LinuxCon Brazil 2010

Linux on Real-Time Embedded Systems

Tips and Tools

Daniel Bristot de Oliveira
<daniel@bristot.eti.br>
bristot@OFTC / #linux-rt

Agenda

Concepts

Common problems

Hardware tips

Real Time Linux implementations

PREMPT-RT/Vanilla Features

Tracing

Tips Tips and more tips

What is Real-Time ?

Event -> Response

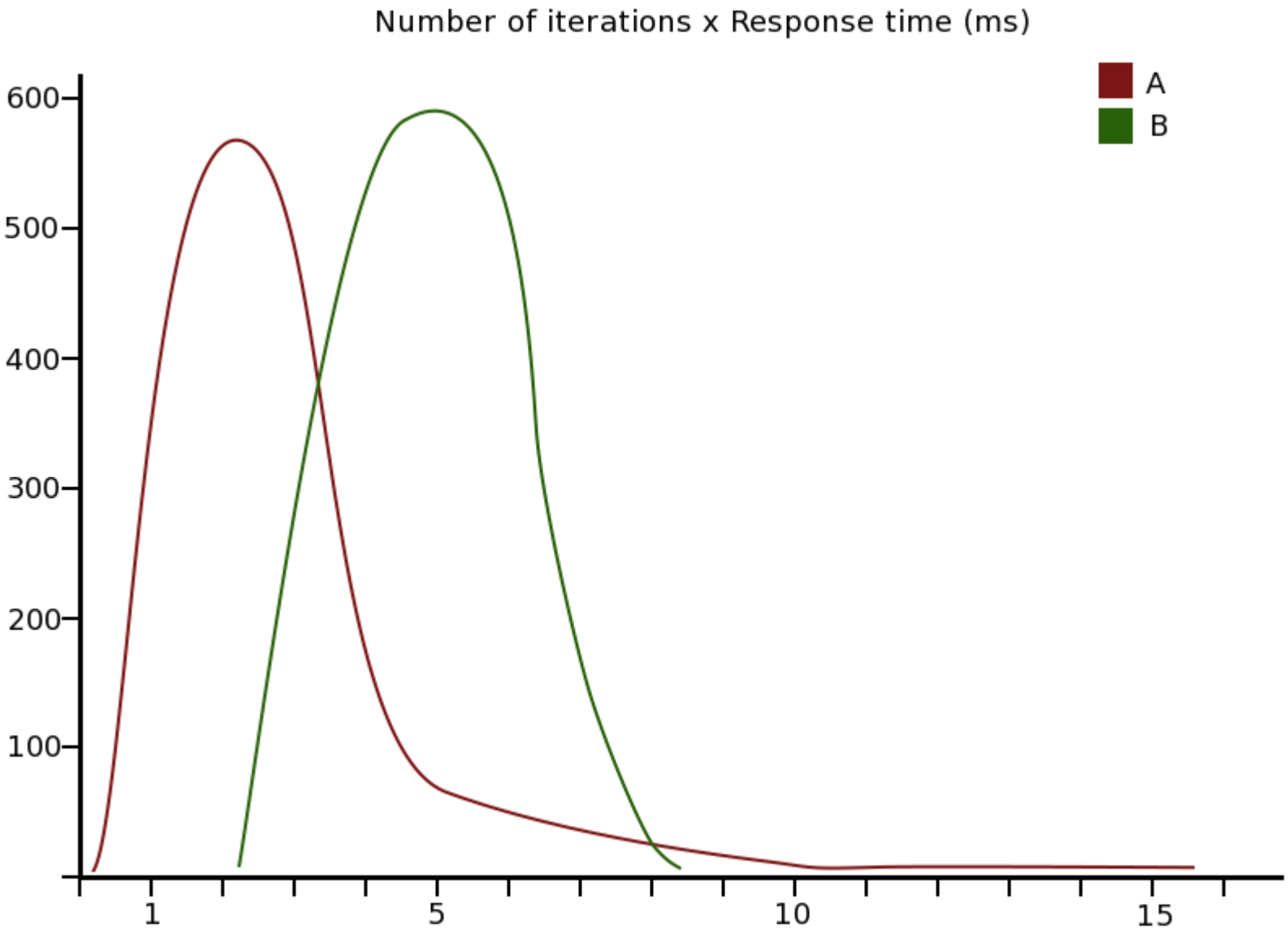
Response time or “latency”

When we have a deadline...

Common examples:

- Video Game
- Telecommunication device
- Air Craft

RT != FAST

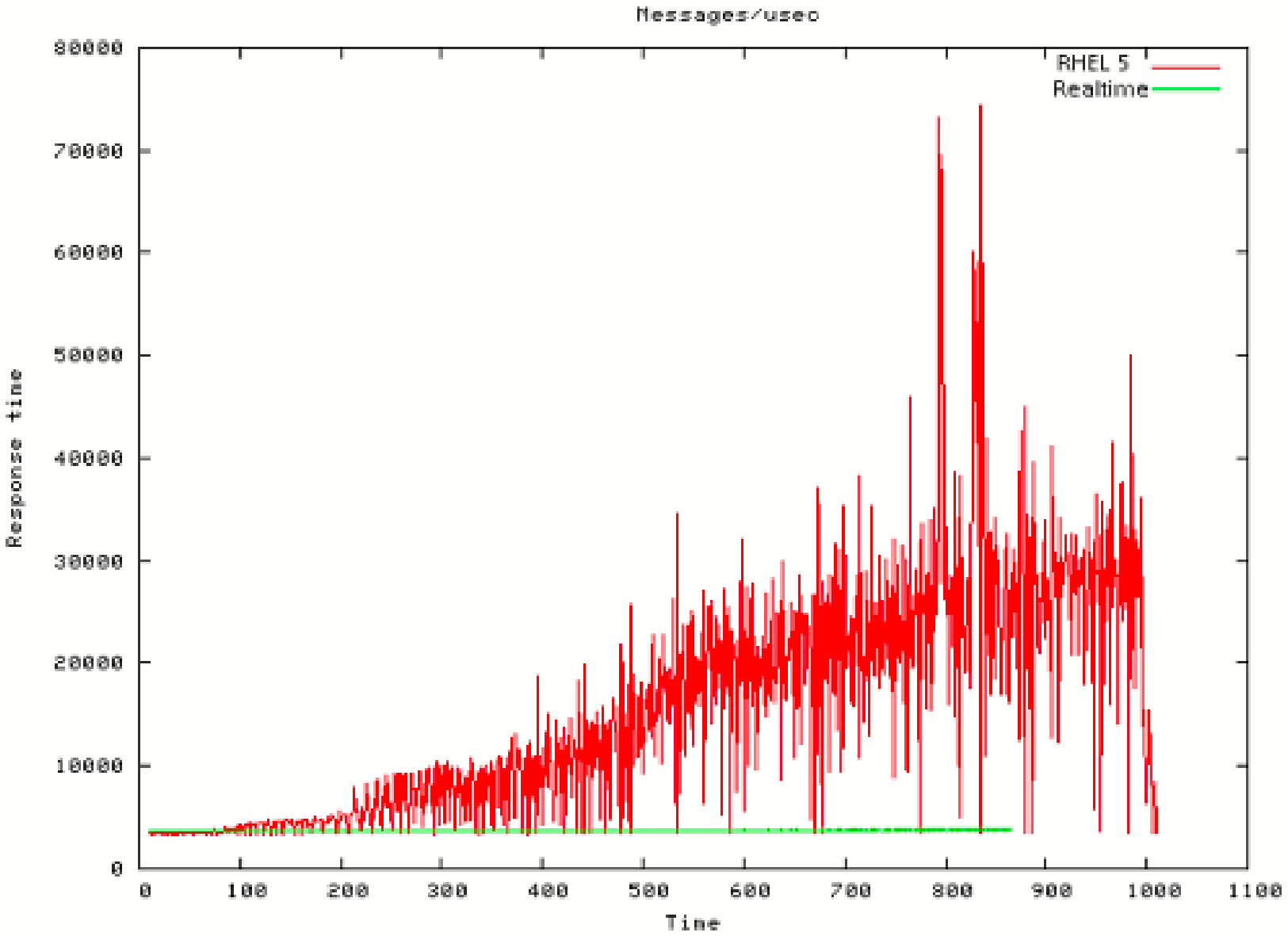

Number of iterations x Response time (ms)
A
B
600
500
400
300
200
100
1
5
10
15

Real-Time means...

Determinism

- Know WCET
- Know reactions time
- Reliable results
- Repeatability

RHEL 5 vs RH MRG



Embedded market changes

Embedded != One application

Multitask

Real-Time applications

- Video, Voice, Chat...

Non Real-Time applications

- Agenda, text editors, E-Mail...

Sharing Resources

Embedded market still

Low cost devices

- Small processor
- Small bandwidth bus/networks
- Low power consumptions

That's the fun :)

Common Problems

Hardware

- Bus sharing / latency
- Timer resolution

Operating Systems

- Sources of indeterminism
- Interrupts Handling
- Process Scheduling

How trace the system

A little about hardware

“Daniel only know how to write to memory address and ask for more resources”
-- The Hardware guy

IRQ Prioritization

DMA Scheduling

- Prioritization
- Burst
- Wait position

High Resolution Timers

What more? Depend of your project...

Linux is a...

Soft Real-Time Operating System...

Why?

But... there are some crazy guys trying to change it...

Welcome :)

And yeah, I still think the hard-RT people are mostly crazy.

-- Linus

Real-Time Linux

Task Forces

RTAI/Xenomai style:

- Trap/interrupt abstraction layer
- Hard Real-Time kernel
- Linux Kernel

PREMPT-RT style:

- Linux kernel
- Change on some sub-systems

Why PREMPT-RT?

All sub-system in Real-Time

Hardware support? All Linux supported

Many Features joined in mainline

Joe Average Programmer :-)

High Resolution Timers

~~The 10ms myth~~

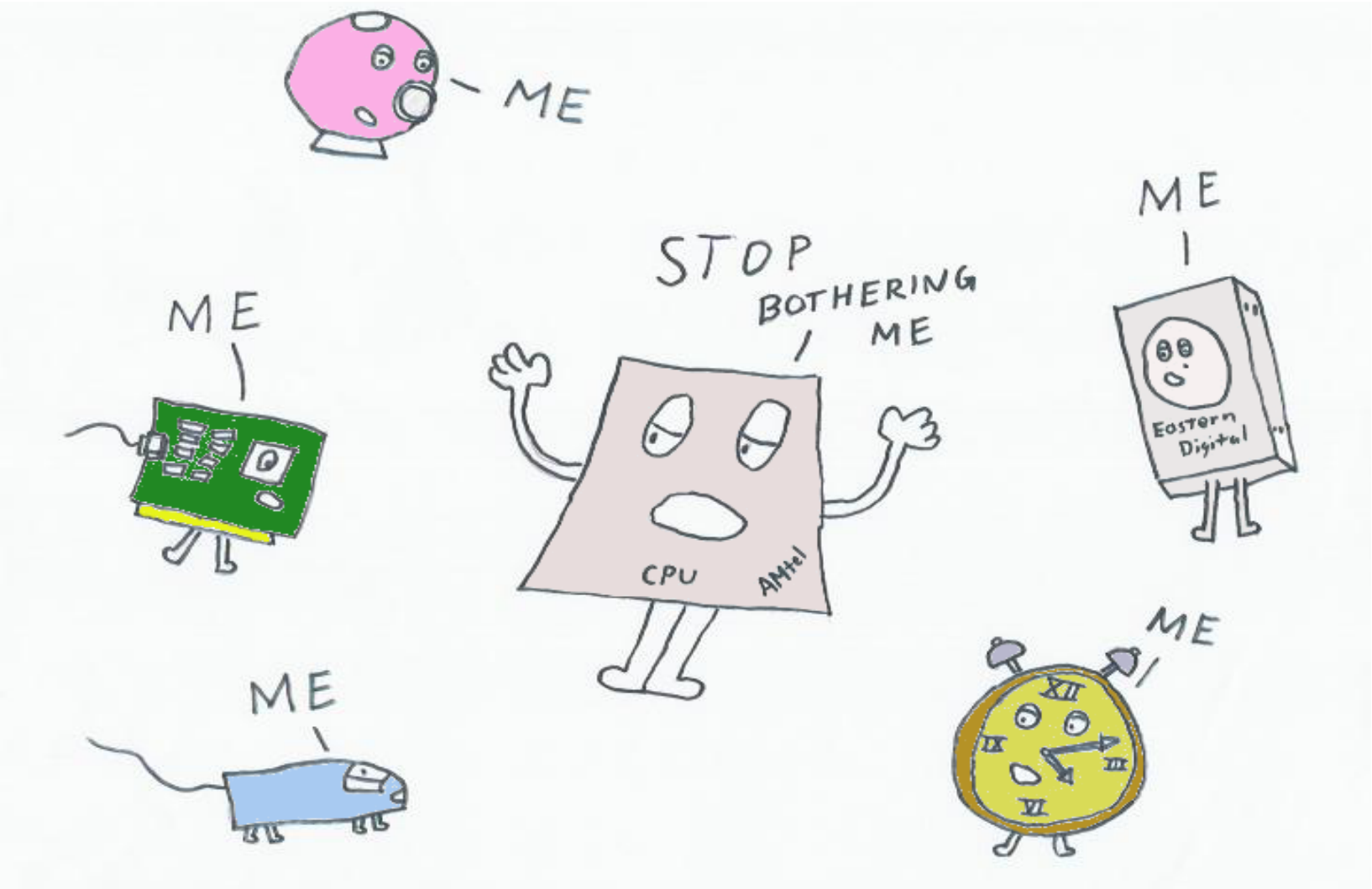
Timer resolution = 1ns

Precision ~ 1us / 20us

Hardware dependent

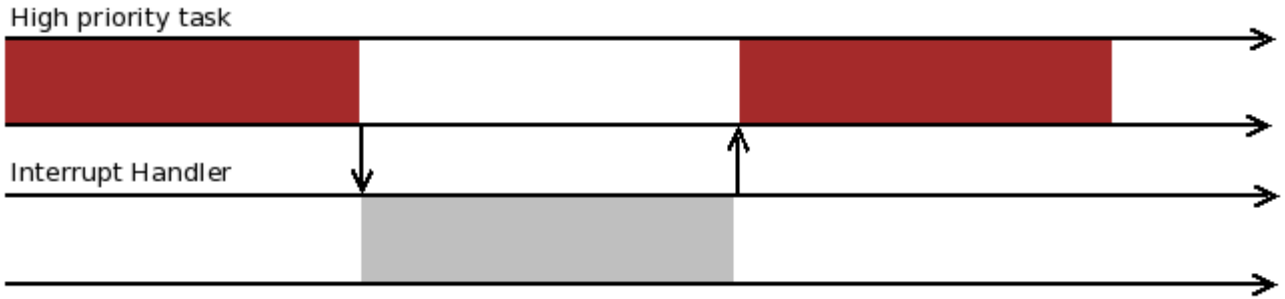
NO_HZ = Power saving without of lost time

Interrupt

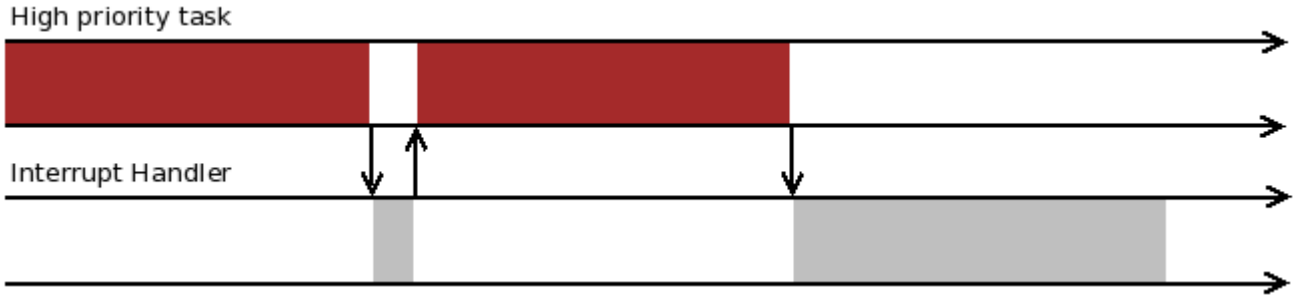


Interrupt as threads

Normal interrupt handling:

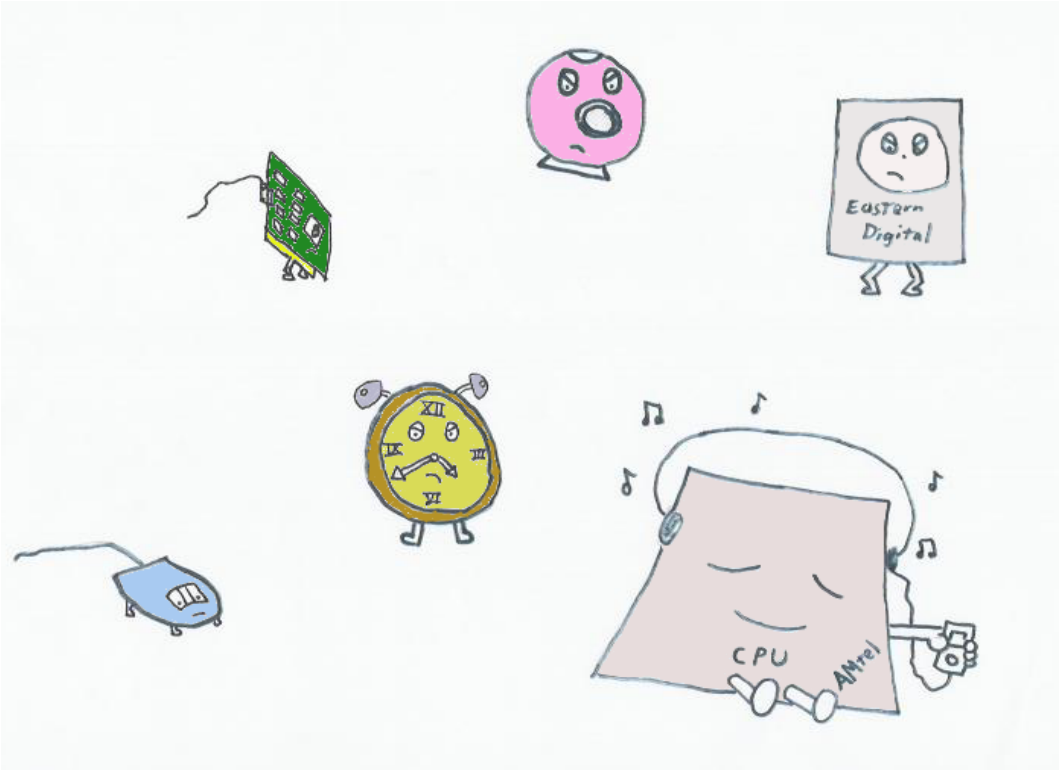


As thread:



Interrupt as threads

```
PID   USER     COMMAND
    1 root     init
          [...]
  314 root     [IRQ-21]
  323 root     [mtdblockd]
  331 root     [IRQ-20]
  340 root     [IRQ-62]
  341 root     [IRQ-63]
  347 root     [IRQ-1]
  349 root     [IRQ-9]
  356 root     [IRQ-13]
```



Linux Schedulers

CFS Sched:

- SCHED_IDLE - BATCH - NORMAL (nice)

Real Time Sched:

- SCHED_FIFO (chrt -f PRIO PID)
- SCHED_RR (chrt -r PRIO PID)

Future:

- SCHED_DEADLINE - IRMOS

ftrace

Kernel internal tracer

PREMPT-RT -> Mainline

- function: Function tracer
- function_graph: tracer with stack
- sched_switch: task context switch
- Irqsoff: Interrupt disabled latency

echo on debugfs files

Function tracer

```
# echo function > current_tracer
# cat trace
# tracer: function
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
           xchat-4418  [000] 25100.401772: finish_task_switch <-thread_return
           xchat-4418  [000] 25100.401772: perf_event_task_sched_in <-finish_task_switch
           xchat-4418  [000] 25100.401773: mmdrop <-finish_task_switch
           xchat-4418  [000] 25100.401773: post_schedule <-thread_return
           xchat-4418  [000] 25100.401774: hrtimer_cancel <-schedule_hrtimeout_range
           xchat-4418  [000] 25100.401775: hrtimer_try_to_cancel <-hrtimer_cancel
           xchat-4418  [000] 25100.401775: lock_hrtimer_base <-hrtimer_try_to_cancel
           xchat-4418  [000] 25100.401775: _spin_lock_irqsave <-lock_hrtimer_base
           xchat-4418  [000] 25100.401776: remove_hrtimer <-hrtimer_try_to_cancel
           xchat-4418  [000] 25100.401776: _spin_unlock_irqrestore <-hrtimer_try_to_cancel
           xchat-4418  [000] 25100.401777: fget_light <-do_sys_poll
           xchat-4418  [000] 25100.401779: sock_poll <-do_sys_poll
           xchat-4418  [000] 25100.401780: unix_poll <-sock_poll
           xchat-4418  [000] 25100.401780: sock_poll_wait <-unix_poll
           xchat-4418  [000] 25100.401781: fget_light <-do_sys_poll
           xchat-4418  [000] 25100.401782: pipe_poll <-do_sys_poll
           xchat-4418  [000] 25100.401783: fget_light <-do_sys_poll
           xchat-4418  [000] 25100.401784: sock_poll <-do_sys_poll
           xchat-4418  [000] 25100.401784: unix_poll <-sock_poll
```

Function Graph

```
# echo function_graph > current_tracer
# Cat trace

# CPU  DURATION                  FUNCTION CALLS
# |     |   |                     |   |   |   |
 1)               |  schedule() {
 1)   0.445 us    |    rcu_sched_qs();
 1)   0.432 us    |    _spin_lock_irq();
 1)   0.480 us    |    update_rq_clock();
 1)   0.424 us    |    put_prev_task_idle();
 1)               |    pick_next_task() {
 1)               |      pick_next_task_rt() {
 1)   0.408 us    |        rt_task_of();
 1)   0.483 us    |        dequeue_pushable_task();
 1)   2.464 us    |      }
 1)   3.408 us    |    }
```

Sched_switch

```
# echo sched_switch > current_tracer
# cat trace
# tracer: sched_switch
#
#           TASK-PID    CPU#    TIMESTAMP  FUNCTION
#              | |       |          |         |
         firefox-5117  [000] 24039.607535:   5117:120:S ==> [000]  1576:120:R Xorg
            Xorg-1576  [000] 24039.607664:   1576:120:R   + [000]  1839:120:R metacity
            Xorg-1576  [000] 24039.607806:   1576:120:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.609537:      0:120:R   + [000]  5117:120:R firefox
          <idle>-0     [000] 24039.609541:      0:120:R ==> [000]  5117:120:R firefox
         firefox-5117  [000] 24039.609745:   5117:120:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.612435:      0:120:R ==> [000]  1847:109:R pulseaudio
      pulseaudio-1847  [000] 24039.612563:   1847:109:R   + [001]  5529:120:R firefox
      pulseaudio-1847  [000] 24039.612637:   1847:109:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.612652:      0:120:R ==> [000]  1847:109:R pulseaudio
      pulseaudio-1847  [000] 24039.612663:   1847:109:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.612676:      0:120:R ==> [000]  1847:109:R pulseaudio
      pulseaudio-1847  [000] 24039.612687:   1847:109:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.612703:      0:120:R ==> [000]  5530:120:R firefox
         firefox-5530  [000] 24039.612792:   5530:120:R   + [001]  5529:120:R firefox
         firefox-5530  [000] 24039.612817:   5530:120:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.612841:      0:120:R ==> [000]  1847:109:R pulseaudio
      pulseaudio-1847  [000] 24039.612889:   1847:109:R   + [001]  5529:120:R firefox
      pulseaudio-1847  [000] 24039.612922:   1847:109:R   + [001]  1849: 94:R pulseaudio
      pulseaudio-1847  [000] 24039.612946:   1847:109:S ==> [000]     0:120:R <idle>
          <idle>-0     [000] 24039.613248:      0:120:R   + [000]  5120:120:R firefox
          <idle>-0     [000] 24039.613255:      0:120:R ==> [000]  5120:120:R firefox
         firefox-5120  [000] 24039.613295:   5120:120:R   + [001]  5117:120:R firefox
```

trace-cmd

Binary tool to ftrace functions

Output on binary file

Can send trace info to network

Ex:

```
[root@kiron trace-cmd]# ./trace-cmd record -e sched
disable all
enable sched
path = /sys/kernel/debug/tracing/events/sched/enable
path = /sys/kernel/debug/tracing/events/*/sched/enable
Hit Ctrl^C to stop recording
```

kernelshark

kernelshark(trace.dat)

File Filter Plots Capture Help

Pointer: 101588.389137 Cursor: 0.0 MarkerA: 0.0 MarkerB: 0.0 A,B Delta: 0.0

101588.450758 101588.666431

CPU 0

CPU 1

CPU 2

CPU 3

sched_stat_runtime
d.h..
task: firefox:14173 runtime: 990423 [ns], vruntime: 2264089857180 [ns]
101588.388898 firefox-14173

Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
0	3	101588.023029	gnome-terminal	2580	d....	sched_wakeup	task simpress.bin:9120 [120] success=1 [00
1	0	101588.023080	<idle>	0	d....	sched_switch	task swapper:0 [120] (R) ==> simpress.bin:
2	0	101588.023168	simpress.bin	9120	d....	sched_stat_runtime	task: simpress.bin:9120 runtime: 92765 [ns
3	0	101588.023174	simpress.bin	9120	d....	sched_switch	task simpress.bin:9120 [120] (S) ==> swapp
4	0	101588.032241	<idle>	0	d.h..	sched_wakeup	task simpress.bin:9120 [120] success=1 [00
5	0	101588.032248	<idle>	0	d....	sched_switch	task swapper:0 [120] (R) ==> simpress.bin:
6	0	101588.032328	simpress.bin	9120	d....	sched_stat_runtime	task: simpress.bin:9120 runtime: 81630 [ns
7	0	101588.032334	simpress.bin	9120	d....	sched_switch	task simpress.bin:9120 [120] (S) ==> swapp
8	0	101588.041357	Xorg	1757	d.h.0	sched_wakeup	task simpress.bin:9120 [120] success=1 [00

Tips... Tips... Tips...

New *vs* old kernel

- Old does not mean stable
- Port vendor kernel to mainline
- Features: UBI/LogFS, KSM, LZMA Compression

Use -f 99 /bin/sh to debug

Be aware of use -f 99 /bin/sh

uClibc linuxthreads / nptl

Network scheduler (htb, cbq, hfsc)

Tips... Tips... Tips...

Cortex A9, Atom... Multi-core is available!

- Sched partitioning/affinity
- IRQ balance -> Off

File System

- Writeback/cache
- Memory Filesystem - Memory lock
- Dynamic libraries loading

Thanks!

Questions?

bristot@oftc / (#linux-rt|#kernel-br)

intelbras