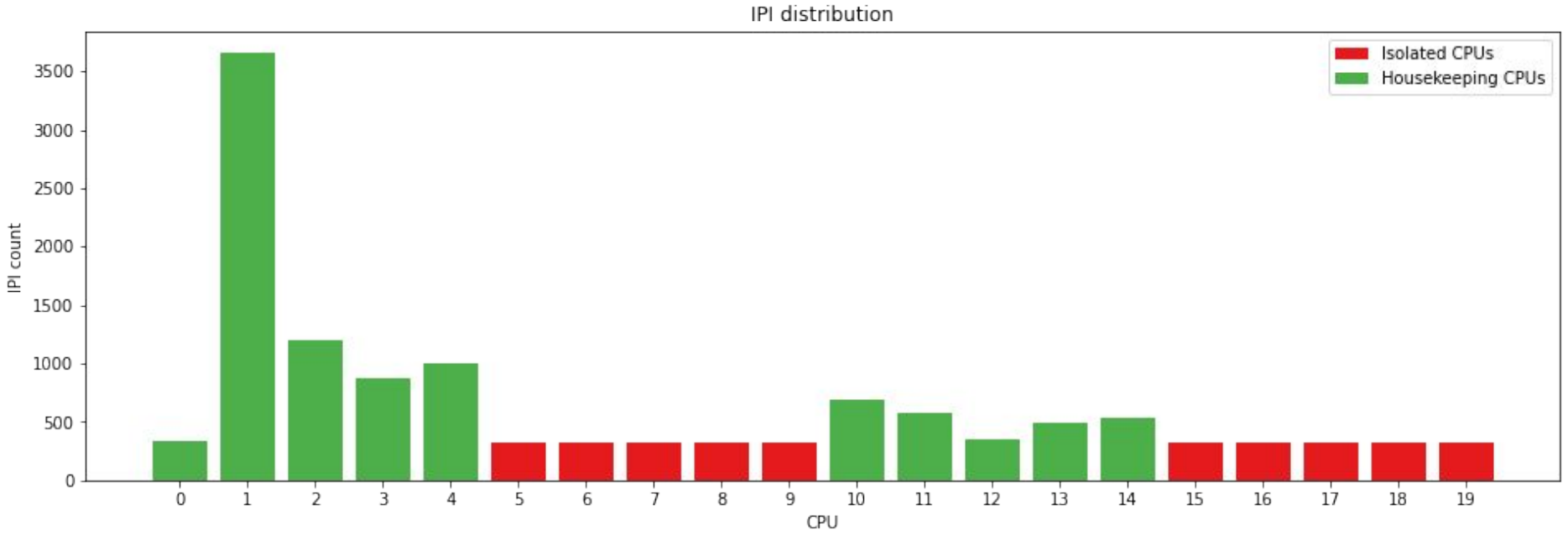
Red Hat

Context

- Isolated CPUs + nohz_full
- Looking at IPIs
  - Software-visible IPIs, so **smp_call** / **irq_work**

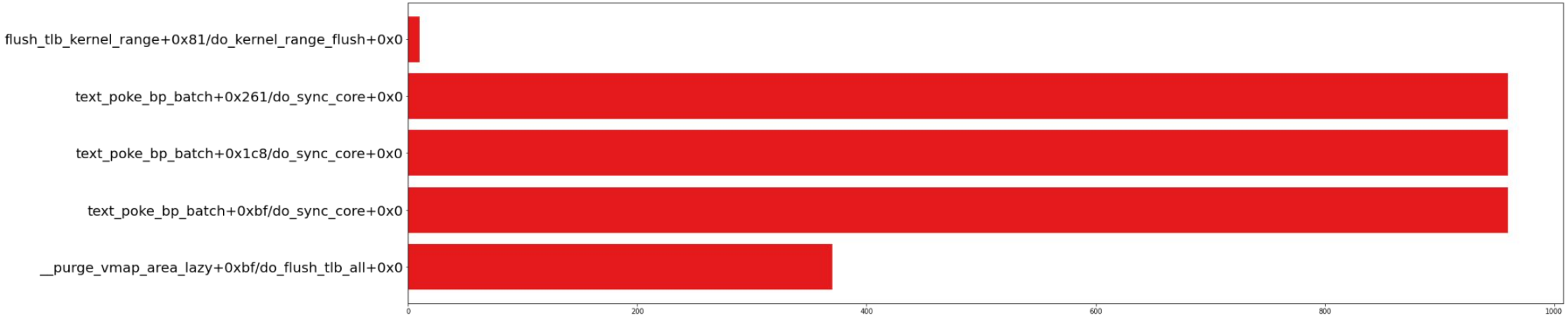
- **rteval** + single userspace **busy-loop** pinned on each isolated CPU

- Expectations: no smp_calls targeting isolated CPUs
- Reality over 20 minutes of tracing **smp_calls** (v5.19-rc2, x86):



[1]: https://lore.kernel.org/all/20220628131619.2109651-1-frederic@kernel.org/
[2]: https://lore.kernel.org/lkml/20220503100051.2799723-1-frederic@kernel.org/

Closer look into the IPIs



- Major culprit is x86 instruction patching (static key)
- Second one is (kernel) TLB invalidation

- IPI deferral patches are out there:
  - [1] rcu/context-tracking: Merge RCU eqs-dynticks counter to context tracking
  - [2] context_tracking,livepatch: Dont disturb NOHZ_FULL

- This is all reactive though:

```
$ git grep -Ir "smp_call*" | wc -l
545
```


[1]: https://lore.kernel.org/all/20220628131619.2109651-1-frederic@kernel.org/
[2]: https://lore.kernel.org/all/20210929151723.162004989@infradead.org/

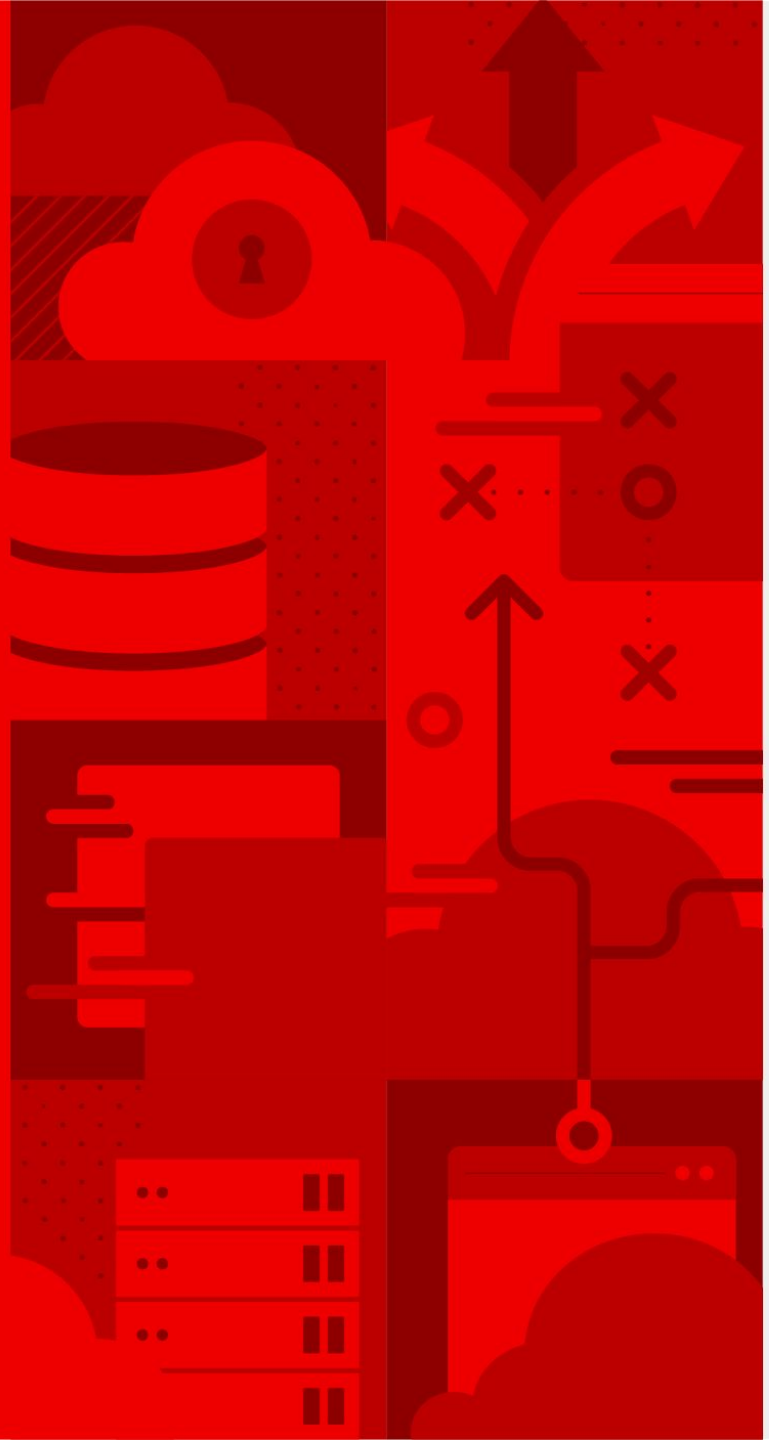
smp_call classification

- **Coccinelle**: Detect isolation cpu(mask) check in callstack leading to smp_call
  - Reduces the search space by… About 3 files
  - Hard to detect “good by construction” callsites (e.g. *sched/rt:tell_cpu_to_push()*)
  - Still > 400 sites

- **Manual classification**
  - Remote data fetch
    - x86: cpuid_read()
    - perf:perf_event_read()
    - arm64: counters_read_on_cpu()
  - Running task synchronization
    - ftrace: event_pid_write()
    - resctrl: rdtgroup_move_task()
  - System-wide synchronization
    - Instruction patching (x86:test_poke_bp_batch())
    - flush_tlb_kernel_range() (x86, mips)
    - mm:do_tune_cpucache()

- “Don’t do this” vs expectations for the kernel

- Encode **intent** in callsite
  - smp_call_data_fetch()
    - Return error/default value
    - Issue IPI WARN_ONCE
  - smp_call_current_task()
    - Wait until next context transition
    - Issue IPI but WARN_ONCE
  - ??? (ask Santa)



(More) Questions?

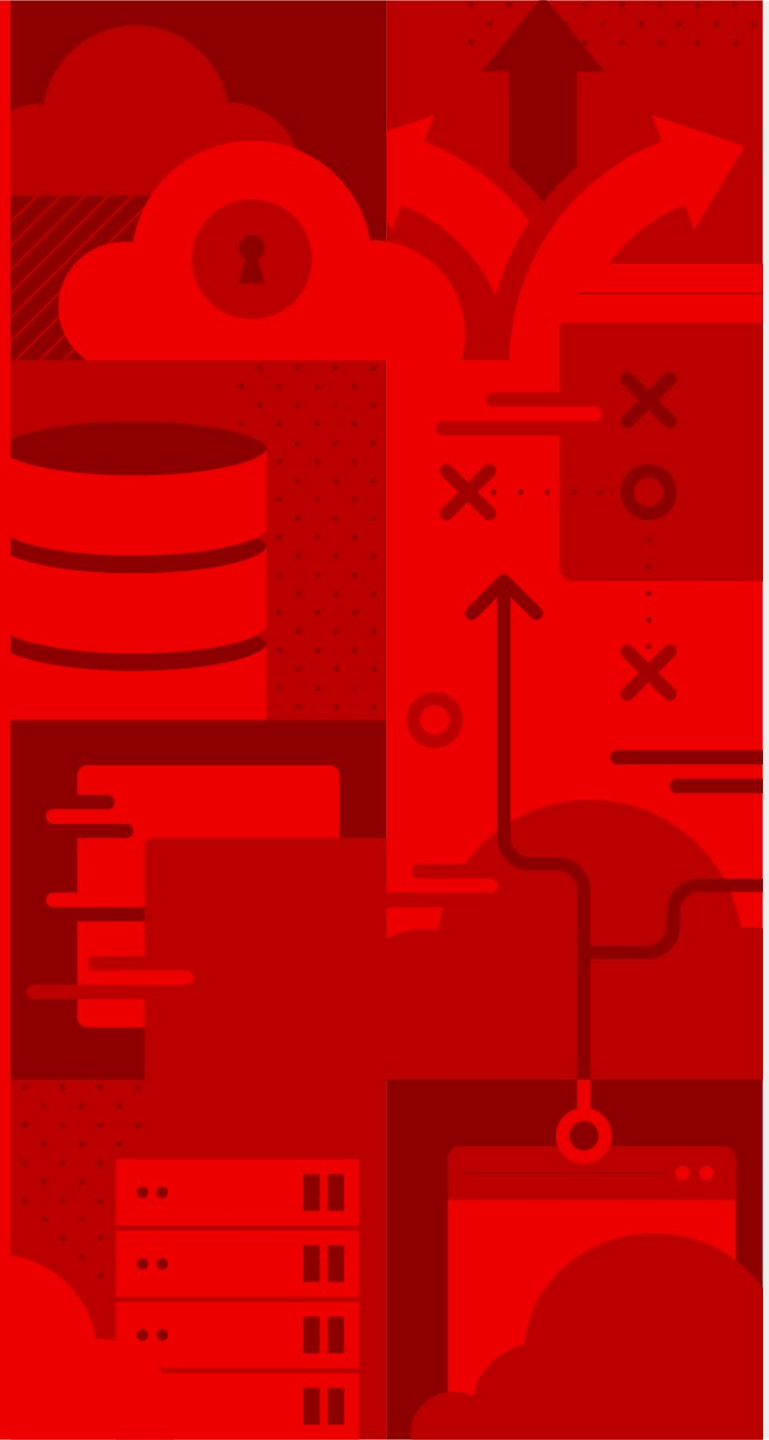
linkedin.com/company/red-hat

facebook.com/redhatinc

youtube.com/user/RedHatVideos

twitter.com/RedHat

Red Hat



Thanks!

Valentin Schneider <vschneid@redhat.com>

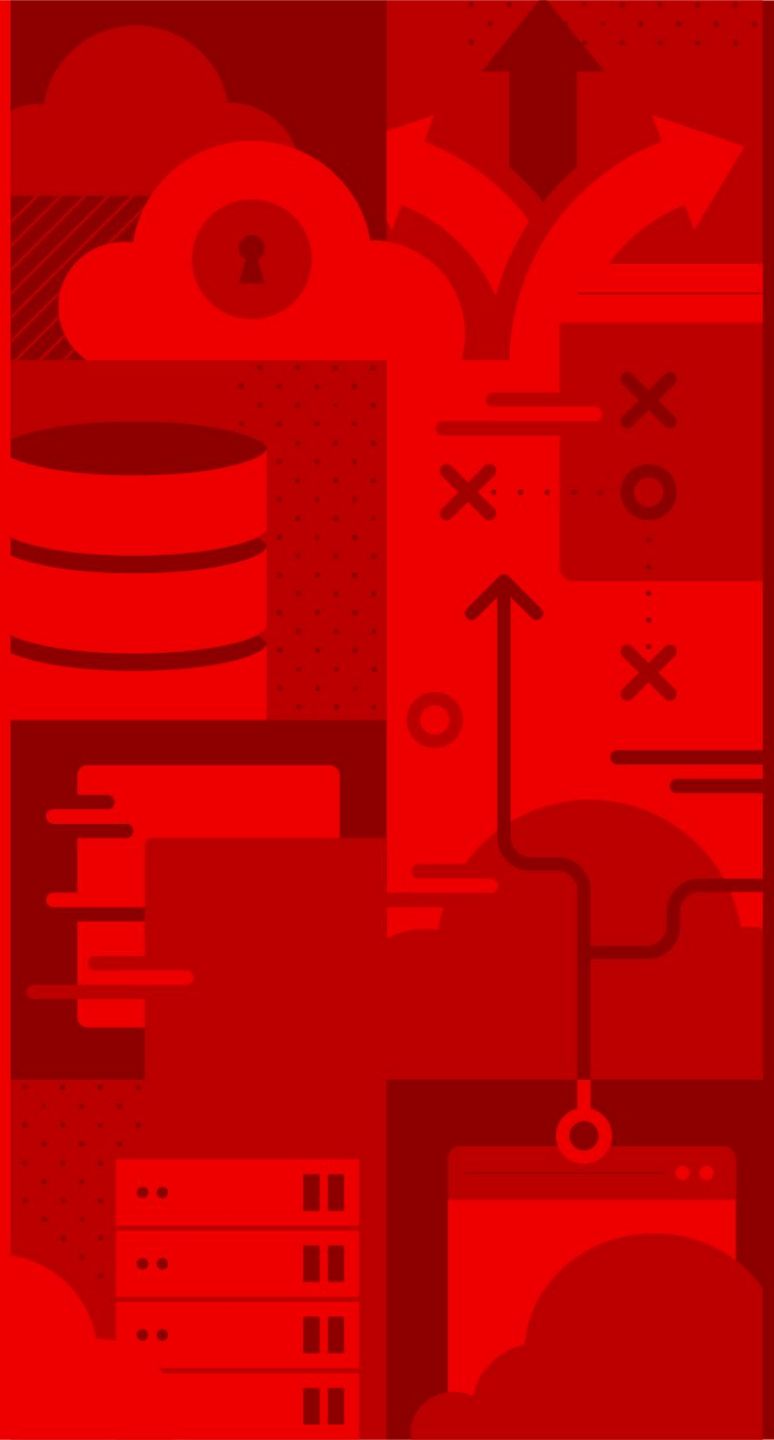
linkedin.com/company/red-hat

facebook.com/redhatinc

youtube.com/user/RedHatVideos

twitter.com/RedHat

Red Hat



Extras

Red Hat

arm64 (Ampere eMAG)