Advanced Features of Ftrace

Presenter:
Steven Rostedt
rostedt@goodmis.org
Red Hat

Ftrace Review

- Function Tracer
  - function graph tracing
  - stack tracing
  - function profiling
- Latency tracers
  - wake up latency
  - irq and preemption latency
- Trace events

Ftrace Debugfs

- Control and I/O files located in debugfs as well as the /proc system
- /proc system holds major switches
  - ftrace_enabled
    - big switch for function tracing
  - stack_trace_enabled
    - start tracing function stack size
- mount -t debugfs nodev /sys/kernel/debug
  - /sys/kernel/debug exists when debugfs is configured

trace-cmd

git://git.kernel.org/pub/scm/linux/kernel/git/rostedt/trace-cmd.git

- command line interface to ftrace debugfs

```
commands:
   record - record a trace into a trace.dat file
   start - start tracing without recording into a file
   extract - extract a trace from the kernel
   stop - stop the kernel from recording trace data
   show - show the contents of the kernel tracing buffer
   reset - disable all kernel tracing and clear the trace buffers
   report - read out the trace stored in a trace.dat file
   hist - show a historgram of the trace.dat information
   split - parse a trace.dat file into smaller file(s)
   options - list the plugin options available for trace-cmd report
   listen - listen on a network socket for trace clients
   list - list the available events, plugins or options
   restore - restore a crashed record
   snapshot - take snapshot of running trace
   stack - output, enable or disable kernel stack tracing
   check-events - parse trace event formats
```

- man trace-cmd

Debugfs

- mount -t debugfs nodev /sys/kernel/debug
- trace-cmd will automatically mount this directory for you when it needs it

The Tracing Directory

```
# ls /sys/kernel/debug/tracing
available_events       max_graph_depth     stack_trace
available_filter_functions                 options
stack_trace_filter     available_tracers   per_cpu
trace                  buffer_size_kb      printk_formats
trace_clock            buffer_total_size_kb
README                 trace_marker        current_tracer
saved_cmdlines         trace_options
dyn_ftrace_total_info  set_event           trace_pipe
enabled_functions      set_ftrace_filter   trace_stat
events                 set_ftrace_notrace  tracing_cpumask
free_buffer            set_ftrace_pid
tracing_max_latency    function_profile_enabled
set_graph_function     tracing_on          instances
set_graph_notrace      tracing_thresh      kprobe_events
snapshot               uprobe_events       kprobe_profile
stack_max_size         uprobe_profile
```

Simple Function Tracing

```
# cd /sys/kernel/debug/tracing
# echo function > current_tracer
# cat trace
# tracer: function
#
# entries-in-buffer/entries-written: 205022/119956607   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
          <idle>-0     [002] dN.1  1781.978299: rcu_eqs_exit <-rcu_idle_exit
          <idle>-0     [002] dN.1  1781.978300: rcu_eqs_exit_common <-rcu_eqs_exit
          <idle>-0     [002] .N.1  1781.978301: arch_cpu_idle_exit <-cpu_startup_entry
          <idle>-0     [002] .N.1  1781.978301: tick_nohz_idle_exit <-cpu_startup_entry
          <idle>-0     [002] dN.1  1781.978301: ktime_get <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978302: update_ts_time_stats <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978302: nr_iowait_cpu <-update_ts_time_stats
          <idle>-0     [002] dN.1  1781.978303: tick_do_update_jiffies64 <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978303: update_cpu_load_nohz <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978303: calc_load_exit_idle <-tick_nohz_idle_exit
```

Simple Function Tracing

```
# cd ~
# trace-cmd start -p function
# trace-cmd show
# tracer: function
#
# entries-in-buffer/entries-written: 205022/119956607   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
          <idle>-0     [002] dN.1  1781.978299: rcu_eqs_exit <-rcu_idle_exit
          <idle>-0     [002] dN.1  1781.978300: rcu_eqs_exit_common <-rcu_eqs_exit
          <idle>-0     [002] .N.1  1781.978301: arch_cpu_idle_exit <-cpu_startup_entry
          <idle>-0     [002] .N.1  1781.978301: tick_nohz_idle_exit <-cpu_startup_entry
          <idle>-0     [002] dN.1  1781.978301: ktime_get <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978302: update_ts_time_stats <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978302: nr_iowait_cpu <-update_ts_time_stats
          <idle>-0     [002] dN.1  1781.978303: tick_do_update_jiffies64 <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978303: update_cpu_load_nohz <-tick_nohz_idle_exit
          <idle>-0     [002] dN.1  1781.978303: calc_load_exit_idle <-tick_nohz_idle_exit
```

Simple Function Tracing

```
# cat trace_pipe
CPU:0 [LOST 191982610 EVENTS]
          <idle>-0     [000] d.h1  1942.474532: wake_up_process <-hrtimer_wakeup
          <idle>-0     [000] d.h1  1942.474533: try_to_wake_up <-wake_up_process
          <idle>-0     [000] d.h1  1942.474533: _raw_spin_lock_irqsave <-try_to_wake_up
          <idle>-0     [000] d.h1  1942.474533: preempt_count_add <-_raw_spin_lock_irqsave
          <idle>-0     [000] d.h2  1942.474534: task_waking_fair <-try_to_wake_up
          <idle>-0     [000] d.h2  1942.474534: select_task_rq_fair <-try_to_wake_up
          <idle>-0     [000] d.h2  1942.474535: __rcu_read_lock <-select_task_rq_fair
          <idle>-0     [000] d.h2  1942.474535: idle_cpu <-select_task_rq_fair
          <idle>-0     [000] d.h2  1942.474536: __rcu_read_unlock <-select_task_rq_fair
          <idle>-0     [000] d.h2  1942.474536: _raw_spin_lock <-try_to_wake_up
          <idle>-0     [000] d.h2  1942.474537: preempt_count_add <-_raw_spin_lock
          <idle>-0     [000] d.h3  1942.474537: ttwu_do_activate.constprop.82 <-try_to_wake_up
          <idle>-0     [000] d.h3  1942.474537: activate_task <-ttwu_do_activate.constprop.82
          <idle>-0     [000] d.h3  1942.474538: enqueue_task <-activate_task
          <idle>-0     [000] d.h3  1942.474538: update_rq_clock <-enqueue_task
          <idle>-0     [000] d.h3  1942.474539: enqueue_task_fair <-enqueue_task
          <idle>-0     [000] d.h3  1942.474539: enqueue_entity <-enqueue_task_fair
          <idle>-0     [000] d.h3  1942.474539: update_curr <-enqueue_entity
          <idle>-0     [000] d.h3  1942.474540: __compute_runnable_contrib <-enqueue_entity
```

Simple Function Tracing

```
# trace-cmd show -p
CPU:0 [LOST 191982610 EVENTS]
          <idle>-0     [000] d.h1  1942.474532: wake_up_process <-hrtimer_wakeup
          <idle>-0     [000] d.h1  1942.474533: try_to_wake_up <-wake_up_process
          <idle>-0     [000] d.h1  1942.474533: _raw_spin_lock_irqsave <-try_to_wake_up
          <idle>-0     [000] d.h1  1942.474533: preempt_count_add <-_raw_spin_lock_irqsave
          <idle>-0     [000] d.h2  1942.474534: task_waking_fair <-try_to_wake_up
          <idle>-0     [000] d.h2  1942.474534: select_task_rq_fair <-try_to_wake_up
          <idle>-0     [000] d.h2  1942.474535: __rcu_read_lock <-select_task_rq_fair
          <idle>-0     [000] d.h2  1942.474535: idle_cpu <-select_task_rq_fair
          <idle>-0     [000] d.h2  1942.474536: __rcu_read_unlock <-select_task_rq_fair
          <idle>-0     [000] d.h2  1942.474536: _raw_spin_lock <-try_to_wake_up
          <idle>-0     [000] d.h2  1942.474537: preempt_count_add <-_raw_spin_lock
          <idle>-0     [000] d.h3  1942.474537: ttwu_do_activate.constprop.82 <-try_to_wake_up
          <idle>-0     [000] d.h3  1942.474537: activate_task <-ttwu_do_activate.constprop.82
          <idle>-0     [000] d.h3  1942.474538: enqueue_task <-activate_task
          <idle>-0     [000] d.h3  1942.474538: update_rq_clock <-enqueue_task
          <idle>-0     [000] d.h3  1942.474539: enqueue_task_fair <-enqueue_task
          <idle>-0     [000] d.h3  1942.474539: enqueue_entity <-enqueue_task_fair
          <idle>-0     [000] d.h3  1942.474539: update_curr <-enqueue_entity
          <idle>-0     [000] d.h3  1942.474540: __compute_runnable_contrib <-enqueue_entity
```

Stopping the Trace

```
# echo nop > current_tracer
# cat trace
# tracer: nop
#
# entries-in-buffer/entries-written: 0/0   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
```

Stopping the Trace

```
# trace-cmd start -p nop
# trace-cmd show
# tracer: nop
#
# entries-in-buffer/entries-written: 0/0   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
```

Stopping the Trace

```
# echo function > current_tracer
# echo 0 > tracing_on
# cat trace
# tracer: function
#
# entries-in-buffer/entries-written: 205023/7067728   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
 gnome-terminal--6467  [003] d..1  2350.726994: _raw_spin_unlock_irqrestore <-eventfd_poll
 gnome-terminal--6467  [003] ...1  2350.726994: preempt_count_sub <-_raw_spin_unlock_irqrestore
 gnome-terminal--6467  [003] ....  2350.726994: fput <-do_sys_poll
 gnome-terminal--6467  [003] ....  2350.726995: __fdget <-do_sys_poll
 gnome-terminal--6467  [003] ....  2350.726995: __fget_light <-__fdget
 gnome-terminal--6467  [003] ....  2350.726995: __fget <-__fget_light
 gnome-terminal--6467  [003] ....  2350.726996: __rcu_read_lock <-__fget
 gnome-terminal--6467  [003] ....  2350.726996: __rcu_read_unlock <-__fget
 gnome-terminal--6467  [003] ....  2350.726996: sock_poll <-do_sys_poll
 gnome-terminal--6467  [003] ....  2350.726997: unix_poll <-sock_poll
 gnome-terminal--6467  [003] ....  2350.726997: __pollwait <-unix_poll
 gnome-terminal--6467  [003] ....  2350.726997: add_wait_queue <-__pollwait
 gnome-terminal--6467  [003] ....  2350.726998: _raw_spin_lock_irqsave <-add_wait_queue
```

Stopping the Trace

```
# trace-cmd start -p function
# trace-cmd stop
# trace-cmd show
# tracer: function
#
# entries-in-buffer/entries-written: 205023/7067728   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
 gnome-terminal--6467  [003] d..1  2350.726994: _raw_spin_unlock_irqrestore <-eventfd_poll
 gnome-terminal--6467  [003] ...1  2350.726994: preempt_count_sub <-_raw_spin_unlock_irqrestore
 gnome-terminal--6467  [003] ....  2350.726994: fput <-do_sys_poll
 gnome-terminal--6467  [003] ....  2350.726995: __fdget <-do_sys_poll
 gnome-terminal--6467  [003] ....  2350.726995: __fget_light <-__fdget
 gnome-terminal--6467  [003] ....  2350.726995: __fget <-__fget_light
 gnome-terminal--6467  [003] ....  2350.726996: __rcu_read_lock <-__fget
 gnome-terminal--6467  [003] ....  2350.726996: __rcu_read_unlock <-__fget
 gnome-terminal--6467  [003] ....  2350.726996: sock_poll <-do_sys_poll
 gnome-terminal--6467  [003] ....  2350.726997: unix_poll <-sock_poll
 gnome-terminal--6467  [003] ....  2350.726997: __pollwait <-unix_poll
 gnome-terminal--6467  [003] ....  2350.726997: add_wait_queue <-__pollwait
 gnome-terminal--6467  [003] ....  2350.726998: _raw_spin_lock_irqsave <-add_wait_queue
```

Clearing the Trace

```
# echo > trace
# cat trace
# tracer: function
#
# entries-in-buffer/entries-written: 0/0   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
```

Function Graph Tracer

```
# echo function_graph > current_tracer
# cat trace
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 2)   7.879 us    |  } /* context_tracking_user_exit */
 2)               |  __do_page_fault() {
 2)   0.070 us    |    down_read_trylock();
 2)   0.057 us    |    __might_sleep();
 2)   0.096 us    |    find_vma();
 2)               |    handle_mm_fault() {
 2)               |      __do_fault() {
 2)               |        filemap_fault() {
 2)               |          find_get_page() {
 2)   0.057 us    |            __rcu_read_lock();
 2)   0.061 us    |            __rcu_read_unlock();
 2)   1.241 us    |          }
 2)   0.074 us    |          __might_sleep();
 2)   2.201 us    |        }
 2)               |        _raw_spin_lock() {
 2)   0.069 us    |          preempt_count_add();
 2)   0.528 us    |        }
 2)   0.063 us    |        add_mm_counter_fast();
 2)   0.070 us    |        page_add_file_rmap();
 2)               |        _raw_spin_unlock() {
 2)   0.070 us    |          preempt_count_sub();
```

Function Graph Tracer

```
# trace-cmd start -p function_graph
# trace-cmd show
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 2)   7.879 us    |  } /* context_tracking_user_exit */
 2)               |  __do_page_fault() {
 2)   0.070 us    |    down_read_trylock();
 2)   0.057 us    |    __might_sleep();
 2)   0.096 us    |    find_vma();
 2)               |    handle_mm_fault() {
 2)               |      __do_fault() {
 2)               |        filemap_fault() {
 2)               |          find_get_page() {
 2)   0.057 us    |            __rcu_read_lock();
 2)   0.061 us    |            __rcu_read_unlock();
 2)   1.241 us    |          }
 2)   0.074 us    |          __might_sleep();
 2)   2.201 us    |        }
 2)               |        _raw_spin_lock() {
 2)   0.069 us    |          preempt_count_add();
 2)   0.528 us    |        }
 2)   0.063 us    |        add_mm_counter_fast();
 2)   0.070 us    |        page_add_file_rmap();
 2)               |        _raw_spin_unlock() {
 2)   0.070 us    |          preempt_count_sub();
```

Dynamic Function Tracing

- set_ftrace_filter
  - only trace functions listed
- set_ftrace_notrace
  - do not trace functions listed
  - overrides set_ftrace_filter
- available_filter_functions
  - list of functions that can be added to the above two files
- set_graph_function
  - Trace what a function does

Dynamic Function Tracing

```
# cat available_filter_functions
run_init_process
try_to_run_init_process
do_one_initcall
match_dev_by_uuid
rootfs_mount
name_to_dev_t
name_to_dev_t
calibrate_delay
start_thread_common.constprop.7
set_personality_ia32
__show_regs
release_thread
start_thread
start_thread_ia32
set_personality_64bit
get_wchan
do_arch_prctl
copy_thread
sys_arch_prctl
KSTK_ESP
restore_sigcontext
setup_sigcontext
do_signal
do_notify_resume
signal_fault
sys_rt_sigreturn
math_state_restore
do_divide_error
do_overflow
do_bounds
[...]
```

Dynamic Function Tracing

```
# trace-cmd list -f
run_init_process
try_to_run_init_process
do_one_initcall
match_dev_by_uuid
rootfs_mount
name_to_dev_t
name_to_dev_t
calibrate_delay
start_thread_common.constprop.7
set_personality_ia32
__show_regs
release_thread
start_thread
start_thread_ia32
set_personality_64bit
get_wchan
do_arch_prctl
copy_thread
sys_arch_prctl
KSTK_ESP
restore_sigcontext
setup_sigcontext
do_signal
do_notify_resume
signal_fault
sys_rt_sigreturn
math_state_restore
do_divide_error
do_overflow
do_bounds
[...]
```

Dynamic Function Tracing

```
# trace-cmd list -f ‘^hrtimer’
hrtimer_init_sleeper
hrtimer_wakeup
hrtimer_forward
hrtimer_get_res
hrtimer_force_reprogram
hrtimer_reprogram.isra.25
hrtimer_rt_defer.part.26
hrtimer_get_remaining
hrtimer_init
hrtimer_try_to_cancel
hrtimers_resume
hrtimer_wait_for_timer
hrtimer_cancel
hrtimer_start
hrtimer_start_range_ns
hrtimer_get_next_event
hrtimer_interrupt
hrtimer_cpu_notify
hrtimer_peek_ahead_timers
hrtimer_run_queues
hrtimer_nanosleep
hrtimer_nanosleep_restart
```

Dynamic Function Tracing

```
# echo '*sched*' > set_ftrace_filter
# echo function > current_tracer
# cat trace
# tracer: function
#
# entries-in-buffer/entries-written: 193727/240417   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
          <idle>-0     [003] d.h3  6325.742705: resched_task <-check_preempt_curr
          <idle>-0     [003] dNh3  6325.742712: native_smp_send_reschedule <-enqueue_task_fair
          <idle>-0     [003] dNh3  6325.742714: resched_task <-check_preempt_curr
          <idle>-0     [003] dN.1  6325.742719: smp_reschedule_interrupt <-reschedule_interrupt
          <idle>-0     [003] dN.1  6325.742720: scheduler_ipi <-smp_reschedule_interrupt
          <idle>-0     [003] dNh1  6325.742722: sched_ttwu_pending <-scheduler_ipi
          <idle>-0     [003] .N.1  6325.742728: schedule_preempt_disabled <-cpu_startup_entry
          <idle>-0     [003] .N..  6325.742729: schedule <-schedule_preempt_disabled
          <idle>-0     [003] .N..  6325.742731: __schedule <-preempt_schedule
          <idle>-0     [003] .N.1  6325.742732: rcu_sched_qs <-rcu_note_context_switch
          <idle>-0     [003] dN.2  6325.742733: pre_schedule_idle <-__schedule
           aprsd-3467  [003] ....  6325.742746: schedule <-do_nanosleep
           aprsd-3467  [003] ....  6325.742747: __schedule <-schedule
           aprsd-3467  [003] ...1  6325.742748: rcu_sched_qs <-rcu_note_context_switch
           aprsd-3454  [003] ....  6325.742767: schedule <-do_nanosleep
           aprsd-3454  [003] ....  6325.742767: __schedule <-schedule
           aprsd-3454  [003] ...1  6325.742768: rcu_sched_qs <-rcu_note_context_switch
     rcu_preempt-9     [003] d..2  6325.742788: smp_reschedule_interrupt <-reschedule_interrupt
     rcu_preempt-9     [003] d..2  6325.742789: scheduler_ipi <-smp_reschedule_interrupt
```

Dynamic Function Tracing

```
# trace-cmd start -p function -l ‘*sched*’
# trace-cmd show
# tracer: function
#
# entries-in-buffer/entries-written: 193727/240417   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
          <idle>-0     [003] d.h3  6325.742705: resched_task <-check_preempt_curr
          <idle>-0     [003] dNh3  6325.742712: native_smp_send_reschedule <-enqueue_task_fair
          <idle>-0     [003] dNh3  6325.742714: resched_task <-check_preempt_curr
          <idle>-0     [003] dN.1  6325.742719: smp_reschedule_interrupt <-reschedule_interrupt
          <idle>-0     [003] dN.1  6325.742720: scheduler_ipi <-smp_reschedule_interrupt
          <idle>-0     [003] dNh1  6325.742722: sched_ttwu_pending <-scheduler_ipi
          <idle>-0     [003] .N.1  6325.742728: schedule_preempt_disabled <-cpu_startup_entry
          <idle>-0     [003] .N..  6325.742729: schedule <-schedule_preempt_disabled
          <idle>-0     [003] .N..  6325.742731: __schedule <-preempt_schedule
          <idle>-0     [003] .N.1  6325.742732: rcu_sched_qs <-rcu_note_context_switch
          <idle>-0     [003] dN.2  6325.742733: pre_schedule_idle <-__schedule
           aprsd-3467  [003] ....  6325.742746: schedule <-do_nanosleep
           aprsd-3467  [003] ....  6325.742747: __schedule <-schedule
           aprsd-3467  [003] ...1  6325.742748: rcu_sched_qs <-rcu_note_context_switch
           aprsd-3454  [003] ....  6325.742767: schedule <-do_nanosleep
           aprsd-3454  [003] ....  6325.742767: __schedule <-schedule
           aprsd-3454  [003] ...1  6325.742768: rcu_sched_qs <-rcu_note_context_switch
     rcu_preempt-9     [003] d..2  6325.742788: smp_reschedule_interrupt <-reschedule_interrupt
     rcu_preempt-9     [003] d..2  6325.742789: scheduler_ipi <-smp_reschedule_interrupt
```

Dynamic Function Tracing

```
# echo SyS_read > set_graph_function
# echo function_graph > current_tracer
# cat trace
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 2)               |  SyS_read() {
 2)   0.341 us    |    __fdget_pos();
 2)               |    vfs_read() {
 2)               |      rw_verify_area() {
 2)               |        security_file_permission() {
 2)   0.180 us    |          cap_file_permission();
 2)   0.175 us    |          __fsnotify_parent();
 2)   0.180 us    |          fsnotify();
 2)   3.466 us    |        }
 2)   4.509 us    |      }
 2)               |      tty_read() {
 2)   0.361 us    |        tty_paranoia_check();
 2)               |        tty_ldisc_ref_wait() {
 2)               |          ldsem_down_read() {
 2)   0.181 us    |            __might_sleep();
 2)   1.815 us    |          }
 2)   3.621 us    |        }
 2)               |        n_tty_read() {
 2)               |          _raw_spin_lock_irq() {
 2)   0.336 us    |            preempt_count_add();
 2)   2.232 us    |          }
 2)               |          _raw_spin_unlock_irq() {
 2)   0.261 us    |            preempt_count_sub();
 2)   2.047 us    |          }
 2)               |          mutex_lock_interruptible() {
 2)   0.476 us    |            __might_sleep();
 2)   2.252 us    |          }
```

Dynamic Function Tracing

```
# trace-cmd start -p function_graph -g SyS_read
# trace-cmd show
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 2)               |  SyS_read() {
 2)   0.341 us    |    __fdget_pos();
 2)               |    vfs_read() {
 2)               |      rw_verify_area() {
 2)               |        security_file_permission() {
 2)   0.180 us    |          cap_file_permission();
 2)   0.175 us    |          __fsnotify_parent();
 2)   0.180 us    |          fsnotify();
 2)   3.466 us    |        }
 2)   4.509 us    |      }
 2)               |      tty_read() {
 2)   0.361 us    |        tty_paranoia_check();
 2)               |        tty_ldisc_ref_wait() {
 2)               |          ldsem_down_read() {
 2)   0.181 us    |            __might_sleep();
 2)   1.815 us    |          }
 2)   3.621 us    |        }
 2)               |        n_tty_read() {
 2)               |          _raw_spin_lock_irq() {
 2)   0.336 us    |            preempt_count_add();
 2)   2.232 us    |          }
 2)               |          _raw_spin_unlock_irq() {
 2)   0.261 us    |            preempt_count_sub();
 2)   2.047 us    |          }
 2)               |          mutex_lock_interruptible() {
 2)   0.476 us    |            __might_sleep();
 2)   2.252 us    |          }
```

Function Triggers

- <function-name>:<trigger>:<count>
  - count is optional
    - unlimited if missing
- try_to_wake_up:traceon:5
- schedule:traceoff:5

Function Triggers

```
# echo 0 > tracing_on
# echo function > current_tracer
# echo ‘try_to_wake_up:traceon:5 schedule:traceoff:5’ > \
    set_ftrace_filter
# cat trace_pipe
          <idle>-0     [003] .N..  6808.634701: schedule <-schedule_preempt_disabled
          <idle>-0     [003] .N..  6808.634702: __schedule <-preempt_schedule
          <idle>-0     [003] .N.1  6808.634702: rcu_sched_qs <-rcu_note_context_switch
          <idle>-0     [003] dN.2  6808.634704: pre_schedule_idle <-__schedule
 panel-19-system-5144  [003] d..3  6808.634933: resched_task <-check_preempt_curr
 panel-19-system-5144  [003] ....  6808.634946: _cond_resched <-task_work_run
 panel-19-system-5144  [003] ....  6808.634961: _cond_resched <-unmap_single_vma
 panel-19-system-5144  [003] ....  6808.635061: _cond_resched <-task_work_run
 panel-19-system-5144  [003] ....  6808.635128: _cond_resched <-task_work_run
 panel-19-system-5144  [003] ....  6808.635135: _cond_resched <-unmap_single_vma
 panel-19-system-5144  [003] d..3  6808.636140: resched_task <-check_preempt_curr
 panel-19-system-5144  [003] ....  6808.636203: poll_schedule_timeout <-do_sys_poll
 panel-19-system-5144  [003] ....  6808.636204: schedule_hrtimeout_range
<-poll_schedule_timeout
 panel-19-system-5144  [003] ....  6808.636204: schedule_hrtimeout_range_clock
<-schedule_hrtimeout_range
 panel-19-system-5144  [003] ....  6808.636205: schedule <-schedule_hrtimeout_range_clock
 panel-19-system-5144  [003] ....  6808.636205: __schedule <-schedule
 panel-19-system-5144  [003] ...1  6808.636205: rcu_sched_qs <-rcu_note_context_switch
```

Function Triggers

```
# trace-cmd start -p function \
    -l ‘try_to_wake_up:traceon:5 schedule:traceoff:5’
# trace-cmd show
          <idle>-0     [003] .N..  6808.634701: schedule <-schedule_preempt_disabled
          <idle>-0     [003] .N..  6808.634702: __schedule <-preempt_schedule
          <idle>-0     [003] .N.1  6808.634702: rcu_sched_qs <-rcu_note_context_switch
          <idle>-0     [003] dN.2  6808.634704: pre_schedule_idle <-__schedule
 panel-19-system-5144  [003] d..3  6808.634933: resched_task <-check_preempt_curr
 panel-19-system-5144  [003] ....  6808.634946: _cond_resched <-task_work_run
 panel-19-system-5144  [003] ....  6808.634961: _cond_resched <-unmap_single_vma
 panel-19-system-5144  [003] ....  6808.635061: _cond_resched <-task_work_run
 panel-19-system-5144  [003] ....  6808.635128: _cond_resched <-task_work_run
 panel-19-system-5144  [003] ....  6808.635135: _cond_resched <-unmap_single_vma
 panel-19-system-5144  [003] d..3  6808.636140: resched_task <-check_preempt_curr
 panel-19-system-5144  [003] ....  6808.636203: poll_schedule_timeout <-do_sys_poll
 panel-19-system-5144  [003] ....  6808.636204: schedule_hrtimeout_range
<-poll_schedule_timeout
 panel-19-system-5144  [003] ....  6808.636204: schedule_hrtimeout_range_clock
<-schedule_hrtimeout_range
 panel-19-system-5144  [003] ....  6808.636205: schedule <-schedule_hrtimeout_range_clock
 panel-19-system-5144  [003] ....  6808.636205: __schedule <-schedule
 panel-19-system-5144  [003] ...1  6808.636205: rcu_sched_qs <-rcu_note_context_switch
```

Function Triggers

```
# cat set_ftrace_filter
#### all functions enabled ####
schedule:traceoff:count=0
try_to_wake_up:traceon:count=0

# echo ‘!schedule:traceoff:count=0’ > set_ftrace_filter
# cat set_ftrace_filter
#### all functions enabled ####
try_to_wake_up:traceon:count=0

# echo ‘!try_to_wake_up:traceon:count=0’ > set_ftrace_filter
# cat set_ftrace_filter
#### all functions enabled ####
```

Function Triggers

- trace-cmd show options coming in trace-cmd version 2.4

```
# trace-cmd show --ftrace_filter
#### all functions enabled ####
schedule:traceoff:count=0
try_to_wake_up:traceon:count=0

# trace-cmd start -p nop -l ‘!schedule:traceoff:count=0’
# trace-cmd show --ftrace_filter
#### all functions enabled ####
try_to_wake_up:traceon:count=0

# trace-cmd start -p nop -l ‘!try_to_wake_up:traceon:count=0’
# trace-cmd show --ftrace_filter
#### all functions enabled ####
```

Function Triggers

```
# echo ‘schedule:traceon’ > set_ftrace_filter
# cat set_ftrace_filter
#### all functions enabled ####
schedule:traceoff:unlimited

# echo ‘!schedule:traceon:unlimited’ > set_ftrace_filter
# cat set_ftrace_filter
#### all functions enabled ####
schedule:traceoff:unlimited
```

- ??
  - Why did that not work?

Function Triggers

```
# echo ‘schedule:traceon’ > set_ftrace_filter
# cat set_ftrace_filter
#### all functions enabled ####
schedule:traceoff:unlimited

# echo ‘!schedule:traceon’ > set_ftrace_filter
# cat set_ftrace_filter
#### all functions enabled ####
```

- Don’t add ‘:unlimited’
  - I plan on fixing this in the near future

Function Triggers

- traceon is usually not helpful
- traceoff, on the other hand, is
    - Set to a function in a error path
    - Will stop tracing when the error is hit

Function Triggers

```
# echo schedule:stacktrace > set_ftrace_filter
# cat trace
# tracer: nop
#
# entries-in-buffer/entries-written: 67843/200785   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
          <idle>-0     [003] .N.2  8202.221929: <stack trace>
 => cpu_startup_entry
 => start_secondary
           aprsd-3454  [003] ...2  8202.221954: <stack trace>
 => hrtimer_nanosleep
 => SyS_nanosleep
 => tracesys
          <idle>-0     [003] .N.2  8202.223021: <stack trace>
 => cpu_startup_entry
 => start_secondary
           aprsd-3454  [003] ...2  8202.223046: <stack trace>
 => hrtimer_nanosleep
 => SyS_nanosleep
 => tracesys
          <idle>-0     [003] .N.2  8202.223736: <stack trace>
 => cpu_startup_entry
 => start_secondary
          chrome-5907  [003] ...2  8202.223840: <stack trace>
 => futex_wait
 => do_futex
 => SyS_futex
 => tracesys
```

Function Triggers

```
# trace-cmd -p nop -l schedule:stacktrace
# cat trace
# tracer: nop
#
# entries-in-buffer/entries-written: 67843/200785   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
          <idle>-0     [003] .N.2  8202.221929: <stack trace>
 => cpu_startup_entry
 => start_secondary
           aprsd-3454  [003] ...2  8202.221954: <stack trace>
 => hrtimer_nanosleep
 => SyS_nanosleep
 => tracesys
          <idle>-0     [003] .N.2  8202.223021: <stack trace>
 => cpu_startup_entry
 => start_secondary
           aprsd-3454  [003] ...2  8202.223046: <stack trace>
 => hrtimer_nanosleep
 => SyS_nanosleep
 => tracesys
          <idle>-0     [003] .N.2  8202.223736: <stack trace>
 => cpu_startup_entry
 => start_secondary
          chrome-5907  [003] ...2  8202.223840: <stack trace>
 => futex_wait
 => do_futex
 => SyS_futex
 => tracesys
```

Other Function Triggers

- dump
  - triggers ftrace_dump_on_oops
  - dumps entire trace buffer to console
- cpudump
  - like dump but only dumps the current CPU buffer to console
- enable_event / disable_event
  - Will describe with event_triggers

Events

```
# ls /sys/kernel/debug/tracing/events
block               i915         nmi           sock
cfg80211            iommu        oom           spi
compaction          irq          pagemap       sunrpc
context_tracking    irq_vectors  power         swiotlb
drm                 jbd          printk        syscalls
enable              jbd2         random        task
exceptions          kmem         ras           timer
ext3                kvm          raw_syscalls   udp
ext4                kvmmmu       rcu           vmscan
filemap             mac80211     regmap        vsyscall
ftrace              mac80211_msg regulator     workqueue
gpio                mce          rpm           writeback
hda                 migrate      sched         xhci-hcd
hda_intel           module       scsi
header_event        napi         signal
header_page         net          skb
```

Events

```
# ls /sys/kernel/debug/tracing/events/sched
enable              sched_process_exit  sched_stat_sleep
filter              sched_process_fork  sched_stat_wait
sched_kthread_stop  sched_process_free  sched_stick_numa
sched_kthread_stop_ret sched_process_hang  sched_swap_numa
sched_migrate_task  sched_process_wait  sched_switch
sched_move_numa     sched_stat_blocked  sched_wait_task
sched_pi_setprio    sched_stat_iowait   sched_wakeup
sched_process_exec  sched_stat_runtime  sched_wakeup_new
```

Events

```
# trace-cmd list -e
hda_intel:azx_pcm_trigger
hda_intel:azx_get_position
hda:hda_send_cmd
hda:hda_get_response
hda:hda_bus_reset
hda:hda_power_down
hda:hda_power_up
hda:hda_power_count
hda:hda_unsol_event
i915:i915_gem_object_create
i915:i915_vma_bind
i915:i915_vma_unbind
i915:i915_gem_object_change_domain
i915:i915_gem_object_pwrite
i915:i915_gem_object_pread
i915:i915_gem_object_fault
i915:i915_gem_object_clflush
i915:i915_gem_object_destroy
i915:i915_gem_evict
i915:i915_gem_evict_everything
```

Events

- -e search, coming in trace-cmd 2.4

```
# trace-cmd list -e sched:
sched:sched_swap_numa
sched:sched_stick_numa
sched:sched_move_numa
sched:sched_pi_setprio
sched:sched_stat_runtime
sched:sched_stat_blocked
sched:sched_stat_iowait
sched:sched_stat_sleep
sched:sched_stat_wait
sched:sched_process_exec
sched:sched_process_fork
sched:sched_process_wait
sched:sched_wait_task
sched:sched_process_exit
sched:sched_process_free
sched:sched_migrate_task
sched:sched_switch
sched:sched_wakeup_new
sched:sched_wakeup
```

Events

```
# ls /debug/tracing/events/sched/sched_switch

enable     filter     format     id     trigger
```

Dynamic Function Tracing and Events

```
# echo ‘do_IRQ’ > set_ftrace_filter
# echo 1 > events/irq/irq_handler_entry/enable
# echo function_graph > current_tracer
# cat trace
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=12 name=i8042 */
 0) + 13.186 us   |  }
 0)   <========== |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=12 name=i8042 */
 0) + 10.287 us   |  }
 0)   <========== |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=12 name=i8042 */
 0) + 10.252 us   |  }
 0)   <========== |
```

Dynamic Function Tracing and Events

```
# trace-cmd start -p function_graph -l ‘do_IRQ’ \
   -e irq_handler_entry
# trace-cmd show
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=12 name=i8042 */
 0) + 13.186 us   |  }
 0)   <========== |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=12 name=i8042 */
 0) + 10.287 us   |  }
 0)   <========== |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=12 name=i8042 */
 0) + 10.252 us   |  }
 0)   <========== |
```

Event Triggers

- Similar to function triggers
  - traceon
  - traceoff
  - stacktrace
  - enable event*
  - disable event*
- More features
  - snapshot
  - conditionals

Event Triggers

```
# echo stacktrace > events/sched/sched_switch/trigger
# cat events/sched/sched_switch/trigger
stacktrace:unlimited


# cat trace
# tracer: nop
#
# entries-in-buffer/entries-written: 66382/179849   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
           aprsd-3467  [003] d..3 93035.966745: <stack trace>
 => __schedule
 => schedule
 => do_nanosleep
 => hrtimer_nanosleep
 => SyS_nanosleep
 => tracesys
          <idle>-0     [003] d..3 93035.967265: <stack trace>
 => __schedule
 => schedule
 => schedule_preempt_disabled
 => cpu_startup_entry
 => start_secondary
 Gamepad polling-5797  [003] d..3 93035.967358: <stack trace>
 => __schedule
 => schedule
 => schedule_hrtimeout_range_clock
 => schedule_hrtimeout_range
```

Event Triggers

- Coming in trace-cmd 2.4

```
# trace-cmd start -v -e sched_switch -R stacktrace
# trace-cmd list -e sched_switch -R
sched:sched_switch
stacktrace:unlimited

# trace-cmd show
# tracer: nop
#
# entries-in-buffer/entries-written: 66382/179849   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
           aprsd-3467  [003] d..3 93035.966745: <stack trace>
 => __schedule
 => schedule
 => do_nanosleep
 => hrtimer_nanosleep
 => SyS_nanosleep
 => tracesys
          <idle>-0     [003] d..3 93035.967265: <stack trace>
 => __schedule
 => schedule
 => schedule_preempt_disabled
```

Event Triggers

- -v is similar to grep -v
    - grep -e match_me -v -e ignore_me
- trace-cmd start -e trace_me -v -e ignore_me
- Useful for ignoring events within a system
- Now useful for enabling a trigger without enabling the event

```
# trace-cmd start -e sched_switch
# trace-cmd start -e sched -v -e sched_switch
# trace-cmd start -v -e sched_switch -R stacktrace
```

Disabling Event Triggers

```
# echo ‘!stacktrace’ > events/sched/sched_switch/trigger
# cat events/sched/sched_switch/trigger
# Available triggers:
# traceon traceoff snapshot stacktrace enable_event disable_event
```

Disabling Event Triggers

```
# trace-cmd start -v -e sched_switch -R ‘!stacktrace’
# trace-cmd list -e sched_switch -R
sched:sched_switch
# Available triggers:
# traceon traceoff snapshot stacktrace enable_event disable_event
```

Event Triggers

```
# cat events/enable
X

# cat events/sched/enable
X

# cat events/sched/sched_wakeup/enable
0*

# cat set_event
sched:sched_wakeup
```

- In future, “set_event” may have “*” by events with triggers.

Event Triggers

- Not very useful if you blindly enable tracing
- Need a way to conditionally enable it
- Need a way to conditionally disable it

Event Format

```
# cat /debug/tracing/events/sched/sched_switch/format
name: sched_switch
ID: 276
format:
    field:unsigned short common_type; offset:0;    size:2;  signed:0;
    field:unsigned char common_flags; offset:2;    size:1;  signed:0;
    field:unsigned char common_preempt_count;  offset:3;    size:1;  signed:0;
    field:int common_pid;         offset:4;    size:4;  signed:1;

    field:char prev_comm[16];     offset:8;    size:16; signed:1;
    field:pid_t prev_pid;         offset:24;   size:4;  signed:1;
    field:int prev_prio;          offset:28;   size:4;  signed:1;
    field:long prev_state;        offset:32;   size:8;  signed:1;
    field:char next_comm[16];     offset:40;   size:16; signed:1;
    field:pid_t next_pid;         offset:56;   size:4;  signed:1;
    field:int next_prio;          offset:60;   size:4;  signed:1;

print fmt: "prev_comm=%s prev_pid=%d prev_prio=%d prev_state=%s%s ==>
next_comm=%s next_pid=%d next_prio=%d", REC->prev_comm, REC->prev_pid,
REC->prev_prio, REC->prev_state & (1024-1) ? __print_flags(REC->prev_state &
(1024-1), "|", { 1, "S"} , { 2, "D" }, { 4, "T" }, { 8, "t" }, { 16, "Z" }, {
32, "X" }, { 64, "x" }, { 128, "K" }, { 256, "W" }, { 512, "P" }) : "R",
REC->prev_state & 1024 ? "+" : "", REC->next_comm, REC->next_pid,
REC->next_prio
```

Event Format

```
# trace-cmd list -e sched_switch -F
system: sched
name: sched_switch
ID: 275
format:
    field:unsigned short common_type; offset:0;    size:2;  signed:0;
    field:unsigned char common_flags; offset:2;    size:1;  signed:0;
    field:unsigned char common_preempt_count;  offset:3;    size:1;  signed:0;
    field:int common_pid; offset:4;    size:4;  signed:1;
    field:unsigned short common_migrate_disable;   offset:8;    size:2;  signed:0;
    field:unsigned short common_padding;   offset:10;   size:2;  signed:0;

    field:char prev_comm[16]; offset:16;   size:16; signed:1;
    field:pid_t prev_pid; offset:32;   size:4;  signed:1;
    field:int prev_prio;  offset:36;   size:4;  signed:1;
    field:long prev_state;    offset:40;   size:8;  signed:1;
    field:char next_comm[16]; offset:48;   size:16; signed:1;
    field:pid_t next_pid; offset:64;   size:4;  signed:1;
    field:int next_prio;  offset:68;   size:4;  signed:1;
```

Event Triggers

```
# echo “traceon if pid==$$” > events/sched/sched_wakeup/trigger
# cat events/sched/sched_wakeup/trigger
traceon:unlimited if pid==7623

# echo “traceoff if next_pid==$$” > events/sched/sched_switch/trigger
# cat events/sched/sched_switch/trigger
traceoff:unlimited if pid==7623

# echo ‘!traceon’ > events/sched/sched_wakeup/trigger
# cat events/sched/sched_wakeup/trigger
# Available triggers:
# traceon traceoff snapshot stacktrace enable_event disable_event

# echo ‘!traceoff’ > events/sched/sched_switch/trigger
# cat events/sched/sched_switch/trigger
# Available triggers:
# traceon traceoff snapshot stacktrace enable_event disable_event
```

Event Triggers

```
# trace-cmd start -v -e sched_wakeup -R “traceon if pid==$$’
# trace-cmd list -e ‘sched_wakeup$’ -R
sched:sched_wakeup
traceon:unlimited if pid==7623

# trace-cmd start -v -e sched_switch -R “traceoff if next_pid==$$”
# trace-cmd list -e sched_switch -R
sched:sched_switch
traceoff:unlimited if pid==7623

# trace-cmd start -v -e sched_wakeup -R ‘!traceon’
# trace-cmd list -e ‘sched_wakeup$’ -R
sched:sched_wakeup
# Available triggers:
# traceon traceoff snapshot stacktrace enable_event disable_event

# trace-cmd start -v -e sched_switch -R ‘!traceoff’
# trace-cmd list -e sched_switch -R
sched:sched_switch
# Available triggers:
# traceon traceoff snapshot stacktrace enable_event disable_event
```

Event Triggers

```
# cat trace
# tracer: nop
#
# entries-in-buffer/entries-written: 3519/3519   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
            bash-7623  [002] d..4  2878.448311: sched_wakeup: comm=rcuop/2 pid=12 prio=120
success=1 target_cpu=001
          <idle>-0     [001] d..3  2878.448355: sched_switch: prev_comm=swapper/1 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=rcuop/2 next_pid=12 next_prio=120
         rcuop/2-12    [001] d..3  2878.448371: sched_switch: prev_comm=rcuop/2 prev_pid=12
prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
            bash-7623  [002] d.h3  2878.448824: sched_wakeup: comm=aprsd pid=3543 prio=120
success=1 target_cpu=001
            bash-7623  [002] d..4  2878.448849: sched_wakeup: comm=kworker/2:0 pid=8390 prio=120
success=1 target_cpu=002
          <idle>-0     [001] d..3  2878.448877: sched_switch: prev_comm=swapper/1 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3543 next_prio=120
            bash-7623  [002] d..3  2878.448888: sched_switch: prev_comm=bash prev_pid=7623
prev_prio=120 prev_state=S ==> next_comm=kworker/2:0 next_pid=8390 next_prio=120
     kworker/2:0-8390  [002] d..4  2878.448904: sched_wakeup: comm=gnome-terminal- pid=5415
prio=120 success=1 target_cpu=002
           aprsd-3543  [001] d..3  2878.448914: sched_switch: prev_comm=aprsd prev_pid=3543
prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
     kworker/2:0-8390  [002] d..3  2878.448916: sched_switch: prev_comm=kworker/2:0 prev_pid=8390
prev_prio=120 prev_state=S ==> next_comm=gnome-terminal- next_pid=5415 next_prio=120
          <idle>-0     [001] dNh4  2878.448928: sched_wakeup: comm=aprsd pid=3556 prio=110
success=1 target_cpu=001
          <idle>-0     [001] d..3  2878.448935: sched_switch: prev_comm=swapper/1 prev_pid=0
```

Event Conditions

- <trigger> “if” <Condition>
- Condition := <cond> | <cond> <bop> <Condition> | “(“ <Condition> “)”
- bop := “&&” | “||”
- cond: = <field> <op> <value>
- field := any field in event format
- op := “==” | “!=” | “<” | “>” | “<=” | “>=” | “&” | “~”

Event Conditions

- Number comparisons
  - “==”, “!=”, “>”, “<”, “>=”, “<=”, “&”
- String comparisons
  - “==”, “!=”, “~”
  - “~” same glob that set_ftrace_filter uses

Enable/Disable Events

- Triggers to enable and disable other events
- Can enable the same event
  - Remember, triggers don’t necessarily trace the event where the trigger lies
- Allow tracing something and then enable/disable something on a condition of an event
  - as suppose to using traceon or traceoff

Enable/Disable Events

- enable_event:<system>:<event>
- disable_event:<system>:<event>

```
# echo do_IRQ:sched:sched_switch > set_ftrace_filter

# echo enable_event:net:net_dev_xmit if irq==51 >
      events/irq/irq_handler_entry/trigger

# cat events/net/net_dev_xmit/enable
1*
```

Enable/Disable Events

- enable_event:<system>:<event>
- disable_event:<system>:<event>

```
# trace-cmd start -p function -l “do_IRQ:sched:sched_switch”

# trace-cmd start -v -e irq_handler_entry \
     -R “enable_event:net:net_dev_xmit if irq==51”

# cat events/net/net_dev_xmit/enable
1*
```

Snapshots

- Uses the latency tracer technology
- Takes a “snapshot” of the current data in the ring buffer
- Snapshot buffer doesn’t get updated, except for performing the snapshot

Snapshot

```
# cat snapshot
# tracer: nop
#
#
# * Snapshot is freed *
#
# Snapshot commands:
# echo 0 > snapshot : Clears and frees snapshot buffer
# echo 1 > snapshot : Allocates snapshot buffer, if not already allocated.
#                      Takes a snapshot of the main buffer.
# echo 2 > snapshot : Clears snapshot buffer (but does not allocate or free)
#                      (Doesn't have to be '2' works with any number that
#                       is not a '0' or '1')
```

Snapshot

```
# trace-cmd show -s
# tracer: nop
#
#
# * Snapshot is freed *
#
# Snapshot commands:
# echo 0 > snapshot : Clears and frees snapshot buffer
# echo 1 > snapshot : Allocates snapshot buffer, if not already allocated.
#                      Takes a snapshot of the main buffer.
# echo 2 > snapshot : Clears snapshot buffer (but does not allocate or free)
#                      (Doesn't have to be '2' works with any number that
#                       is not a '0' or '1')
```

Snapshot

```
# trace-cmd snapshot
# tracer: nop
#
#
# * Snapshot is freed *
#
# Snapshot commands:
# echo 0 > snapshot : Clears and frees snapshot buffer
# echo 1 > snapshot : Allocates snapshot buffer, if not already allocated.
#                       Takes a snapshot of the main buffer.
# echo 2 > snapshot : Clears snapshot buffer (but does not allocate or free)
#                       (Doesn't have to be '2' works with any number that
#                        is not a '0' or '1')
```

Snapshot

```
# echo 1 > snapshot
# cat snapshot
# tracer: nop
#
# entries-in-buffer/entries-written: 1747/1747   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
            bash-7623  [003] d..4  3563.846447: sched_wakeup: comm=kworker/3:0 pid=8766 
prio=120 success=1 target_cpu=003
            bash-7623  [003] d..3  3563.846471: sched_switch: prev_comm=bash prev_pid=7623 
prev_prio=120 prev_state=S ==> next_comm=kworker/3:0 next_pid=8766 next_prio=120
     kworker/3:0-8766  [003] d..4  3563.846480: sched_wakeup: comm=gnome-terminal-
pid=5415 prio=120 success=1 target_cpu=001
          <idle>-0     [001] d..3  3563.846489: sched_switch: prev_comm=swapper/1 
prev_pid=0 prev_prio=120 prev_state=R ==> next_comm=gnome-terminal- next_pid=5415 
next_prio=120
     kworker/3:0-8766  [003] d..3  3563.846492: sched_switch: prev_comm=kworker/3:0 
prev_pid=8766 prev_prio=120 prev_state=S ==> next_comm=swapper/3 next_pid=0 next_prio=120
 gnome-terminal--5415  [001] d..3  3563.846639: sched_switch: prev_comm=gnome-terminal-
prev_pid=5415 prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
          <idle>-0     [001] dNh4  3563.846817: sched_wakeup: comm=aprsd pid=3556 prio=110 
success=1 target_cpu=001
          <idle>-0     [001] d..3  3563.846824: sched_switch: prev_comm=swapper/1 
prev_pid=0 prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3556 next_prio=110
```

Snapshot

```
# trace-cmd snapshot -s
# trace-cmd snapshot
# tracer: nop
#
# entries-in-buffer/entries-written: 1747/1747   #P:4
#
#                              _-----=> irqs-off
#                             / _----=> need-resched
#                            | / _---=> hardirq/softirq
#                            || / _--=> preempt-depth
#                            ||| /     delay
#           TASK-PID   CPU#  ||||    TIMESTAMP  FUNCTION
#              | |       |   ||||       |         |
            bash-7623  [003] d..4  3563.846447: sched_wakeup: comm=kworker/3:0 pid=8766
prio=120 success=1 target_cpu=003
            bash-7623  [003] d..3  3563.846471: sched_switch: prev_comm=bash prev_pid=7623
prev_prio=120 prev_state=S ==> next_comm=kworker/3:0 next_pid=8766 next_prio=120
     kworker/3:0-8766  [003] d..4  3563.846480: sched_wakeup: comm=gnome-terminal-
pid=5415 prio=120 success=1 target_cpu=001
          <idle>-0     [001] d..3  3563.846489: sched_switch: prev_comm=swapper/1
prev_pid=0 prev_prio=120 prev_state=R ==> next_comm=gnome-terminal- next_pid=5415
next_prio=120
     kworker/3:0-8766  [003] d..3  3563.846492: sched_switch: prev_comm=kworker/3:0
prev_pid=8766 prev_prio=120 prev_state=S ==> next_comm=swapper/3 next_pid=0 next_prio=120
 gnome-terminal--5415  [001] d..3  3563.846639: sched_switch: prev_comm=gnome-terminal-
prev_pid=5415 prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
          <idle>-0     [001] dNh4  3563.846817: sched_wakeup: comm=aprsd pid=3556 prio=110
success=1 target_cpu=001
          <idle>-0     [001] d..3  3563.846824: sched_switch: prev_comm=swapper/1
prev_pid=0 prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3556 next_prio=110
```

Snapshot

```
# echo 2 > snapshot
# cat snapshot
# tracer: nop
#
#
# * Snapshot is allocated *
#
# Snapshot commands:
# echo 0 > snapshot : Clears and frees snapshot buffer
# echo 1 > snapshot : Allocates snapshot buffer, if not already allocated.
#                      Takes a snapshot of the main buffer.
# echo 2 > snapshot : Clears snapshot buffer (but does not allocate or free)
#                      (Doesn't have to be '2' works with any number that
#                       is not a '0' or '1')
```

Snapshot

```
# trace-cmd snapshot -r
# trace-cmd snapshot
# tracer: nop
#
#
# * Snapshot is allocated *
#
# Snapshot commands:
# echo 0 > snapshot : Clears and frees snapshot buffer
# echo 1 > snapshot : Allocates snapshot buffer, if not already allocated.
#                      Takes a snapshot of the main buffer.
# echo 2 > snapshot : Clears snapshot buffer (but does not allocate or free)
#                      (Doesn't have to be '2' works with any number that
#                       is not a '0' or '1')
```

Snapshot

```
# trace-cmd snapshot -f
# trace-cmd snapshot
# tracer: nop
#
#
# * Snapshot is freed *
#
# Snapshot commands:
# echo 0 > snapshot : Clears and frees snapshot buffer
# echo 1 > snapshot : Allocates snapshot buffer, if not already allocated.
#                      Takes a snapshot of the main buffer.
# echo 2 > snapshot : Clears snapshot buffer (but does not allocate or free)
#                      (Doesn't have to be '2' works with any number that
#                       is not a '0' or '1')
```

Snapshot Trigger

```
# echo 1 > events/sched/sched_switch/enable
# echo 1 > events/irq/irq_handler_entry/enable
# echo ‘snapshot:1 if irq==50’ >   \
      events/irq/irq_handler_exit/trigger
# cat snapshot | tail
          <idle>-0     [000] d..3   350.826053: sched_switch: prev_comm=swapper/0 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=jbd2/dm-1-8 next_pid=337 next_prio=120
           <...>-5504  [001] d..3   350.826066: sched_switch: prev_comm=soffice.bin prev_pid=5504
prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
      goa-daemon-5249  [003] d..3   350.826143: sched_switch: prev_comm=goa-daemon prev_pid=5249
prev_prio=120 prev_state=S ==> next_comm=swapper/3 next_pid=0 next_prio=120
     jbd2/dm-1-8-337   [000] d..3   350.826163: sched_switch: prev_comm=jbd2/dm-1-8 prev_pid=337
prev_prio=120 prev_state=D ==> next_comm=kworker/0:3 next_pid=1059 next_prio=120
          <idle>-0     [001] d..3   350.826508: sched_switch: prev_comm=swapper/1 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3419 next_prio=110
          <idle>-0     [002] d..3   350.826524: sched_switch: prev_comm=swapper/2 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3406 next_prio=120
           aprsd-3419  [001] d..3   350.826541: sched_switch: prev_comm=aprsd prev_pid=3419
prev_prio=110 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
           aprsd-3406  [002] d..3   350.826561: sched_switch: prev_comm=aprsd prev_pid=3406
prev_prio=120 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
           <...>-1059  [000] d..3   350.826956: sched_switch: prev_comm=kworker/0:3 prev_pid=1059
prev_prio=120 prev_state=S ==> next_comm=swapper/0 next_pid=0 next_prio=120
          <idle>-0     [000] d.h2   350.827526: irq_handler_entry: irq=50 name=ahci
```

Snapshot Trigger

```
# trace-cmd start -e sched_switch -e irq_handler_entry \
   -v -e irq_handler_exit -R ‘snapshot:1 if irq==50’
# trace-cmd snapshot | tail
          <idle>-0     [000] d..3   350.826053: sched_switch: prev_comm=swapper/0 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=jbd2/dm-1-8 next_pid=337 next_prio=120
           <...>-5504  [001] d..3   350.826066: sched_switch: prev_comm=soffice.bin prev_pid=5504
prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
      goa-daemon-5249  [003] d..3   350.826143: sched_switch: prev_comm=goa-daemon prev_pid=5249
prev_prio=120 prev_state=S ==> next_comm=swapper/3 next_pid=0 next_prio=120
     jbd2/dm-1-8-337   [000] d..3   350.826163: sched_switch: prev_comm=jbd2/dm-1-8 prev_pid=337
prev_prio=120 prev_state=D ==> next_comm=kworker/0:3 next_pid=1059 next_prio=120
          <idle>-0     [001] d..3   350.826508: sched_switch: prev_comm=swapper/1 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3419 next_prio=110
          <idle>-0     [002] d..3   350.826524: sched_switch: prev_comm=swapper/2 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3406 next_prio=120
           aprsd-3419  [001] d..3   350.826541: sched_switch: prev_comm=aprsd prev_pid=3419
prev_prio=110 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
           aprsd-3406  [002] d..3   350.826561: sched_switch: prev_comm=aprsd prev_pid=3406
prev_prio=120 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
           <...>-1059  [000] d..3   350.826956: sched_switch: prev_comm=kworker/0:3 prev_pid=1059
prev_prio=120 prev_state=S ==> next_comm=swapper/0 next_pid=0 next_prio=120
          <idle>-0     [000] d.h2   350.827526: irq_handler_entry: irq=50 name=ahci
```

Snapshot Trigger

```
# trace-cmd start -e sched_switch -e irq_handler_entry \
   -v -e irq_handler_exit -R ‘snapshot:1 if irq==50’
# trace-cmd snapshot | tail
          <idle>-0     [000] d..3   350.826053: sched_switch: prev_comm=swapper/0 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=jbd2/dm-1-8 next_pid=337 next_prio=120
           <...>-5504  [001] d..3   350.826066: sched_switch: prev_comm=soffice.bin prev_pid=5504
prev_prio=120 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
      goa-daemon-5249  [003] d..3   350.826143: sched_switch: prev_comm=goa-daemon prev_pid=5249
prev_prio=120 prev_state=S ==> next_comm=swapper/3 next_pid=0 next_prio=120
     jbd2/dm-1-8-337   [000] d..3   350.826163: sched_switch: prev_comm=jbd2/dm-1-8 prev_pid=337
prev_prio=120 prev_state=D ==> next_comm=kworker/0:3 next_pid=1059 next_prio=120
          <idle>-0     [001] d..3   350.826508: sched_switch: prev_comm=swapper/1 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3419 next_prio=110
          <idle>-0     [002] d..3   350.826524: sched_switch: prev_comm=swapper/2 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3406 next_prio=120
           aprsd-3419  [001] d..3   350.826541: sched_switch: prev_comm=aprsd prev_pid=3419
prev_prio=110 prev_state=S ==> next_comm=swapper/1 next_pid=0 next_prio=120
           aprsd-3406  [002] d..3   350.826561: sched_switch: prev_comm=aprsd prev_pid=3406
prev_prio=120 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
           <...>-1059  [000] d..3   350.826956: sched_switch: prev_comm=kworker/0:3 prev_pid=1059
prev_prio=120 prev_state=S ==> next_comm=swapper/0 next_pid=0 next_prio=120
          <idle>-0     [000] d.h2   350.827526: irq_handler_entry: irq=50 name=ahci
```

Multiple Trace Buffers

- Created from the “instances” director
  - mkdir “clown”
- Creates a independent trace environment
- Only can enable events (for now)

Multiple Trace Buffers

```
# cd instances
# mkdir clown
# ls clown
buffer_size_kb          free_buffer  snapshot    trace_marker    tracing_cpumask
buffer_total_size_kb  per_cpu     trace trace_options   tracing_on
events                 set_event    trace_clock     trace_pipe
```

Multiple Trace Buffers

```
# cd instances
# mkdir clown
# mkdir car
# echo 1 > clown/events/sched/enable
# echo 1 > car/events/irq/enable
# echo 1 > car/events/sched/sched_wakeup/enable
# cat clown/trace_pipe
CPU:2 [LOST 233789 EVENTS]
          <idle>-0     [002] dN.3  5840.309621: sched_stat_wait: comm=aprsd pid=3556 delay=0 [ns]
          <idle>-0     [002] d..3  5840.309623: sched_switch: prev_comm=swapper/2 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3556 next_prio=110
           aprsd-3556  [002] d..3  5840.309633: sched_stat_runtime: comm=aprsd pid=3556
runtime=21002 [ns] vruntime=340429305929 [ns]
           aprsd-3556  [002] d..3  5840.309635: sched_switch: prev_comm=aprsd prev_pid=3556
prev_prio=110 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
          <idle>-0     [002] d.s4  5840.309665: sched_stat_sleep: comm=rcu_preempt pid=9
delay=2816303 [ns]
          <idle>-0     [002] dNs4  5840.309667: sched_wakeup: comm=rcu_preempt pid=9 prio=120
success=1 target_cpu=002
          <idle>-0     [002] dN.3  5840.309682: sched_stat_wait: comm=rcu_preempt pid=9 delay=8089
[ns]
          <idle>-0     [002] d..3  5840.309684: sched_switch: prev_comm=swapper/2 prev_pid=0
prev_prio=120 prev_state=R ==> next_comm=rcu_preempt next_pid=9 next_prio=120
     rcu_preempt-9     [002] d..3  5840.309692: sched_stat_runtime: comm=rcu_preempt pid=9
runtime=20489 [ns] vruntime=340429372766 [ns]
     rcu_preempt-9     [002] d..3  5840.309704: sched_switch: prev_comm=rcu_preempt prev_pid=9
prev_prio=120 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
          <idle>-0     [002] d.h4  5840.310138: sched_stat_sleep: comm=aprsd pid=3543
delay=1182053 [ns]
```

Multiple Trace Buffers

```
# trace-cmd start -B clown -e sched \
     -B car -e irq -e sched_wakeup
# trace-cmd show -B clown -p
CPU:2 [LOST 233789 EVENTS]
          <idle>-0     [002] dN.3  5840.309621: sched_stat_wait: comm=aprsd pid=3556 delay=0 [ns]
          <idle>-0     [002] d..3  5840.309623: sched_switch: prev_comm=swapper/2 prev_pid=0 
prev_prio=120 prev_state=R ==> next_comm=aprsd next_pid=3556 next_prio=110
           aprsd-3556  [002] d..3  5840.309633: sched_stat_runtime: comm=aprsd pid=3556 
runtime=21002 [ns] vruntime=340429305929 [ns]
           aprsd-3556  [002] d..3  5840.309635: sched_switch: prev_comm=aprsd prev_pid=3556 
prev_prio=110 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
          <idle>-0     [002] d.s4  5840.309665: sched_stat_sleep: comm=rcu_preempt pid=9 
delay=2816303 [ns]
          <idle>-0     [002] dNs4  5840.309667: sched_wakeup: comm=rcu_preempt pid=9 prio=120 
success=1 target_cpu=002
          <idle>-0     [002] dN.3  5840.309682: sched_stat_wait: comm=rcu_preempt pid=9 delay=8089 
[ns]
          <idle>-0     [002] d..3  5840.309684: sched_switch: prev_comm=swapper/2 prev_pid=0 
prev_prio=120 prev_state=R ==> next_comm=rcu_preempt next_pid=9 next_prio=120
     rcu_preempt-9     [002] d..3  5840.309692: sched_stat_runtime: comm=rcu_preempt pid=9 
runtime=20489 [ns] vruntime=340429372766 [ns]
     rcu_preempt-9     [002] d..3  5840.309704: sched_switch: prev_comm=rcu_preempt prev_pid=9 
prev_prio=120 prev_state=S ==> next_comm=swapper/2 next_pid=0 next_prio=120
          <idle>-0     [002] d.h4  5840.310138: sched_stat_sleep: comm=aprsd pid=3543 
delay=1182053 [ns]
```

Multiple Trace Buffers

```
# cat car/trace_pipe
           <...>-7623  [000] d..4  6024.888836: sched_wakeup: comm=rcuop/0 pid=10 prio=120 
success=1 target_cpu=003
          <idle>-0     [001] dNh4  6024.888936: sched_wakeup: comm=aprsd pid=3556 prio=110 
success=1 target_cpu=001
          <idle>-0     [003] dNh4  6024.888940: sched_wakeup: comm=aprsd pid=3543 prio=120 
success=1 target_cpu=003
           <...>-7623  [000] d.h1  6024.888944: irq_handler_entry: irq=52 
name=i915@pci:0000:00:02.0
           <...>-7623  [000] d.h4  6024.888951: sched_wakeup: comm=Xorg pid=4573 prio=120 
success=1 target_cpu=002
           <...>-7623  [000] d.h1  6024.888955: irq_handler_exit: irq=52 ret=handled
           <...>-7623  [000] d..4  6024.889027: sched_wakeup: comm=kworker/0:1 pid=10186 prio=120 
success=1 target_cpu=000
          <idle>-0     [000] d.h2  6024.889229: irq_handler_entry: irq=51 name=iwlwifi
          <idle>-0     [000] d.h2  6024.889233: irq_handler_exit: irq=51 ret=handled
          <idle>-0     [000] dNh4  6024.889240: sched_wakeup: comm=irq/51-iwlwifi pid=1194 prio=49 
success=1 target_cpu=000
  irq/51-iwlwifi-1194  [000] d.s5  6024.889290: sched_wakeup: comm=ax25spyd pid=3287 prio=120 
success=1 target_cpu=001
          <idle>-0     [000] d.h2  6024.889658: irq_handler_entry: irq=52 
name=i915@pci:0000:00:02.0
          <idle>-0     [000] d.h5  6024.889664: sched_wakeup: comm=Xorg pid=4573 prio=120 
success=1 target_cpu=002
          <idle>-0     [000] d.h2  6024.889671: irq_handler_exit: irq=52 ret=handled
          <idle>-0     [003] dNh4  6024.890010: sched_wakeup: comm=aprsd pid=3543 prio=120 
success=1 target_cpu=003
          <idle>-0     [001] dNh4  6024.890010: sched_wakeup: comm=aprsd pid=3556 prio=110 
success=1 target_cpu=001
          <idle>-0     [000] d.h2  6024.890399: irq_handler_entry: irq=52 
name=i915@pci:0000:00:02.0
          <idle>-0     [000] d.h5  6024.890405: sched_wakeup: comm=Xorg pid=4573 prio=120
```

Multiple Trace Buffers

```
# trace-cmd show -B car -p
           <...>-7623  [000] d..4  6024.888836: sched_wakeup: comm=rcuop/0 pid=10 prio=120
success=1 target_cpu=003
          <idle>-0     [001] dNh4  6024.888936: sched_wakeup: comm=aprsd pid=3556 prio=110
success=1 target_cpu=001
          <idle>-0     [003] dNh4  6024.888940: sched_wakeup: comm=aprsd pid=3543 prio=120
success=1 target_cpu=003
           <...>-7623  [000] d.h1  6024.888944: irq_handler_entry: irq=52
name=i915@pci:0000:00:02.0
           <...>-7623  [000] d.h4  6024.888951: sched_wakeup: comm=Xorg pid=4573 prio=120
success=1 target_cpu=002
           <...>-7623  [000] d.h1  6024.888955: irq_handler_exit: irq=52 ret=handled
           <...>-7623  [000] d..4  6024.889027: sched_wakeup: comm=kworker/0:1 pid=10186 prio=120
success=1 target_cpu=000
          <idle>-0     [000] d.h2  6024.889229: irq_handler_entry: irq=51 name=iwlwifi
          <idle>-0     [000] d.h2  6024.889233: irq_handler_exit: irq=51 ret=handled
          <idle>-0     [000] dNh4  6024.889240: sched_wakeup: comm=irq/51-iwlwifi pid=1194 prio=49
success=1 target_cpu=000
  irq/51-iwlwifi-1194  [000] d.s5  6024.889290: sched_wakeup: comm=ax25spyd pid=3287 prio=120
success=1 target_cpu=001
          <idle>-0     [000] d.h2  6024.889658: irq_handler_entry: irq=52
name=i915@pci:0000:00:02.0
          <idle>-0     [000] d.h5  6024.889664: sched_wakeup: comm=Xorg pid=4573 prio=120
success=1 target_cpu=002
          <idle>-0     [000] d.h2  6024.889671: irq_handler_exit: irq=52 ret=handled
          <idle>-0     [003] dNh4  6024.890010: sched_wakeup: comm=aprsd pid=3543 prio=120
success=1 target_cpu=003
          <idle>-0     [001] dNh4  6024.890010: sched_wakeup: comm=aprsd pid=3556 prio=110
success=1 target_cpu=001
          <idle>-0     [000] d.h2  6024.890399: irq_handler_entry: irq=52
name=i915@pci:0000:00:02.0
          <idle>-0     [000] d.h5  6024.890405: sched_wakeup: comm=Xorg pid=4573 prio=120
```

Multiple Trace Buffers

- “triggers” file exists, but!
- It affects the main buffer
- Expect this to change in 3.16 or 3.17
  - Will only affect current instance

Other tricks

- Buffer size
- Per cpu
- trace_marker
- trace_clock

Changing Buffer Size

```
# cat buffer_size_kb
7 (expanded: 1408)

# cat buffer_total_size_kb
28 (expanded: 5632)

# echo 1000 > buffer_size_kb
# cat buffer_size_kb
1000
```

Per CPU

```
# ls per_cpu/
cpu0  cpu1  cpu2  cpu3

# ls per_cpu/cpu0/
buffer_size_kb   snapshot_raw  trace   trace_pipe_raw
snapshot  stats        trace_pipe

# cat per_cpu/cpu0/stats
entries: 35944
overrun: 5068447
commit overrun: 0
bytes: 1441704
oldest event ts:  9303.580084
now ts:  9304.425873
dropped events: 0
read events: 0
```

Trace Marker

```
# echo ‘hello Japan!’ > trace_marker
# cat trace
# tracer: nop
#
# entries-in-buffer/entries-written: 1/1   #P:8
#
#                              _-------=> irqs-off
#                            /  _------=> need-resched
#                            |/  _-----=> need-resched_lazy
#                            ||/  _----=> hardirq/softirq
#                            |||/  _---=> preempt-depth
#                            ||||/  _--=> preempt-lazy-depth
#                            ||||| / _-=> migrate-disable
#                            |||||| /     delay
#           TASK-PID   CPU#  ||||||  TIMESTAMP  FUNCTION
#              | |       |   ||||||     |         |
            bash-24555 [001] ......1 209648.661564: tracing_mark_write: hello Japan!
```

trace_clock

```
# ls trace_clock
[local] global counter uptime perf x86-tsc

# echo counter > trace_clock
# echo function > current_tracer
# cat trace_pipe
         rcuop/2-12    [001] d..2     65492961: preempt_count_sub <-_raw_spin_unlock_irqrestore
         rcuop/2-12    [001] d..1     65492963: rcu_irq_exit <-irq_exit
         rcuop/2-12    [001] ...1     65492966: preempt_count_sub <-_raw_spin_unlock_irqrestore
         rcuop/2-12    [001] ....     65492967: trace_rcu_future_gp.isra.6 <-rcu_nocb_kthread
         rcuop/2-12    [001] ....     65492968: prepare_to_wait_event <-rcu_nocb_kthread
         rcuop/2-12    [001] ....     65492969: _raw_spin_lock_irqsave <-prepare_to_wait_event
         rcuop/2-12    [001] d...     65492970: preempt_count_add <-_raw_spin_lock_irqsave
         rcuop/2-12    [001] d..1     65492972: _raw_spin_unlock_irqrestore
<-prepare_to_wait_event
         rcuop/2-12    [001] ...1     65492973: preempt_count_sub <-_raw_spin_unlock_irqrestore
         rcuop/2-12    [001] ....     65492974: schedule <-rcu_nocb_kthread
         rcuop/2-12    [001] ....     65492975: __schedule <-schedule
         rcuop/2-12    [001] ....     65492977: preempt_count_add <-__schedule
```

Coming in 3.15

- “current_tracer” in instance
- Only allow function tracer
- Can specify specific functions in specific instances

Coming in 3.16

- Different tracers in different instances
- Enable wakeup in one instance
- Enable preemptirqsoff in another
- Limited
  - Some can not be done at same time
  - irqsoff, preemptoff and irqsoff
  - wakeup and wakeup_rt

Questions?

Questions?

Yeah right!
Like we have time