Ftrace: Latency Tracing

Steven Rostedt
srostedt@redhat.com
<rostedt@goodmis.org>
http://people.redhat.com/srostedt/
ftrace-latency-osadl-2009.odp

Ftrace

- Based off of the -rt patch's latency-trace and my own logdev
- Started mainly as a function tracer
- added plugins
- added events
- a basic frame work for kernel tracing

The Debugfs

- Officially mounted at
  - /sys/kernel/debug
- I prefer
  - mkdir /debug
  - mount -t debugfs nodev /debug
  - This presentation will use /debug
- Do what you want

The Tracing Directory

```
# ls /debug/tracing
available_events              printk_formats      trace
available_filter_functions    README              trace_clock
available_tracers             saved_cmdlines      trace_marker
buffer_size_kb                set_event           trace_options
current_tracer                set_ftrace_filter   trace_pipe
dyn_ftrace_total_info         set_ftrace_notrace  trace_stat
events                        set_ftrace_pid
tracing_cpumask               failures
set_graph_function            tracing_enabled
function_profile_enabled      stack_max_size
tracing_max_latency           options             stack_trace
tracing_on                    per_cpu
sysprof_sample_period         tracing_thresh
```

Tracer Plugins

- Found in available_tracers
  - function
  - function_graph
  - wakeup and wakeup_rt
  - irqsoff, preemptoff, preemtirqsoff
  - mmiotrace
  - sched_switch
  - nop

The Function Tracer

```
[root@frodo tracing]# echo function > current_tracer
[root@frodo tracing]# cat trace | head -15
# tracer: function
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
    simpress.bin-2792  [000]   634.280032: unix_poll <-sock_poll
    simpress.bin-2792  [000]   634.280033: sock_poll_wait <-unix_poll
    simpress.bin-2792  [000]   634.280033: fput <-do_sys_poll
    simpress.bin-2792  [000]   634.280034: fget_light <-do_sys_poll
    simpress.bin-2792  [000]   634.280035: sock_poll <-do_sys_poll
    simpress.bin-2792  [000]   634.280035: unix_poll <-sock_poll
    simpress.bin-2792  [000]   634.280036: sock_poll_wait <-unix_poll
    simpress.bin-2792  [000]   634.280037: fput <-do_sys_poll
    simpress.bin-2792  [000]   634.280038: fget_light <-do_sys_poll
    simpress.bin-2792  [000]   634.280038: sock_poll <-do_sys_poll
    simpress.bin-2792  [000]   634.280039: unix_poll <-sock_poll
```

set_ftrace_filter

```
[root@frodo tracing]# echo schedule > set_ftrace_filter
[root@frodo tracing]# cat set_ftrace_filter
schedule
[root@frodo tracing]# echo function > current_tracer
[root@frodo tracing]# cat trace | head -15
# tracer: function
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
            Xorg-1849  [001]   883.657737: schedule <-schedule_hrtimeout_range
          <idle>-0     [001]   883.658534: schedule <-cpu_idle
            Xorg-1849  [001]   883.658612: schedule <-__cond_resched
     kondemand/1-1239  [001]   883.658632: schedule <-worker_thread
            Xorg-1849  [001]   883.659384: schedule <-sysret_careful
            Xorg-1849  [001]   883.659479: schedule <-schedule_hrtimeout_range
  gnome-terminal-2112  [001]   883.660053: schedule <-schedule_hrtimeout_range
            Xorg-1849  [001]   883.660281: schedule <-schedule_hrtimeout_range
            Xorg-1849  [001]   883.660293: schedule <-schedule_hrtimeout_range
  gnome-terminal-2112  [001]   883.660409: schedule <-schedule_hrtimeout_range
            Xorg-1849  [001]   883.660458: schedule <-sysret_careful
```

set_ftrace_filter (Continued)

```
[root@frodo tracing]# echo schedule_tail >> set_ftrace_filter
[root@frodo tracing]# cat set_ftrace_filter
schedule_tail
schedule
[root@frodo tracing]# echo 'sched*' > set_ftrace_filter
[root@frodo tracing]# cat set_ftrace_filter | head -10
sched_avg_update
sched_group_shares
sched_group_rt_runtime
sched_group_rt_period
sched_slice
sched_rt_can_attach
sched_feat_open
sched_debug_open
sched_feat_show
sched_feat_write
```

Acceptable Globs

- match*
  - Selects all functions starting with “match”
- *match
  - Selects all functions ending with “match”
- *match*
  - Selects all functions with “match” in its name

set_ftrace_notrace

```
[root@frodo tracing]# echo > set_ftrace_filter
[root@frodo tracing]# echo '*lock*' > set_ftrace_notrace
[root@frodo tracing]# cat set_ftrace_notrace | head -10
xen_pte_unlock
alternatives_smp_unlock
user_enable_block_step
__acpi_release_global_lock
__acpi_acquire_global_lock
unlock_vector_lock
lock_vector_lock
parse_no_kvmclock
kvm_set_wallclock
kvm_register_clock
```

The Function Graph Tracer

```
[root@frodo tracing]# echo function_graph > current_tracer
[root@frodo tracing]# cat trace | head -20
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 1)               |  down_read_trylock() {
 1)   0.487 us    |    _spin_lock_irqsave();
 1)   0.409 us    |    _spin_unlock_irqrestore();
 1)   2.519 us    |  }
 1)   0.420 us    |  __might_sleep();
 1)   0.415 us    |  _cond_resched();
 1)   0.415 us    |  find_vma();
 1)               |  handle_mm_fault() {
 1)   0.421 us    |    pud_alloc();
 1)   0.409 us    |    pmd_alloc();
 1)               |    __do_fault() {
 1)               |      filemap_fault() {
 1)               |        find_get_page() {
 1)   0.571 us    |          page_cache_get_speculative();
 1)   1.630 us    |        }
 1)               |        lock_page() {
```

What Does That Function Call?

```
[root@frodo tracing]# echo sys_read > set_graph_function
[root@frodo tracing]# cat trace | head -20
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 1)   1.888 us    |          fsnotify();
 1) + 10.016 us   |        }
 1) ! 116.994 us  |      }
 1)   0.920 us    |      fput_light();
 1) ! 122.158 us  |    }
 1)               |    sys_read() {
 1)   1.149 us    |      fget_light();
 1)               |      vfs_read() {
 1)               |        rw_verify_area() {
 1)               |          security_file_permission() {
 1)               |            selinux_file_permission() {
 1)   0.781 us    |              avc_policy_seqno();
 1)   2.435 us    |            }
 1)   4.046 us    |          }
 1)   5.675 us    |        }
 1)               |        tty_read() {
```

Interrupts in Function Graph

```
 1)               |            __sock_recvmsg() {
 1)               |              security_socket_recvmsg() {
 1)               |                selinux_socket_recvmsg() {
 1)   ==========> |
 1)               |                  smp_apic_timer_interrupt() {
 1)               |                    apic_write() {
 1)   0.769 us    |                      native_apic_mem_write();
 1)   2.441 us    |                    }
 1)   0.800 us    |                    exit_idle();
 1)               |                    irq_enter() {
 1)   0.806 us    |                      rcu_irq_enter();
 1)   0.782 us    |                      idle_cpu();
 1)   3.991 us    |                    }
 1)               |                    hrtimer_interrupt() {
 1)               |                      ktime_get() {
 1)               |                        timekeeping_get_ns() {
 1)   1.773 us    |                          read_hpet();
                        [...]
 1) + 36.095 us   |                      }
 1) + 37.934 us   |                    }
 1)   0.800 us    |                    rcu_irq_exit();
 1)   0.776 us    |                    idle_cpu();
 1) + 42.803 us   |                  }
 1) ! 169.218 us  |                }
 1)   <========== |
 1)               |                socket_has_perm() {
 1)               |                  avc_has_perm() {
```

Trace Events

```
[root@frodo tracing]# ls events
block   ext4    header_event  irq   kmem  kvmmmu  sched  syscalls
enable  ftrace  header_page   jbd2  kvm   module  skb    workqueue

[root@frodo tracing]# ls events/sched/
enable                 sched_process_exit  sched_stat_iowait  sched_wakeup
filter                 sched_process_fork  sched_stat_sleep
sched_wakeup_new
sched_kthread_stop     sched_process_free  sched_stat_wait
sched_kthread_stop_ret sched_process_wait  sched_switch
sched_migrate_task     sched_signal_send   sched_wait_task

[root@frodo tracing]# ls events/sched/sched_wakeup
enable  filter  format  id
```

Enable a Single Event

```
[root@frodo tracing]# echo 1 > events/sched/sched_wakeup/enable
[root@frodo tracing]# cat trace | head -10
# tracer: nop
#
#          TASK-PID    CPU#    TIMESTAMP  FUNCTION
#             | |       |          |         |
          bash-2613  [001]  425.078164: sched_wakeup: task bash:2613 [120] success=0 [001]
          bash-2613  [001]  425.078184: sched_wakeup: task bash:2613 [120] success=0 [001]
          bash-2613  [001]  425.078572: sched_wakeup: task bash:2613 [120] success=0 [001]
          bash-2613  [001]  425.078660: sched_wakeup: task bash:2613 [120] success=0 [001]
        <idle>-0     [001]  425.078930: sched_wakeup: task events/1:10 [120] success=1 [001]
      events/1-10    [001]  425.078941: sched_wakeup: task gnome-terminal:2162 [120]
success=1 [001]
```

Enable All Subsystem Events

```
[root@frodo tracing]# echo 1 > events/sched/enable
[root@frodo tracing]# cat trace | head -10
# tracer: nop
#
#       TASK-PID    CPU#    TIMESTAMP  FUNCTION
#         | |       |          |         |
    events/0-9     [000]   638.042792: sched_switch: task events/0:9 [120] (S) ==> kondemand/0:1305 [120]
 kondemand/0-1305  [000]   638.042796: sched_stat_wait: task: restorecond:1395 wait: 15023 [ns]
 kondemand/0-1305  [000]   638.042797: sched_switch: task kondemand/0:1305 [120] (S) ==> restorecond:1395 [120]
 restorecond-1395  [000]   638.051758: sched_stat_wait: task: restorecond:1395 wait: 0 [ns]
 restorecond-1395  [000]   638.052758: sched_stat_sleep: task: kondemand/0:1305 sleep: 9966692 [ns]
 restorecond-1395  [000]   638.052760: sched_wakeup: task kondemand/0:1305 [120] success=1 [000]
```

Enable All Events

```
[root@frodo tracing]# echo 1 > events/enable
[root@frodo tracing]# cat trace | head -10
# tracer: nop
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
           acpid-1470  [001]   794.947181: kfree: call_site=ffffffff810c996d
ptr=(null)
           acpid-1470  [001]   794.947182: sys_read -> 0x1
           acpid-1470  [001]   794.947183: sys_exit: NR 0 = 1
           acpid-1470  [001]   794.947184: sys_read(fd: 3, buf: 7f4ebb32ac50,
count: 1)
           acpid-1470  [001]   794.947185: sys_enter: NR 0 (3, 7f4ebb32ac50,
1, 8, 40, 101010101010101)
           acpid-1470  [001]   794.947186: kfree: call_site=ffffffff810c996d
ptr=(null)
```

Enable Multiple Events

```
[root@frodo tracing]# echo 1 > events/sched/sched_wakeup/enable
[root@frodo tracing]# echo 1 > events/sched/sched_wakeup_new/enable
[root@frodo tracing]# echo 1 > events/sched/sched_switch/enable
[root@frodo tracing]# cat trace | head -15
# tracer: nop
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
            bash-2913  [001]   574.988228: sched_wakeup: task bash:2913 [120] success=0 [001]
            bash-2913  [001]   574.988264: sched_wakeup: task bash:2913 [120] success=0 [001]
            bash-2913  [001]   574.988425: sched_wakeup: task bash:2913 [120] success=0 [001]
            bash-2913  [001]   574.988440: sched_switch: task bash:2913 [120] (S) ==> swapper:0 [140]
          <idle>-0     [001]   574.988744: sched_wakeup: task events/1:10 [120] success=1 [001]
          <idle>-0     [001]   574.988754: sched_switch: task swapper:0 [140] (R) ==> events/1:10 [120]
        events/1-10    [001]   574.988760: sched_wakeup: task gnome-terminal:2158 [120] success=1 [001]
        events/1-10    [001]   574.988764: sched_switch: task events/1:10 [120] (S) ==> gnome-terminal:2158
[120]
  gnome-terminal-2158  [001]   574.988855: sched_switch: task gnome-terminal:2158 [120] (S) ==> swapper:0
[140]
          <idle>-0     [000]   574.991204: sched_wakeup: task phy0:1041 [120] success=1 [000]
          <idle>-0     [000]   574.991211: sched_switch: task swapper:0 [140] (R) ==> phy0:1041 [120]
```

Plugins vs Events

- Plugins are set via current_tracer
  - Events are enabled via the event directory or the set_event file
- Plugins are listed via the available_tracers file
  - Events are listed by the event directory or the available_events file
- Only one plugin at a time
  - Any number of events can be enabled
  - They show up in any trace

Latency Tracers

- wakeup
  - trace wake up time high highest prio task
- wakeup_rt
  - trace wake up time of highest prio RT task
- irqsoff
  - trace time interrupts is disabled
- preemptoff
  - trace time preemption is disabled
- preemptirqsoff
  - trace time preemption or interrupts disabled

Mixing Events With Plugins

- Latency tracers work best with seeing what is happening
- Function tracer is too verbose. Although filtering may help
- Just start and stop is not enough

irqsoff default

```
[root@frodo tracing]# echo irqsoff > current_tracer
[root@frodo tracing]# cat trace
# tracer: irqsoff
#
# irqsoff latency trace v1.1.5 on 2.6.31-git
# --------------------------------------------------------------------
# latency: 366 us, #82/82, CPU#1 | (M:desktop VP:0, KP:0, SP:0 HP:0 #P:2)
#    -----------------
#    | task: -13867 (uid:500 nice:0 policy:0 rt_prio:0)
#    -----------------
#  => started at: save_args
#  => ended at:   call_softirq
#
#
#                  _------=> CPU#
#                 / _-----=> irqs-off
#                | / _----=> need-resched
#                || / _---=> hardirq/softirq
#                ||| / _--=> preempt-depth
#                |||| /_--=> lock-depth
#                |||||/     delay
#  cmd     pid   |||||| time  |   caller
#     \   /      ||||||   \   |   /
     cc1-13867   1d....    0us : trace_hardirqs_off_thunk <-save_args
     cc1-13867   1d....    0us : smp_apic_timer_interrupt <-apic_timer_interrupt
     cc1-13867   1d....    1us : apic_write <-smp_apic_timer_interrupt
     cc1-13867   1d....    1us : native_apic_mem_write <-apic_write
     cc1-13867   1d....    1us : exit_idle <-smp_apic_timer_interrupt
     cc1-13867   1d....    2us : irq_enter <-smp_apic_timer_interrupt
<70 lines deleted>
     cc1-13867   1dNh..  365us : irq_exit <-smp_apic_timer_interrupt
     cc1-13867   1dN...  365us : do_softirq <-irq_exit
     cc1-13867   1dN...  365us : __do_softirq <-call_softirq
     cc1-13867   1dN...  366us : __local_bh_disable <-__do_softirq
     cc1-13867   1dNs..  366us : __do_softirq <-call_softirq
     cc1-13867   1dNs..  367us : trace_hardirqs_on <-call_softirq
```

Disable Function Tracer

```
[root@frodo tracing]# echo 0 > /proc/sys/kernel/ftrace_enabled
[root@frodo tracing]# echo 0 > tracing_max_latency
[root@frodo tracing]# cat trace
# tracer: irqsoff
#
# irqsoff latency trace v1.1.5 on 2.6.31-git
# --------------------------------------------------------------------
# latency: 426 us, #3/3, CPU#0 | (M:desktop VP:0, KP:0, SP:0 HP:0
#P:2)
#    -----------------
#    | task: -1844 (uid:0 nice:0 policy:0 rt_prio:0)
#    -----------------
#  => started at: kcalloc.clone.1
#  => ended at:   kcalloc.clone.1
#
#
#                  _------=> CPU#
#                 / _-----=> irqs-off
#                | / _----=> need-resched
#                || / _---=> hardirq/softirq
#                ||| / _--=> preempt-depth
#                |||| /_--=> lock-depth
#                |||||/     delay
#  cmd     pid   |||||| time  |   caller
#     \   /      ||||||   \   |   /
    Xorg-1844    0d..0.    0us!: __kmalloc <-kcalloc.clone.1
    Xorg-1844    0d..0.  426us : __kmalloc <-kcalloc.clone.1
    Xorg-1844    0d..0.  427us : trace_hardirqs_on <-kcalloc.clone.1
```

Events with irqsoff

```
[root@frodo tracing]# echo 1 > events/enable
[root@frodo tracing]# echo 0 > tracing_max_latency
[root@frodo tracing]# cat trace
# tracer: irqsoff
#
# irqsoff latency trace v1.1.5 on 2.6.31-git
# --------------------------------------------------------------------
# latency: 336 us, #6/6, CPU#0 | (M:desktop VP:0, KP:0, SP:0 HP:0 #P:2)
#    -----------------
#    | task: -33 (uid:0 nice:0 policy:0 rt_prio:0)
#    -----------------
#  => started at: __queue_work
#  => ended at:   __queue_work
#
#
#                  _------=> CPU#
#                 / _-----=> irqs-off
#                | / _----=> need-resched
#                || / _---=> hardirq/softirq
#                ||| / _--=> preempt-depth
#                |||| /_--=> lock-depth
#                |||||/     delay
#  cmd     pid   |||||| time  |   caller
#     \   /      ||||||   \   |   /
 kswapd0-33      0d.s..    0us+: _spin_lock_irqsave <-__queue_work
 kswapd0-33      0d.s..    2us+: workqueue_insertion: thread=kondemand/0:1233
func=do_dbs_timer+0x0/0x273 [cpufreq_ondemand]
 kswapd0-33      0d.s..    6us : sched_stat_sleep: task: kondemand/0:1233 sleep:
9906010 [ns]
 kswapd0-33      0d.s..    7us!: sched_wakeup: task kondemand/0:1233 [120] success=1
[000]
 kswapd0-33      0dNs..  336us : _spin_unlock_irqrestore <-__queue_work
 kswapd0-33      0dNs..  337us : trace_hardirqs_on <-__queue_work
```

```
[root@frodo tracing]# echo 0 > /proc/sys/kernel/ftrace_enabled
[root@frodo tracing]# echo wakeup > current_tracer
[root@frodo tracing]# echo 1 > events/enable
[root@frodo tracing]# echo 0 > tracing_max_latency
[root@frodo tracing]# cat trace
# tracer: wakeup
#
# wakeup latency trace v1.1.5 on 2.6.31
# --------------------------------------------------------------------
# latency: 1440 us, #356/356, CPU#0 | (M:preempt VP:0, KP:0, SP:0 HP:0 #P:2)
#    -----------------
#    | task: -1255 (uid:0 nice:0 policy:0 rt_prio:0)
#    -----------------
#
#                  _------=> CPU#
#                 / _-----=> irqs-off
#                | / _----=> need-resched
#                || / _---=> hardirq/softirq
#                ||| / _--=> preempt-depth
#                |||| /_--=> lock-depth
#                |||||/     delay
#  cmd     pid   |||||| time  |   caller
#     \   /      ||||||   \   |   /
  <idle>-0       0dNs6.    1us :      0:140:R   + [000]  1255:120:S kondemand/0
  <idle>-0       0dNs6.    1us : default_wake_function <-autoremove_wake_function
  <idle>-0       0dNs6.    2us+: sched_wakeup: task kondemand/0:1255 [120] success=1 [000]
  <idle>-0       0.Ns3.    5us : softirq_exit: softirq=1 action=TIMER
  <idle>-0       0.Ns3.    6us+: softirq_entry: softirq=9 action=RCU
  <idle>-0       0.Ns3.    8us+: softirq_exit: softirq=9 action=RCU
  <idle>-0       0d..3.   16us+: sched_stat_wait: task: multiload-apple:2178 wait: 38125 [ns]
  <idle>-0       0d..3.   19us+: sched_switch: task swapper:0 [140] (R) ==> multiload-apple:2178 [120]
multiloa-2178    0...1.   25us+: kmem_cache_free: call_site=ffffffffa046df20 ptr=ffff880068b993c0
multiloa-2178    0...1.   27us : kfree: call_site=ffffffff810c4ba4 ptr=(null)
multiloa-2178    0d....   28us : sys_statfs -> 0x0

[172 lines deleted]

multiloa-2178    0d..3.  627us+: sched_stat_wait: task: Xorg:1859 wait: 53354 [ns]
multiloa-2178    0d..3.  630us+: sched_switch: task multiload-apple:2178 [120] (S) ==> Xorg:1859 [120]
    Xorg-1859    0...1.  674us : kfree: call_site=ffffffff810c4ba4 ptr=(null)
    Xorg-1859    0d....  675us : sys_select -> 0x1
    Xorg-1859    0...1.  677us+: sys_exit: NR 23 = 1

[82 lines deleted]

    Xorg-1859    0d..3. 1426us+: sched_stat_wait: task: multiload-apple:2178 wait: 710554 [ns]
    Xorg-1859    0d..3. 1428us+: sched_switch: task Xorg:1859 [120] (S) ==> multiload-apple:2178 [120]
multiloa-2178    0d..3. 1437us+: sched_stat_wait: task: kondemand/0:1255 wait: 1439391 [ns]
multiloa-2178    0d..3. 1438us : schedule <-schedule_hrtimeout_range
multiloa-2178    0d..3. 1439us :   2178:120:S ==> [000]  1255:120:R kondemand/0
```

Tracing RT Task Wakeups

```
# tracer: wakeup_rt
#
# wakeup_rt latency trace v1.1.5 on 2.6.31
# --------------------------------------------------------------------
# latency: 21 us, #4/4, CPU#1 | (M:preempt VP:0, KP:0, SP:0 HP:0 #P:2)
#    -----------------
#    | task: -6 (uid:0 nice:0 policy:1 rt_prio:99)
#    -----------------
#
#                 _------=> CPU#
#                / _-----=> irqs-off
#               | / _----=> need-resched
#               || / _---=> hardirq/softirq
#               ||| / _--=> preempt-depth
#               |||| /_--=> lock-depth
#               |||||/     delay
#  cmd     pid  |||||| time  |   caller
#     \   /     ||||||  \    |   /
  <idle>-0       0d.s4.    2us+:      0:140:R   + [001]     6:  0:S migration/1
  <idle>-0       0d.s4.    6us+: wake_up_process <-rebalance_domains
multiloa-2260    1d..3.   17us+: schedule <-retint_careful
multiloa-2260    1d..3.   19us :   2260:120:R ==> [001]     6:  0:R migration/1
```

Trace Options

```
[root@frodo tracing]# ls options/
annotate  context-info    latency-format   sched-tree  sym-offset    verbose
bin       ftrace_preempt  printk-msg-only  sleep-time  sym-userobj
block     graph-time      print-parent     stacktrace  trace_printk
branch    hex             raw              sym-addr    userstacktrace
```

latency-format

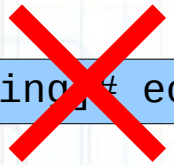
```
# tracer: nop
#
# nop latency trace v1.1.5 on 2.6.31-git
# --------------------------------------------------------------------
# latency: 0 us, #52053/59398, CPU#0 | (M:desktop VP:0, KP:0, SP:0 HP:0 #P:2)
#    -----------------
#    | task: -0 (uid:0 nice:0 policy:0 rt_prio:0)
#    -----------------
#
#                  _------=> CPU#
#                 / _-----=> irqs-off
#                | / _----=> need-resched
#                || / _---=> hardirq/softirq
#                ||| / _--=> preempt-depth
#                |||| /_--=> lock-depth
#                |||||/     delay
#  cmd     pid   |||||| time  |   caller
#     \   /      ||||||   \   |   /
    bash-2758    0..... 149210us+: kmalloc: call_site=ffffffff810d982c
ptr=ffff88002e53c600 bytes_req=49 bytes_alloc=64 gfp_flags=GFP_KERNEL
    bash-2758    0..... 149214us+: kmalloc: call_site=ffffffff810d94d4
ptr=ffff88002f854540 bytes_req=32 bytes_alloc=32 gfp_flags=GFP_KERNEL
```

tracing_on

```
[root@frodo tracing]# echo 0 > tracing_on
```

```
[root@frodo tracing]# echo 1 > tracing_on
```

```
[root@frodo tracing]# echo 0> tracing_on
```



```
[root@frodo tracing]# echo 0 > tracing_on; run_test; echo 0 > tracing_off
```

trace_marker

```
[root@frodo tracing]# echo 'Hallo Deutschland!' > trace_marker
[root@frodo tracing]# cat trace
# tracer: nop
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
           bash-3798  [001]  1739.916994: 0: Hallo Deutschland!
```

tracing_on and trace_marker

```
int trace_on_fd;
int trace_mark_fd;

int main(int argc, char *argv[])
{
        char buf[BUFSIZ];

        [...]

        find_debugfs(buf);
        strcat(buf, "/tracing/tracing_on");
        trace_on_fd = open(buf, O_WRONLY);

        find_debugfs(buf);
        strcat(buf, "/tracing/trace_marker");
        trace_mark_fd = open(buf, O_WRONLY);

        [...]

        write(trace_mark_fd,
              "Testing for error\n", 18);

        if (error_detected()) {
                /* hit bug */
                write(trace_on_fd, "0", 1);
        }
```

traceon / traceoff

- Uses the function tracer to start or stop tracing
- Can start or stop a given number of times
- Add to functions called in the error path

traceon / traceoff

```
[root@frodo tracing]# echo bad_inode_create:traceoff > set_ftrace_filter
[root@frodo tracing]# cat set_ftrace_filter
#### all functions enabled ####
bad_inode_create:traceoff:unlimited
```

Function Latencies

```
[root@frodo tracing]# echo do_IRQ > set_ftrace_filter
[root@frodo tracing]# echo function_graph > current_tracer
[root@frodo tracing]# echo 1 > events/irq/irq_handler_entry/enable
[root@frodo tracing]# cat trace | head -20
# tracer: function_graph
#
# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=0 handler=timer */
 0) + 36.827 us   |  }
 0)   <========== |
 ------------------------------------------
 0)    <idle>-0    =>   ssh-16695
 ------------------------------------------

 0)   ==========> |
 0)               |  do_IRQ() {
 0)               |  /* irq_handler_entry: irq=14 handler=ata_piix */
 0) + 41.619 us   |  }
 0)   <========== |
 ------------------------------------------
 0)   ssh-16695    =>    <idle>-0
```

Future of Ftrace

- Dynamic trace points.
  - Work of Masami Hiramatsu
  - Like a trace printk but no need to recompile

No need anymore to write some printk to debug, worrying, sweating, feeling guilty because we know we'll need yet another printk() after the reboot, and we even already know where while it is compiling.

We would build less kernels, then drink less coffee, becoming less nervous, more friendly. Everyone will offer flowers in the street, the icebergs will grow back and white bears will...

And eventually we'll be inspired enough to write perf love, the more than expected tool to post process ftrace "love" events.

-- Frederic Weisbecker

Future of Ftrace

- Visual (GUI) tools
- More connectivity with perf

Questions?

- Yeah right, like we have time!