

Ftrace: Now and Then

Steven Rostedt

srostedt@redhat.com

rostedt@goodmis.org

[http://people.redhat.com/srostedt/
trace-cmd-linuxcon-2010.odp](http://people.redhat.com/srostedt/trace-cmd-linuxcon-2010.odp)

Who am I?



Ftrace Review

- Internal kernel tracing
 - Function tracing
 - Event tracing
 - Latency tracing
 - Stack 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
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
```

```
#
```

	TASK-PID	CPU#	TIMESTAMP	FUNCTION
	simpres	bin-2792	[000]	634.280032: unix_poll <-sock_poll
	simpres	bin-2792	[000]	634.280033: sock_poll_wait <-unix_poll
	simpres	bin-2792	[000]	634.280033: fput <-do_sys_poll
	simpres	bin-2792	[000]	634.280034: fget_light <-do_sys_poll
	simpres	bin-2792	[000]	634.280035: sock_poll <-do_sys_poll
	simpres	bin-2792	[000]	634.280035: unix_poll <-sock_poll
	simpres	bin-2792	[000]	634.280036: sock_poll_wait <-unix_poll
	simpres	bin-2792	[000]	634.280037: fput <-do_sys_poll
	simpres	bin-2792	[000]	634.280038: fget_light <-do_sys_poll
	simpres	bin-2792	[000]	634.280038: sock_poll <-do_sys_poll
	simpres	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_hrttimeout_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_hrttimeout_range
gnome-terminal-2112 [001]  883.660053:  schedule <-schedule_hrttimeout_range
Xorg-1849  [001]  883.660281:  schedule <-schedule_hrttimeout_range
Xorg-1849  [001]  883.660293:  schedule <-schedule_hrttimeout_range
gnome-terminal-2112 [001]  883.660409:  schedule <-schedule_hrttimeout_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)      |      |      |      |      |
1) 0.487 us |      |      |      |      |
1) 0.409 us |      |      |      |      |
1) 2.519 us |      |      |      |      |
1) 0.420 us |      |      |      |      |
1) 0.415 us |      |      |      |      |
1) 0.415 us |      |      |      |      |
1)      |      |      |      |      |
1) 0.421 us |      |      |      |      |
1) 0.409 us |      |      |      |      |
1)      |      |      |      |      |
1)      |      |      |      |      |
1)      |      |      |      |      |
1) 0.571 us |      |      |      |      |
1) 1.630 us |      |      |      |      |
1)      |      |      |      |      |
```

#	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)			}
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)          |                      avc_policy_seqno();
1)    0.781 us     |                      }
1)    2.435 us     |                  }
1)    4.046 us     |              }
1)    5.675 us     |          }
1)          |      tty_read() {
```

sched_switch

- Soon to be obsolete (tell you why later)

```
[root@frodo tracing]# echo sched_switch > current_tracer
```

```
[root@frodo tracing]# cat trace | head -20
```

```
# tracer: sched_switch
```

```
#
```

```
#          TASK-PID      CPU#    TIMESTAMP  FUNCTION
```

```
#          | |          |          |          |
```

	bash-3002	[000]	484.062117:	3002:120:S	+	[000]	3002:120:S	bash
	bash-3002	[000]	484.062138:	3002:120:S	+	[000]	3002:120:S	bash
	bash-3002	[000]	484.062258:	3002:120:S	+	[000]	3002:120:S	bash
	bash-3002	[000]	484.062267:	3002:120:S	==>	[000]	0:140:R	<idle>
	<idle>-0	[000]	484.062770:	0:140:R	+	[000]	9:120:R	events/0
	<idle>-0	[000]	484.062775:	0:140:R	==>	[000]	9:120:R	events/0
	events/0-9	[000]	484.062780:	9:120:R	+	[001]	2112:120:R	gnome-ter
	events/0-9	[000]	484.062791:	9:120:S	==>	[000]	0:140:R	<idle>
	Xorg-1849	[001]	484.063096:	1849:120:R	==>	[001]	2112:120:R	gnome-ter
gnome-terminal-2112		[001]	484.063155:	2112:120:S	==>	[001]	1849:120:R	Xorg
	Xorg-1849	[001]	484.063552:	1849:120:R	+	[001]	2112:120:R	gnome-ter
	Xorg-1849	[001]	484.063557:	1849:120:R	==>	[001]	2112:120:R	gnome-ter
gnome-terminal-2112		[001]	484.063564:	2112:120:S	==>	[001]	1849:120:R	Xorg
	Xorg-1849	[001]	484.064117:	1849:120:R	+	[001]	2112:120:R	gnome-ter
	Xorg-1849	[001]	484.064122:	1849:120:R	==>	[001]	2112:120:R	gnome-ter
gnome-terminal-2112		[001]	484.064130:	2112:120:S	==>	[001]	1849:120:R	Xorg

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

Latency Tracers

```
[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#
#           /_-----=> irqsoff
#           | /_-----=> 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
[...]
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
```

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
```

```
#
```

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
#           | |          |          |          |
ptr=(null)  acpid-1470    [001]    794.947181:    kfree: call_site=ffffffff810c996d
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]
```

Enable Multiple Events

- Looks a lot like sched_switch plugin

```
[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]
```

Event Directory or File

- `set_event` shows all events enabled
- `available_events` shows what events are available
- `echo 1 > events/sched/enable`
 - same as “`echo sched > set_event`”

```
[root@frodo tracing]# echo 1 > events/irq/enable
[root@frodo tracing]# cat set_event
irq:irq_handler_entry
irq:irq_handler_exit
irq:softirq_entry
irq:softirq_exit
```

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 (the plugins will show the event output)

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	

stacktrace

```
      <idle>-0      [001]    973.876728: kmem_cache_free:
call_site=ffffffff81103e40 ptr=ffff88002f4aa240
      <idle>-0      [001]    973.876728: <stack trace>
=> kmem_cache_free
=> mempool_free_slab
=> mempool_free
=> bio_free
=> dm_bio_destructor
=> bio_put
=> clone_endio
=> bio_endio
```


Filters

- Filter any trace event or ftrace entry
- Use equal '==' and logic descriptors '||' and '&&'
- Filter on any field in the format file
 - i.e. sched_switch's prev_state

Filter on sched_switch

```
[root@frodo tracing]# echo "prev_state == 0" > events/sched/sched_switch/filter
[root@frodo tracing]# cat trace | head -15
# tracer: nop
#
#          TASK-PID    CPU#    TIMESTAMP    FUNCTION
#          | |          |          |          |
<idle>-0    [001]    1408.013962: sched_switch: task swapper:0 [140] (R) ==> events/1:10 [120]
<idle>-0    [001]    1408.015127: sched_switch: task swapper:0 [140] (R) ==> Xorg:1840 [120]
Xorg-1840    [001]    1408.015222: sched_switch: task Xorg:1840 [120] (R) ==> gnome-settings-:2133
[120]
Xorg-1840    [001]    1408.015625: sched_switch: task Xorg:1840 [120] (R) ==> metacity:2139 [120]
Xorg-1840    [001]    1408.015709: sched_switch: task Xorg:1840 [120] (R) ==> wnck-applet:2220 [120]
Xorg-1840    [001]    1408.015913: sched_switch: task Xorg:1840 [120] (R) ==> wnck-applet:2220 [120]
Xorg-1840    [001]    1408.015940: sched_switch: task Xorg:1840 [120] (R) ==> wnck-applet:2220 [120]
Xorg-1840    [001]    1408.016029: sched_switch: task Xorg:1840 [120] (R) ==> metacity:2139 [120]
Xorg-1840    [001]    1408.016057: sched_switch: task Xorg:1840 [120] (R) ==> metacity:2139 [120]
Xorg-1840    [001]    1408.016181: sched_switch: task Xorg:1840 [120] (R) ==> wnck-applet:2220 [120]
Xorg-1840    [001]    1408.016205: sched_switch: task Xorg:1840 [120] (R) ==> wnck-applet:2220 [120]
```

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 1 > tracing_on; run_test; echo 0 > tracing_on
```

stack_trace

- echo 1 >
/proc/sys/kernel/stack_tracer_enabled
- kernel command line “stacktrace”

stack_trace

```
[root@frodo tracing]# cat stack_trace
```

	Depth	Size	Location (45 entries)
	-----	-----	-----
0)	4048	112	ftrace_call+0x5/0x2b
1)	3936	64	update_curr+0x10a/0x12b
2)	3872	64	enqueue_entity+0x31/0x20f
3)	3808	48	enqueue_task_fair+0x3d/0x98
4)	3760	48	enqueue_task+0x6b/0x8d
[...]			
28)	1936	96	sr_test_unit_ready+0x72/0xec
29)	1840	144	sr_media_change+0x57/0x264
30)	1696	64	media_changed+0x63/0xb2
31)	1632	32	cdrom_media_changed+0x44/0x5e
32)	1600	32	sr_block_media_changed+0x2c/0x42
33)	1568	48	check_disk_change+0x3c/0x85
34)	1520	512	cdrom_open+0x8d9/0x96b
35)	1008	80	sr_block_open+0x9f/0xd2
36)	928	112	__blkdev_get+0xde/0x37c
37)	816	32	blkdev_get+0x23/0x39
38)	784	64	blkdev_open+0x85/0xd1
39)	720	96	__dentry_open+0x14b/0x28f
40)	624	48	nameidata_to_filp+0x51/0x76
41)	576	320	do_filp_open+0x514/0x9bc
42)	256	96	do_sys_open+0x71/0x131
43)	160	32	sys_open+0x33/0x49
44)	128	128	system_call_fastpath+0x16/0x1b

trace-cmd

- Version 1.1-rc1

[git://git.kernel.org/pub/scm/linux/kernel/git/rostedt/trace-cmd.git](https://git.kernel.org/pub/scm/linux/kernel/git/rostedt/trace-cmd.git)

trace-cmd

- binary tool to read Ftrace's buffers
 - Records into a trace.dat file for later reads
 - Reads the trace.dat file
 - Can record on big endian, read in little, and vice versa
 - Reads the raw buffers using splice
 - Will automatically mount debugfs if it is not mounted
 - Must have root access (sudo)

trace-cmd record

- Default, writes to “trace.dat”

```
[root@frodo ~]# trace-cmd record -e sched ls -ltr /usr > /dev/null
disable all
enable sched
offset=2f2000
offset=2f4000
```

trace-cmd record

- Default, writes to “trace.dat”

```
[root@frodo ~]# trace-cmd record -e sched ls -ltr /usr > /dev/null
disable all
enable sched
offset=2f2000
offset=2f4000
[root@frodo ~]# trace-cmd record -o func.dat -p function ls -ltr /usr > /dev/null
plugin function
disable all
offset=2f2000
offset=412000
```

trace-cmd record

- Default, writes to “trace.dat”

```
[root@frodo ~]# trace-cmd record -e sched ls -ltr /usr > /dev/null
disable all
enable sched
offset=2f2000
offset=2f4000
[root@frodo ~]# trace-cmd record -o func.dat -p function ls -ltr /usr > /dev/null
plugin function
disable all
offset=2f2000
offset=412000
[root@frodo ~]# trace-cmd record -o fgraph.dat -p function_graph ls -ltr /usr \
> /dev/null
plugin function_graph
disable all
offset=2f2000
offset=460000
```

trace-cmd record

- Default, writes to “trace.dat”

```
[root@frodo ~]# trace-cmd record -e sched ls -ltr /usr > /dev/null
disable all
enable sched
offset=2f2000
offset=2f4000
[root@frodo ~]# trace-cmd record -o func.dat -p function ls -ltr /usr > /dev/null
plugin function
disable all
offset=2f2000
offset=412000
[root@frodo ~]# trace-cmd record -o fgraph.dat -p function_graph ls -ltr /usr \
> /dev/null
plugin function_graph
disable all
offset=2f2000
offset=460000
[root@frodo ~]# trace-cmd record -o fgraph-events.dat -e sched -p function_graph \
ls -ltr /usr > /dev/null
plugin function_graph
disable all
enable sched
offset=2f2000
offset=461000
```

Filters, and Options

```
[root@frodo ~]# trace-cmd record -e sched_switch -f 'prev_prio < 100'
[root@frodo ~]# trace-cmd record -p function_graph -O nograph-time
[root@frodo ~]# trace-cmd record -p function_graph -g sys_read
[root@frodo ~]# trace-cmd record -p function_graph -l do_IRQ -l timer_interrupt
[root@frodo ~]# trace-cmd record -p function_graph -n '*lock*'
```

- -f : filter
- -O : option
- -g : same as echoing into set_graph_function
- -l : same as echoing into set_fttrace_filter
- -n : same as echoing into set_fttrace_notrace

trace-cmd report

- Default, reads from “trace.dat”

```
[root@frodo ~]# trace-cmd report | head -15
```

```
version = 6
```

```
cpus=2
```

trace-cmd-6157	[000]	83.713584:	sched_stat_runtime:	task: trace-cmd:61
trace-cmd-6157	[000]	83.713591:	sched_switch:	6157:120:S ==> 0:1
<idle>-0	[000]	83.713646:	sched_stat_wait:	task: trace-cmd:61
<idle>-0	[000]	83.713648:	sched_switch:	0:120:R ==> 6158:1
ls-6158	[001]	83.713934:	sched_wakeup:	6158:?:? + 5900:
ls-6158	[001]	83.713935:	sched_stat_runtime:	task: trace-cmd:61
ls-6158	[001]	83.713937:	sched_stat_runtime:	task: trace-cmd:61
ls-6158	[001]	83.713938:	sched_switch:	6158:120:R ==> 590
migration/1-5900	[001]	83.713941:	sched_stat_wait:	task: trace-cmd:61
migration/1-5900	[001]	83.713942:	sched_migrate_task:	task trace-cmd:615
migration/1-5900	[001]	83.713947:	sched_switch:	5900:0:S ==> 0:120
ls-6158	[000]	83.714067:	sched_stat_runtime:	task: ls:6158 runt
ls-6158	[000]	83.714636:	sched_stat_runtime:	task: ls:6158 runt

trace-cmd report (continued)

```
[root@frodo ~]# trace-cmd report -i func.dat | head -15
```

```
version = 6
```

```
cpus=2
```

ls-6178	[000]	137.259033:	function:	fsnotify_modify <-- vfs_write
ls-6178	[000]	137.259035:	function:	inotify_inode_queue_event <-- fsn
ls-6178	[000]	137.259035:	function:	fsnotify_parent <-- fsnotify_modi
ls-6178	[000]	137.259035:	function:	__fsnotify_parent <-- fsnotify_pa
ls-6178	[000]	137.259036:	function:	inotify_dentry_parent_queue_event
ls-6178	[000]	137.259036:	function:	fsnotify <-- fsnotify_modify
ls-6178	[000]	137.259036:	function:	fput_light <-- sys_write
ls-6178	[000]	137.259037:	function:	audit_syscall_exit <-- sysret_aud
ls-6178	[000]	137.259037:	function:	audit_get_context <-- audit_sysca
ls-6178	[000]	137.259037:	function:	audit_free_names <-- audit_syscal
ls-6178	[000]	137.259038:	function:	path_put <-- audit_free_names
ls-6178	[000]	137.259038:	function:	dput <-- path_put
ls-6178	[000]	137.259038:	function:	mntput <-- path_put

trace-cmd report (continued)

```
[root@frodo ~]# trace-cmd report -i fgraph.dat | head -15 | cut -c32-43 --complement
version = 6
cpus=2
ls-6186 [000] funcgraph_entry: | fsnotify_modify() {
ls-6186 [000] funcgraph_entry: 0.709 us |   inotify_inode_queue_event();
ls-6186 [000] funcgraph_entry: |   fsnotify_parent() {
ls-6186 [000] funcgraph_entry: 0.397 us |     __fsnotify_parent();
ls-6186 [000] funcgraph_entry: 0.385 us |     inotify_dentry_parent_queu
ls-6186 [000] funcgraph_exit: 1.942 us |   }
ls-6186 [000] funcgraph_entry: 0.390 us |   fsnotify();
ls-6186 [000] funcgraph_exit: 7.064 us | }
ls-6186 [000] funcgraph_entry: 0.403 us | fput_light();
ls-6186 [000] funcgraph_entry: | audit_syscall_exit() {
ls-6186 [000] funcgraph_entry: 0.396 us |   audit_get_context();
ls-6186 [000] funcgraph_entry: |   audit_free_names() {
ls-6186 [000] funcgraph_entry: |     path_put() {
```

trace-cmd report (continued)

```
[root@frodo ~]# trace-cmd report -i fgraph-events.dat | head -15 | \
  cut -c32-43  --complement
version = 6
cpus=2
ls-6209 [001] funcgraph_entry:0.385 us | task_of();
ls-6209 [001] funcgraph_entry: | ftrace_raw
ls-6209 [001] sched_stat_wait: task: phy0:861 wait: 56445 [ns]
ls-6209 [001] funcgraph_exit: 0.613 us | }
ls-6209 [001] funcgraph_exit: 2.291 us | }
ls-6209 [001] funcgraph_entry: 0.439 us | __dequeue_entity(
ls-6209 [001] funcgraph_exit: 4.064 us | }
ls-6209 [001] funcgraph_entry: 0.379 us | task_of();
ls-6209 [001] funcgraph_entry: 0.373 us | hrtick_start_fair
ls-6209 [001] funcgraph_exit: 6.415 us | }
ls-6209 [001] funcgraph_exit: 7.202 us | }
ls-6209 [001] funcgraph_entry: 0.396 us | perf_event_task_sch
ls-6209 [001] funcgraph_entry: | ftrace_raw_event_sc
ls-6209 [001] sched_switch: 6209:120:R ==> 861:120: phy0
ls-6209 [001] funcgraph_exit: 0.637 us | }
```

trace-cmd start

- Using start is like echoing into debugfs
 - `trace-cmd start -e all`
 - same as “`echo 1 > events/enable`”
- Uses the same options as `trace-cmd record`
 - `trace-cmd start -p function_graph`
 - `trace-cmd start -p function -e sched_switch`

trace-cmd stop / extract

- trace-cmd stop
 - stops the tracer from writing:
 - same as “echo 0 > tracing_on”
- trace-cmd extract -o output.dat
 - Makes a “dat” file that trace-cmd report can use
 - Without “-o ...” will create “trace.dat”

trace-cmd reset

- trace-cmd stop does not stop the overhead of tracing
- trace-cmd reset disables all tracing
 - trace-cmd reset
- Removes trace data from kernel
 - Do the extract before doing the reset

trace-cmd list

- See the trace options, events or plugins
 - trace-cmd list -o
 - shows list of trace options
 - these options are used by trace-cmd record -O option
 - trace-cmd list -p
 - available plugins
 - trace-cmd list -e
 - available events

trace-cmd split

- Split by time, events, CPU
 - `trace-cmd split 258.121328`
 - splits from timestamp to end of file
 - `trace-cmd split -e 1000`
 - splits out the first 1000 events
 - `trace-cmd split -m 1 -r 258.121328 259.000000`
 - split 1 millisecond starting at first timestamp to second timestamp repeatedly
 - `trace.dat.1`, `trace.dat.2`, ...

trace-cmd listen

- listen for connections from other boxes
 - `trace-cmd listen -p 5678 -d`
- Record can now send to that box
 - `trace-cmd record -N host:5678 -e all`
 - use “-t” to force TCP otherwise trace data is sent via UDP

A cute little trick

- Finding high latency interrupts

```
[root@frodo ~]# trace-cmd record -p function_graph -l do_IRQ \  
-e irq_handler_entry  
plugin function_graph  
disable all  
enable irq_handler_entry  
path = /debug/tracing/events/irq_handler_entry/enable  
path = /debug/tracing/events/*/irq_handler_entry/enable  
Hit Ctrl^C to stop recording
```

A cute little trick

- Finding high latency interrupts

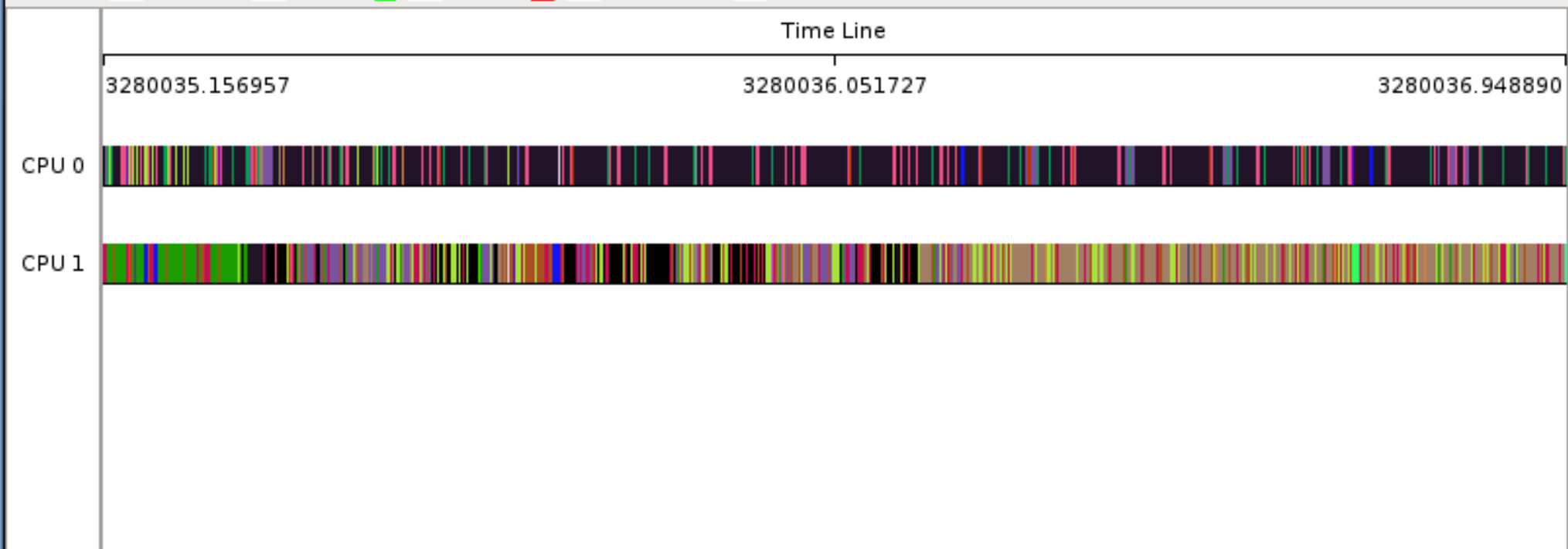
```
[root@frodo ~]# trace-cmd report | cut -c32-43 --complement
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=0 handler=timer
<idle>-0      [000]   funcgraph_exit:        + 29.013 us | }
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=30 handler=iwl3945
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=30 handler=iwl3945
<idle>-0      [000]   funcgraph_exit:        + 22.580 us | }
<idle>-0      [000]   funcgraph_exit:        ! 175.404 us | }
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=0 handler=timer
<idle>-0      [000]   funcgraph_exit:        + 27.239 us | }
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=0 handler=timer
<idle>-0      [000]   funcgraph_exit:        + 28.537 us | }
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=0 handler=timer
<idle>-0      [000]   funcgraph_exit:        + 29.157 us | }
<idle>-0      [000]   funcgraph_entry:                | do_IRQ() {
<idle>-0      [000]   irq_handler_entry:      irq=0 handler=timer
<idle>-0      [000]   funcgraph_exit:        + 21.522 us | }
```

KernelShark

- A front end reader of the trace-cmd trace.dat file
- Graph view
- List view
- Simple and Advance filtering
- Still in Beta (for now)

File Filter Plots Help

Pointer: 0.0 Cursor: 0.0 Marker A: 0.0 Marker B: 0.0 A,B Delta: 0.0



Page 1 Search: Column: # contains graph follows

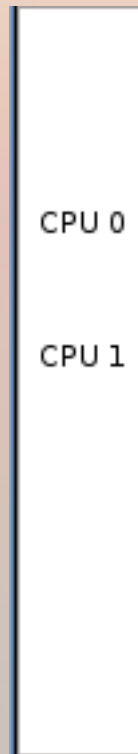
#	CPU	Time Stamp	Task	PID	Latency	Event	Info
0	1	3280035.156957	trace-cmd	25900	sys_exit	NR 42 = 0
1	0	3280035.156958	ls	25901	sys_exit	NR 4 = 1
2	1	3280035.156965	trace-cmd	25900	mm_page_alloc	page=0xffffea00009c3dc8 pfn=10239432 or
3	1	3280035.156971	trace-cmd	25900	sys_enter	NR 162 (ff9aec1c, 0, 806c484, 0, 0, ff9
4	1	3280035.156974	trace-cmd	25900	hrtimer_init	hrtimer 0xffff88007ce35ea8, clockid CLO
5	1	3280035.156980	trace-cmd	25900	d....	hrtimer_start	hrtimer=0xffff88007ce35ea8 function=hrt
6	0	3280035.156991	ls	25901	sys_enter	NR 11 (ff9aec0d, ff9afddc, ff9afdec, ff
7	0	3280035.156994	ls	25901	kmem_cache_alloc	(getname+0x23) call_site=810f559d ptr=0
8	0	3280035.156997	ls	25901	kmem_cache_alloc	(compat_do_execve+0x43) call_site=8111c
9	1	3280035.156997	trace-cmd	25900	d....	sched_switch	25900:44:S ==> 0:140: swapper

Graph Info Area

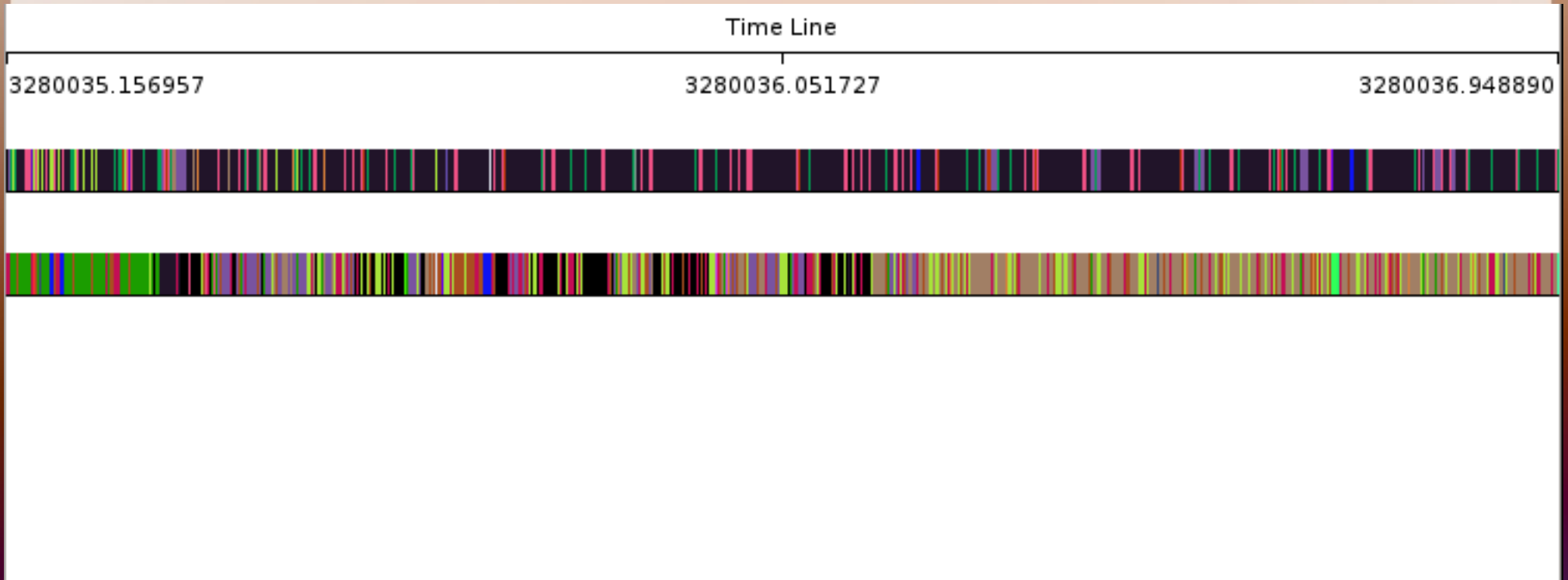
- Time Stamps
 - Pointer – where the mouse is located
 - Cursor – double click
 - Marker A – left mouse click
 - Marker B – left mouse click with shift
 - Delta between A and B

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

Plot Title



Plot Area

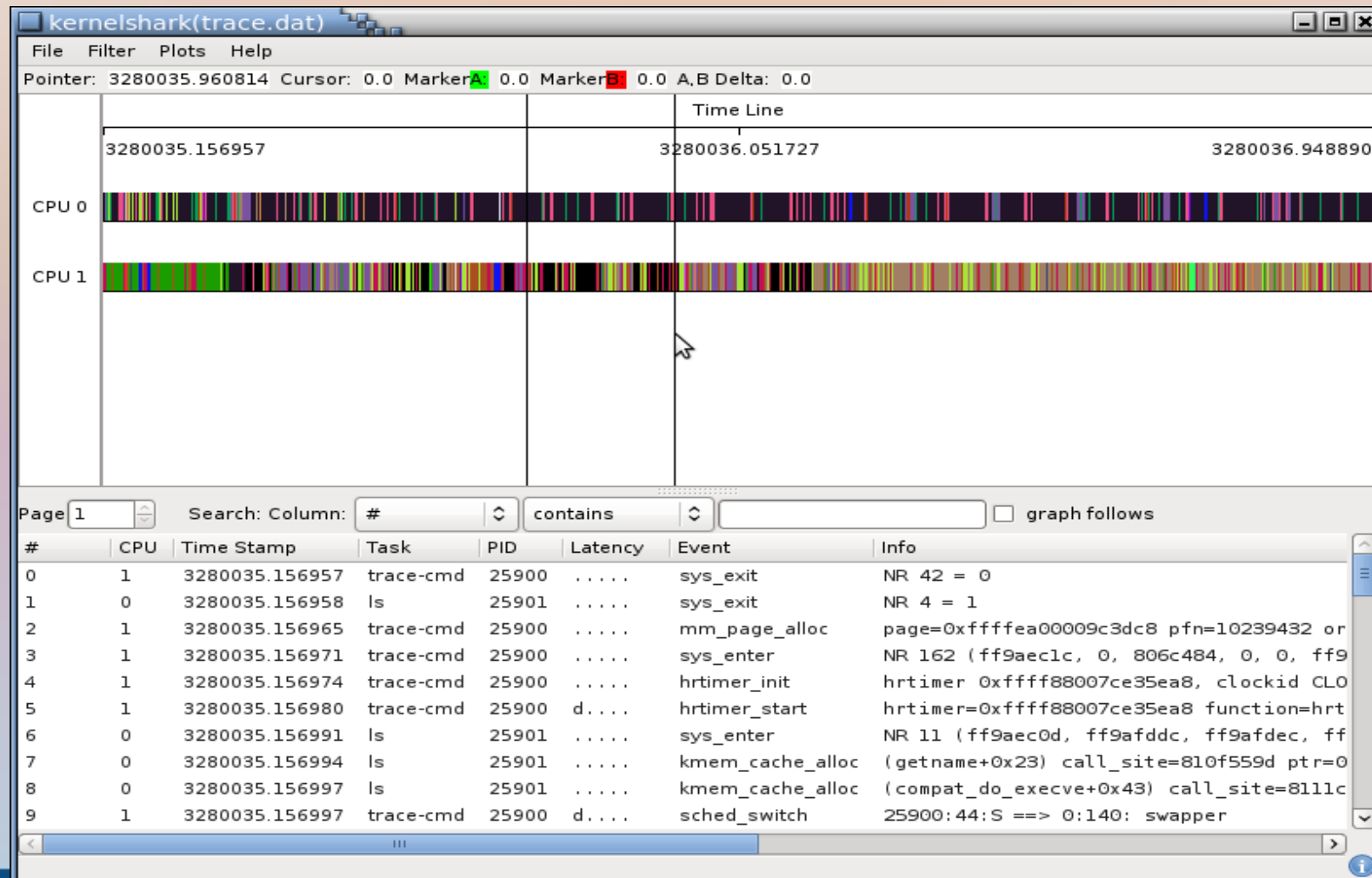


List Area

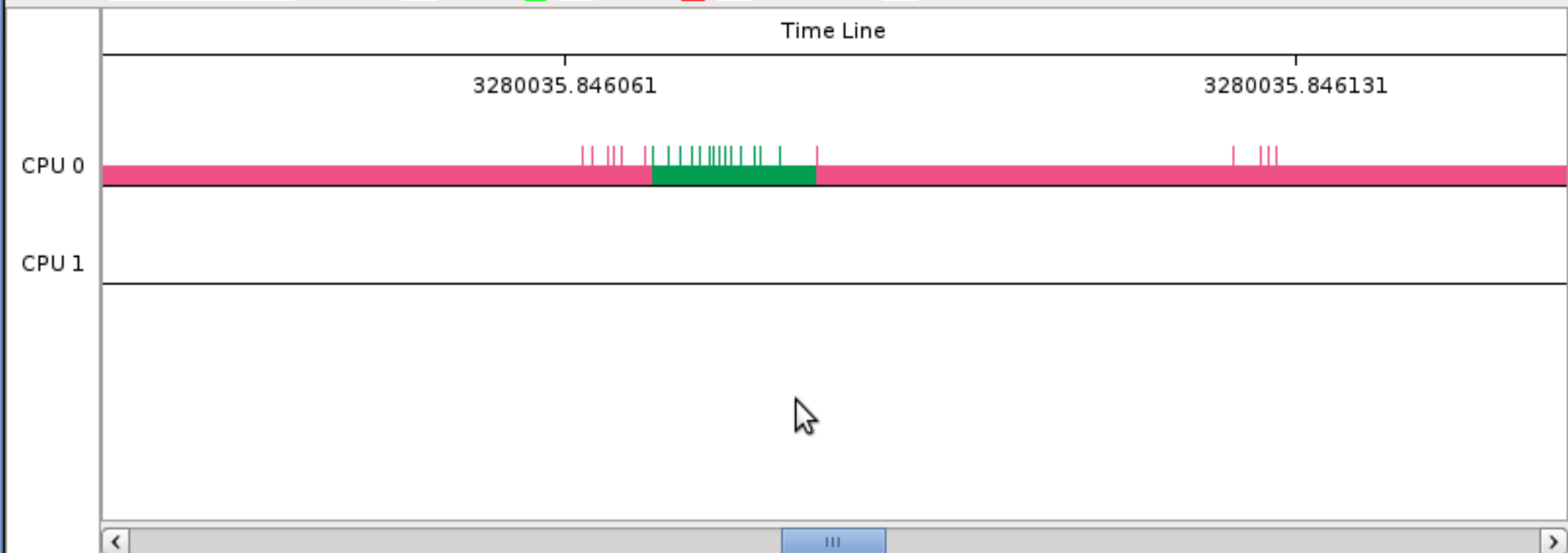
Page 1 Search: Column: # contains ☐ graph follows

Zooming In

- Left click and drag to the right



Pointer: 3280035.846083 Cursor: 0.0 Marker A: 0.0 Marker B: 0.0 A,B Delta: 0.0

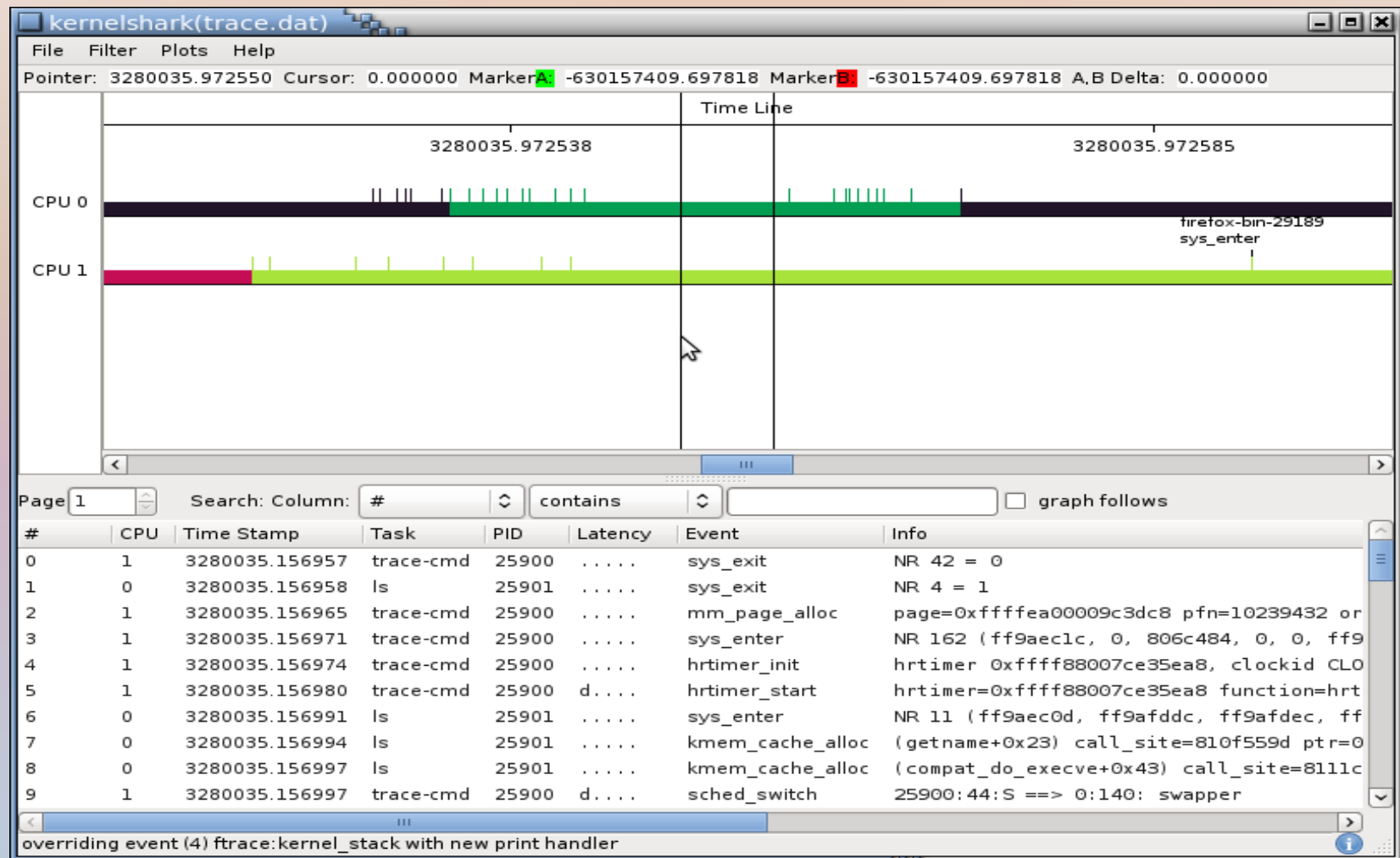


Page 1 Search: Column: # contains graph follows

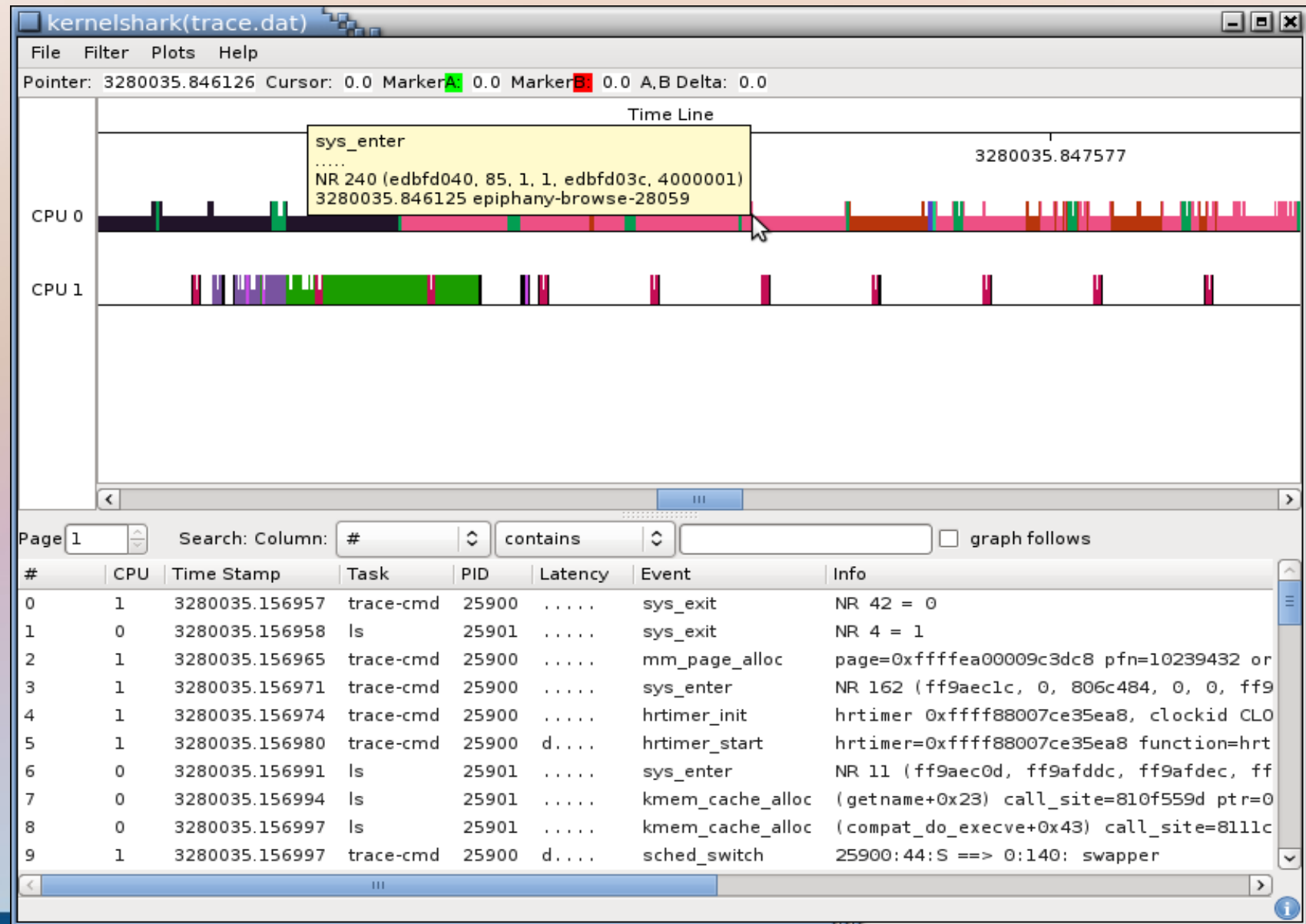
#	CPU	Time Stamp	Task	PID	Latency	Event	Info
0	1	3280035.156957	trace-cmd	25900	sys_exit	NR 42 = 0
1	0	3280035.156958	ls	25901	sys_exit	NR 4 = 1
2	1	3280035.156965	trace-cmd	25900	mm_page_alloc	page=0xffffea00009c3dc8 pfn=10239432 or
3	1	3280035.156971	trace-cmd	25900	sys_enter	NR 162 (ff9aec1c, 0, 806c484, 0, 0, ff9
4	1	3280035.156974	trace-cmd	25900	hrtimer_init	hrtimer 0xffff88007ce35ea8, clockid CLO
5	1	3280035.156980	trace-cmd	25900	d....	hrtimer_start	hrtimer=0xffff88007ce35ea8 function=hrt
6	0	3280035.156991	ls	25901	sys_enter	NR 11 (ff9aec0d, ff9afddc, ff9afdec, ff
7	0	3280035.156994	ls	25901	kmem_cache_alloc	(getname+0x23) call_site=810f559d ptr=0
8	0	3280035.156997	ls	25901	kmem_cache_alloc	(compat_do_execve+0x43) call_site=8111c
9	1	3280035.156997	trace-cmd	25900	d....	sched_switch	25900:44:S ==> 0:140: swapper

Zoom Out

- Left click and drag left



Event Info Tool Tip

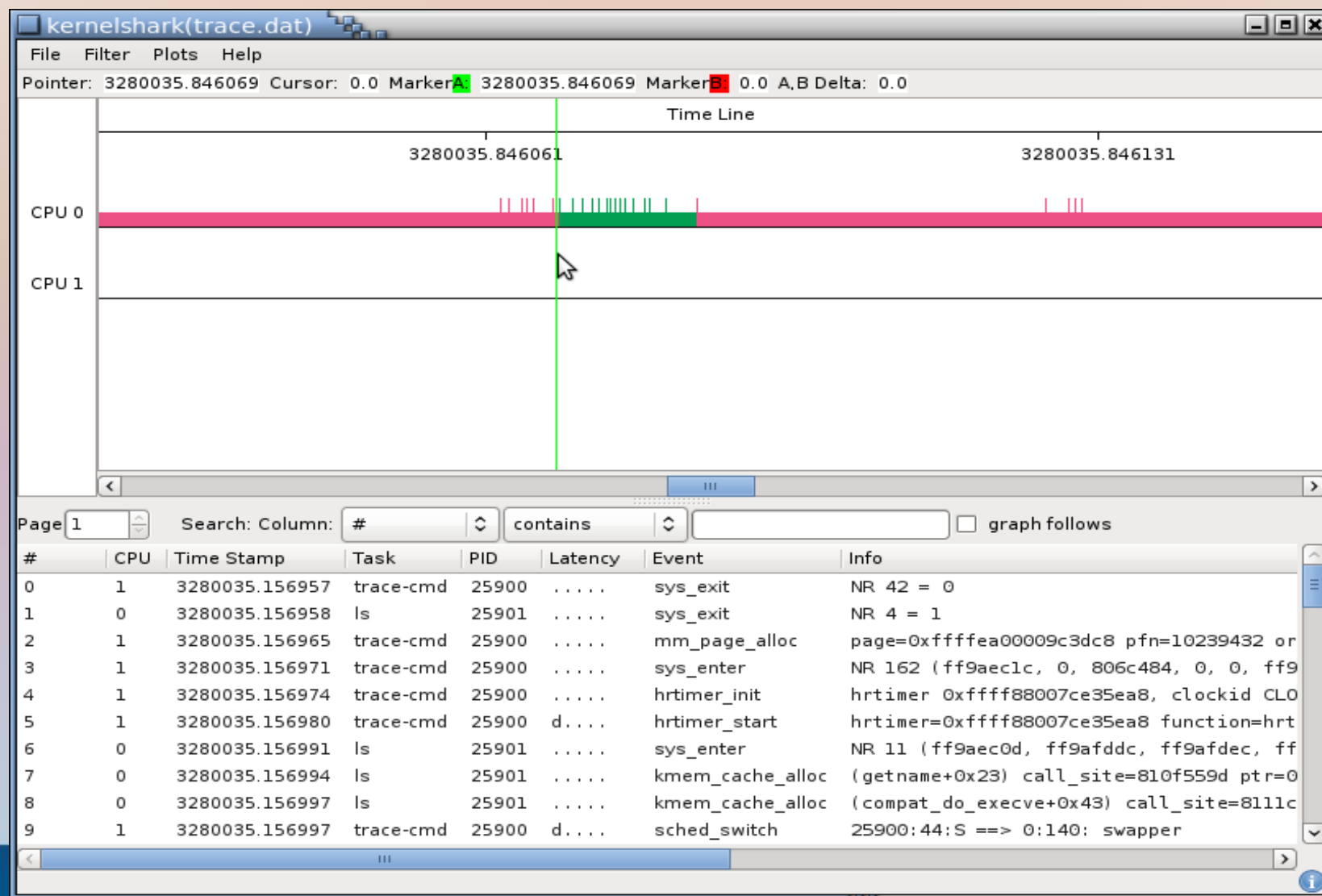


Graph Markers

- Marker A and B
- Used to calculate the deltas

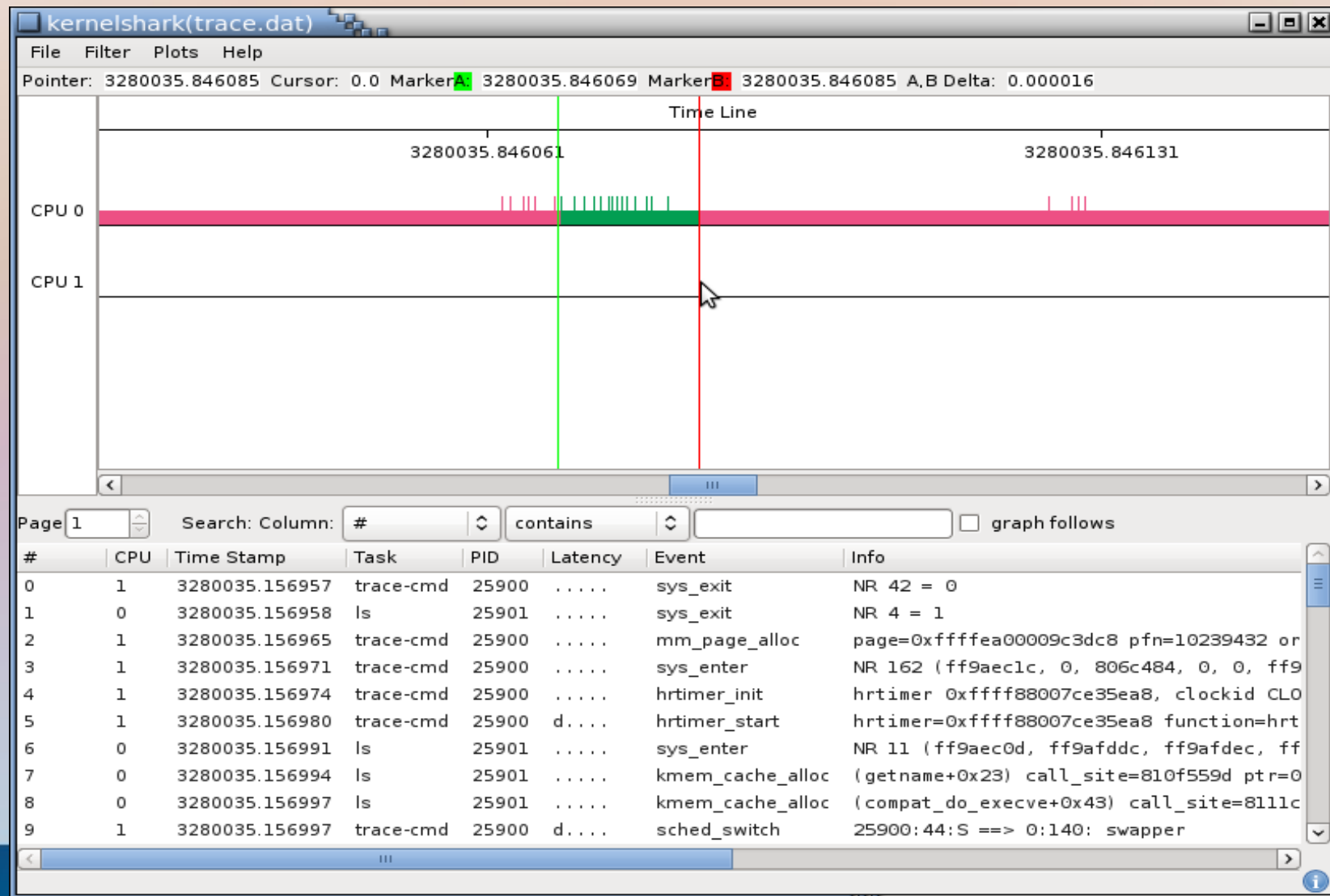
Marker A

- Left mouse click



Marker B

- Left mouse click with shift key held

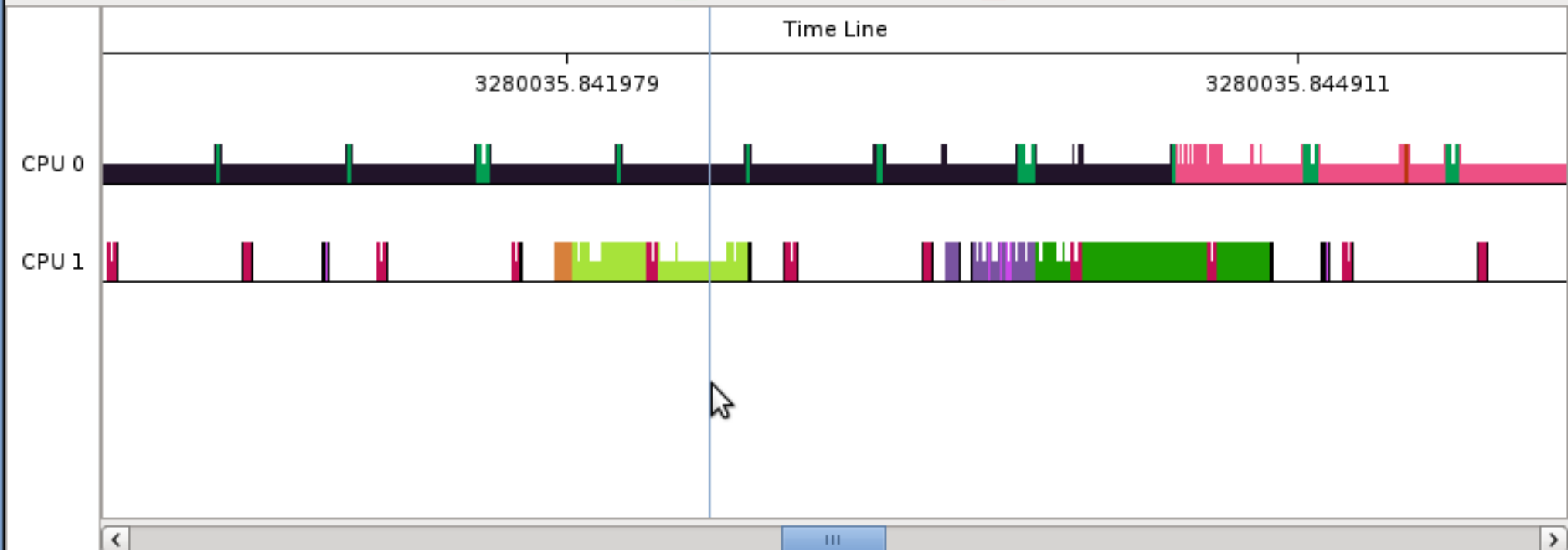


Graph Cursor

- Double click on graph
- Moves the list view to the closest event to the timestamp on where the cursor is.
- Can be used for marking location on zooming in and out

File Filter Plots Help

Pointer: 3280035.842559 Cursor: 3280035.842559 Marker A: 3280035.842559 Marker B: 0.0 A,B Delta: 0.0



Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
217439	1	3280035.842339	trace-cmd	25900	mm_page_free_direct	page=0xffffea0001abb6e0 pfn=28030688
217440	1	3280035.842340	trace-cmd	25900	mm_page_free_direct	page=0xffffea0001abb6e0 pfn=28030688
217441	1	3280035.842341	trace-cmd	25900	kfree	(buffer_pipe_buf_release+0x2d) call_s
217442	1	3280035.842341	trace-cmd	25900	sys_exit	NR 313 = 4096
217443	1	3280035.842342	trace-cmd	25900	sys_enter	NR 162 (ff9aec1c, 0, 806c484, 0, 0, 1
217444	1	3280035.842343	trace-cmd	25900	hrtimer_init	hrtimer 0xfffff88007ce35ea8, clockid (
217445	1	3280035.842345	trace-cmd	25900	d....	hrtimer_start	hrtimer=0xfffff88007ce35ea8 function=f
217446	1	3280035.842349	trace-cmd	25900	d....	sched_switch	25900:44:S ==> 29189:120: firefox-bir
217447	1	3280035.842418	firefox-bin	29189	sys_enter	NR 78 (ff9b8ea8, 0, f6595a9c, f4240,
217448	1	3280035.842421	firefox-bin	29189	sys_exit	NR 78 = 0

Graph Plots

- CPU Plots
 - colors change depending on what task is running
- Task Plots
 - colors change depending on what CPU the task is on
 - shows wake up latency (hollow green box)
 - shows preempt latency (hollow red box)
 - can also be opened by menu option when mouse is over a task in the CPU plot

File Filter Plots Help

Pointer: 328 CPUs Cursor: 3280035.846078 Marker A: 3280035.846069 Marker B: 3280035.846085 A,B Delta: 0.000015

Tasks

Time Line

3280035.156957

3280036.051727

3280036.948890

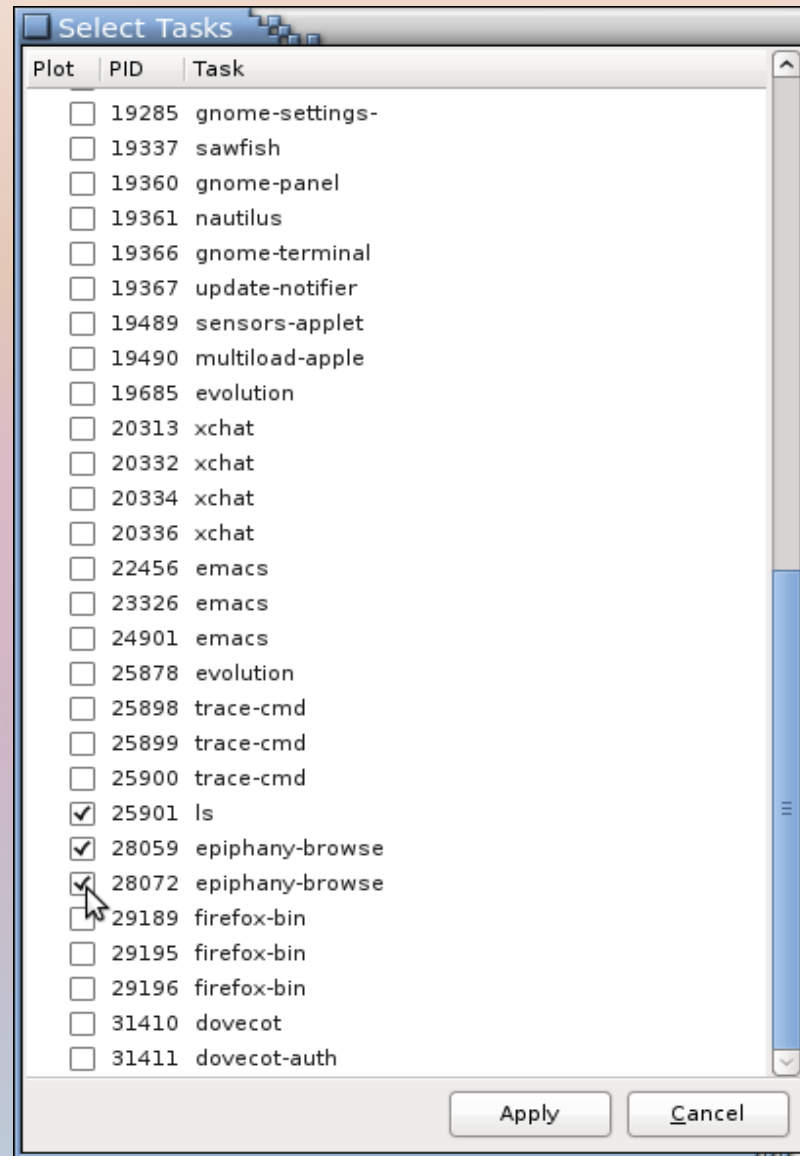
CPU 0

CPU 1

Page 1 Search: Column: # contains graph follows

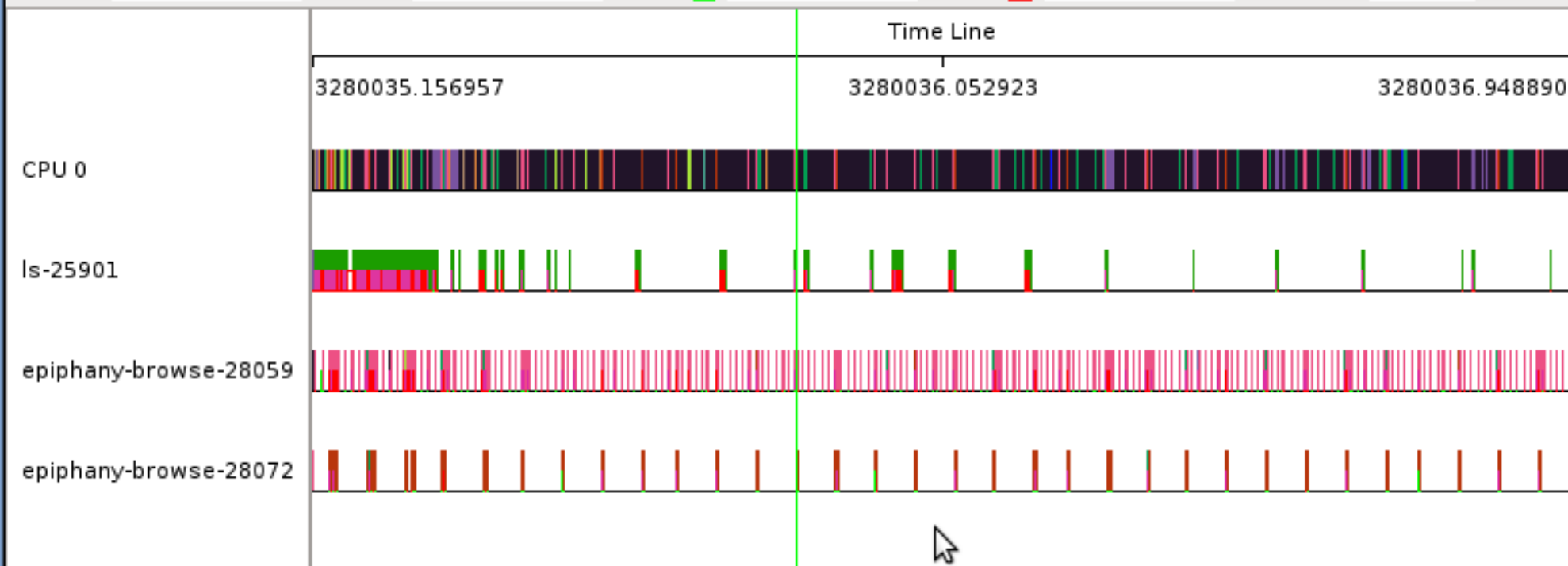
#	CPU	Time Stamp	Task	PID	Latency	Event	Info
218483	0	3280035.846071	trace-cmd	25899	sys_exit	NR 162 = 0
218484	0	3280035.846072	trace-cmd	25899	sys_enter	NR 313 (5, 0, 7, 0, 1000, 1)
218485	0	3280035.846073	trace-cmd	25899	kmalloc	(tracing_buffers_splice_read+0x121) c
218486	0	3280035.846074	trace-cmd	25899	mm_page_alloc	page=0xffffea00008a02c0 pfn=9044672 c
218487	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218488	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218489	0	3280035.846076	trace-cmd	25899	kfree	(tracing_buffers_splice_read+0x180) c
218490	0	3280035.846076	trace-cmd	25899	sys_exit	NR 313 = 0
218491	0	3280035.846077	trace-cmd	25899	sys_enter	NR 313 (6, 0, 4, 0, 1000, 3)
218492	0	3280035.846078	trace-cmd	25899	sys_exit	NR 313 = -11

List of Tasks to plot



File Filter Plots Help

Pointer: 3280036.041510 Cursor: 3280035.846078 Marker A: 3280035.846069 Marker B: 3280035.846085 A,B Delta: 0.000015

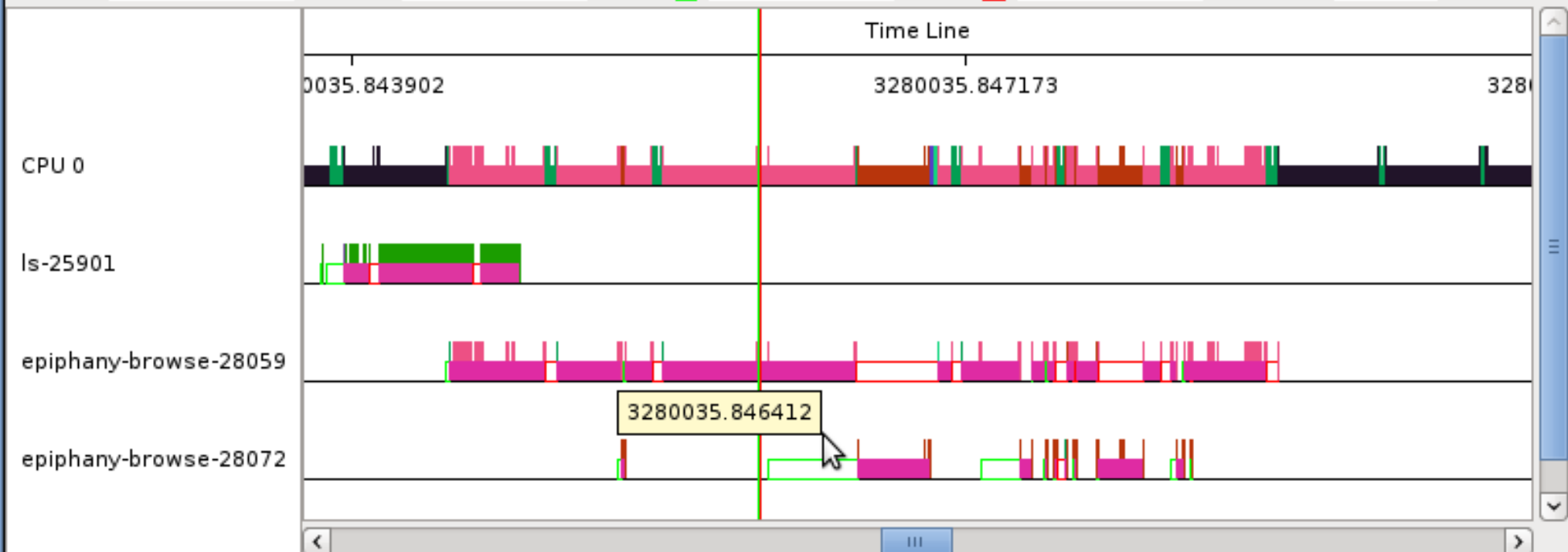


Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
218483	0	3280035.846071	trace-cmd	25899	sys_exit	NR 162 = 0
218484	0	3280035.846072	trace-cmd	25899	sys_enter	NR 313 (5, 0, 7, 0, 1000, 1)
218485	0	3280035.846073	trace-cmd	25899	kmalloc	(tracing_buffers_splice_read+0x121) c
218486	0	3280035.846074	trace-cmd	25899	mm_page_alloc	page=0xffffea00008a02c0 pfn=9044672 c
218487	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218488	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218489	0	3280035.846076	trace-cmd	25899	kfree	(tracing_buffers_splice_read+0x180) c
218490	0	3280035.846076	trace-cmd	25899	sys_exit	NR 313 = 0
218491	0	3280035.846077	trace-cmd	25899	sys_enter	NR 313 (6, 0, 4, 0, 1000, 3)
218492	0	3280035.846078	trace-cmd	25899	sys_exit	NR 313 = -11

File Filter Plots Help

Pointer: 3280035.846412 Cursor: 3280035.846078 Marker A: 3280035.846069 Marker B: 3280035.846085 A,B Delta: 0.000015

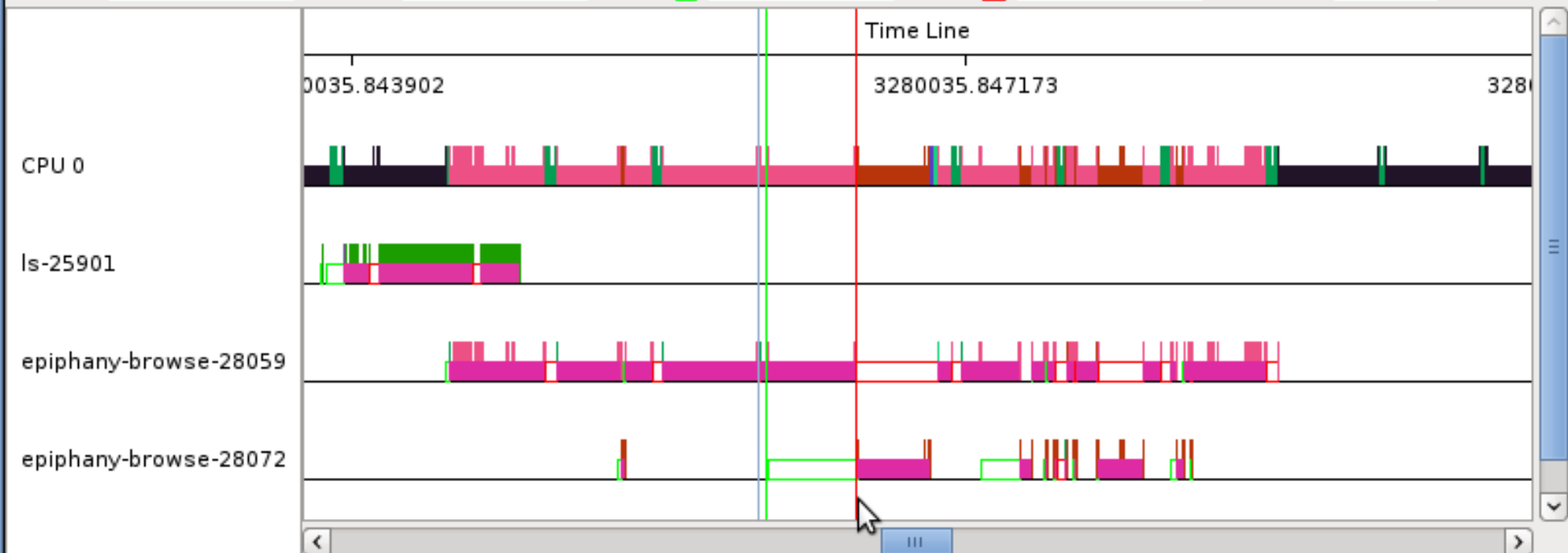


Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
218483	0	3280035.846071	trace-cmd	25899	sys_exit	NR 162 = 0
218484	0	3280035.846072	trace-cmd	25899	sys_enter	NR 313 (5, 0, 7, 0, 1000, 1)
218485	0	3280035.846073	trace-cmd	25899	kmalloc	(tracing_buffers_splice_read+0x121) c
218486	0	3280035.846074	trace-cmd	25899	mm_page_alloc	page=0xffffea00008a02c0 pfn=9044672 c
218487	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218488	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218489	0	3280035.846076	trace-cmd	25899	kfree	(tracing_buffers_splice_read+0x180) c
218490	0	3280035.846076	trace-cmd	25899	sys_exit	NR 313 = 0
218491	0	3280035.846077	trace-cmd	25899	sys_enter	NR 313 (6, 0, 4, 0, 1000, 3)
218492	0	3280035.846078	trace-cmd	25899	sys_exit	NR 313 = -11

File Filter Plots Help

Pointer: 3280035.846600 Cursor: 3280035.846078 Marker A: 3280035.846121 Marker B: 3280035.846600 A,B Delta: 0.000479

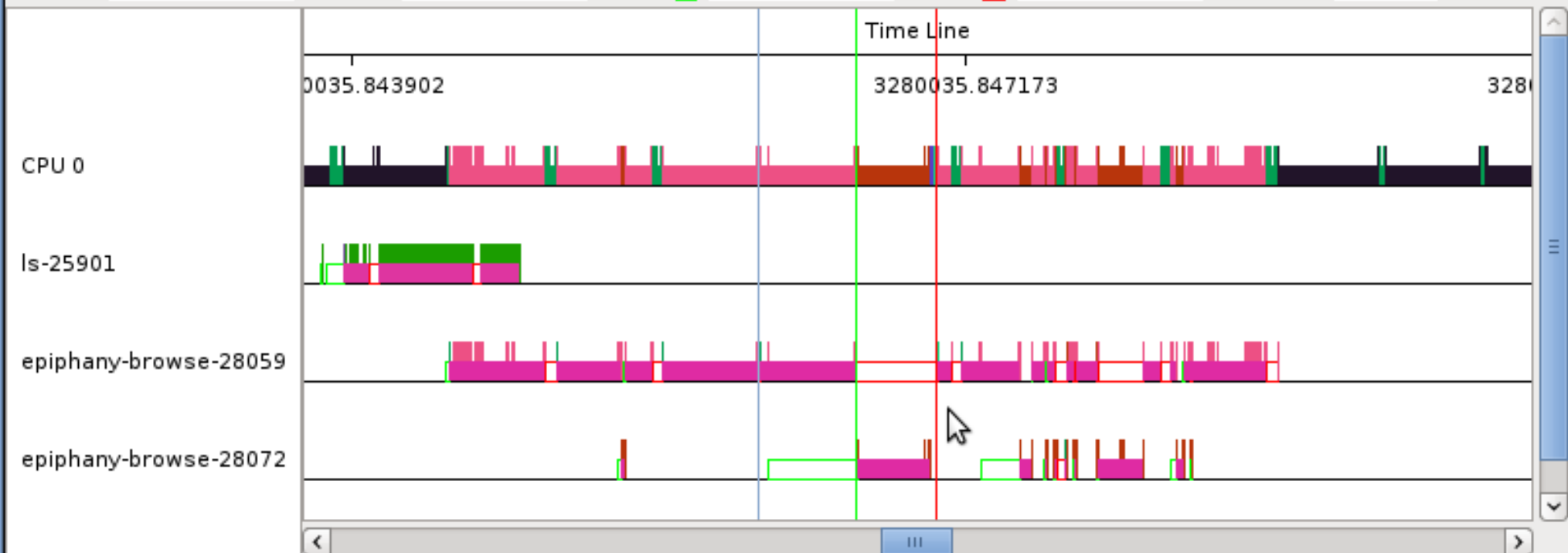


Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
218483	0	3280035.846071	trace-cmd	25899	sys_exit	NR 162 = 0
218484	0	3280035.846072	trace-cmd	25899	sys_enter	NR 313 (5, 0, 7, 0, 1000, 1)
218485	0	3280035.846073	trace-cmd	25899	kmalloc	(tracing_buffers_splice_read+0x121) c
218486	0	3280035.846074	trace-cmd	25899	mm_page_alloc	page=0xffffea00008a02c0 pfn=9044672 c
218487	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218488	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218489	0	3280035.846076	trace-cmd	25899	kfree	(tracing_buffers_splice_read+0x180) c
218490	0	3280035.846076	trace-cmd	25899	sys_exit	NR 313 = 0
218491	0	3280035.846077	trace-cmd	25899	sys_enter	NR 313 (6, 0, 4, 0, 1000, 3)
218492	0	3280035.846078	trace-cmd	25899	sys_exit	NR 313 = -11

File Filter Plots Help

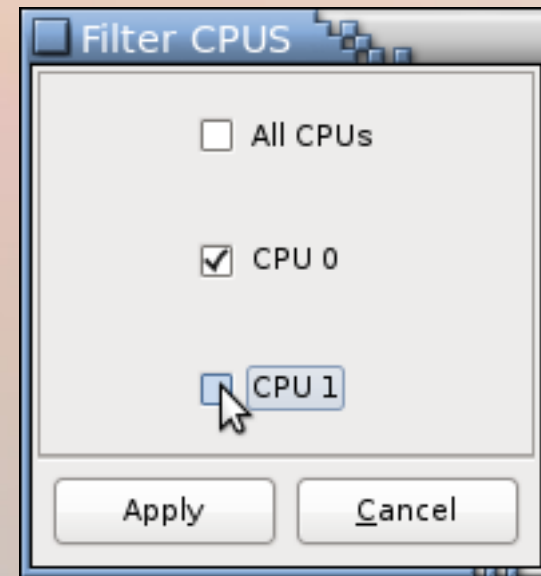
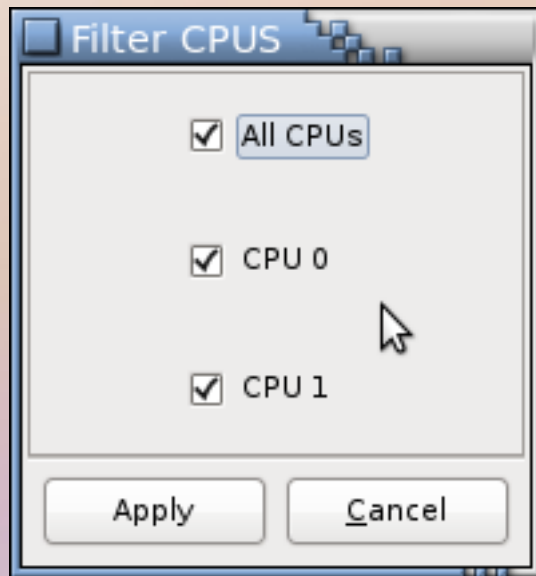
Pointer: 3280035.846923 Cursor: 3280035.846078 Marker A: 3280035.846600 Marker B: 3280035.847027 A,B Delta: 0.000427



Page 1 Search: Column: # contains graph follows

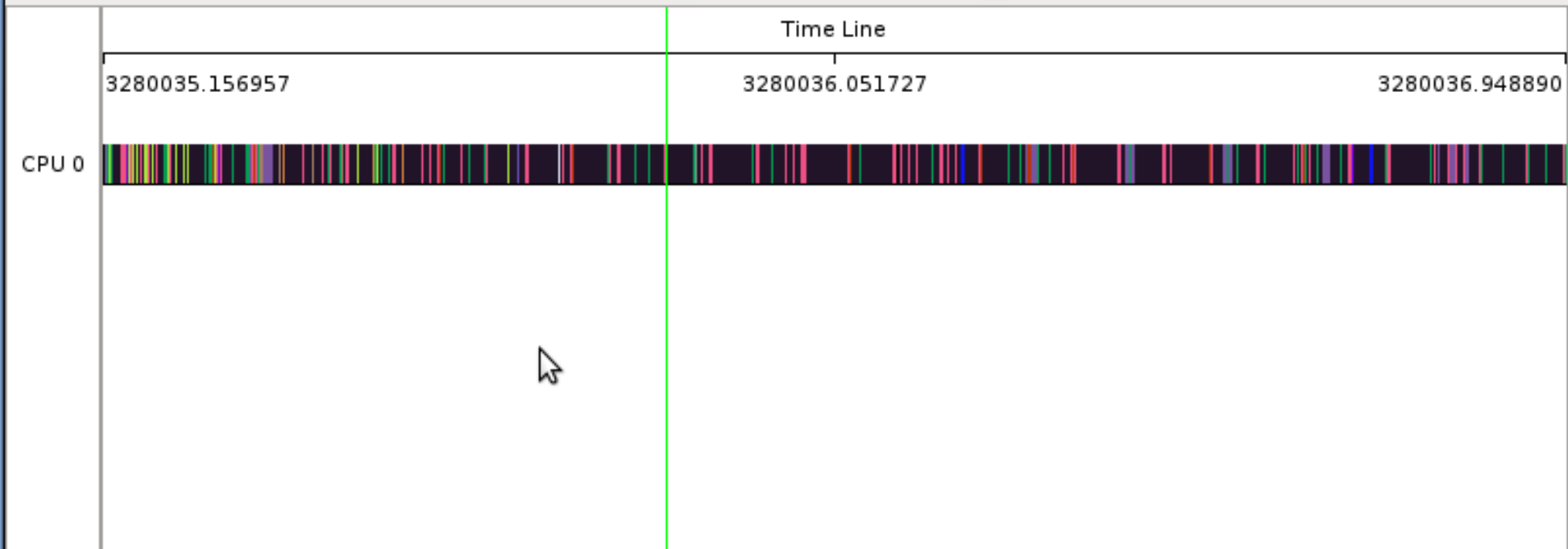
#	CPU	Time Stamp	Task	PID	Latency	Event	Info
218483	0	3280035.846071	trace-cmd	25899	sys_exit	NR 162 = 0
218484	0	3280035.846072	trace-cmd	25899	sys_enter	NR 313 (5, 0, 7, 0, 1000, 1)
218485	0	3280035.846073	trace-cmd	25899	kmalloc	(tracing_buffers_splice_read+0x121) c
218486	0	3280035.846074	trace-cmd	25899	mm_page_alloc	page=0xffffea00008a02c0 pfn=9044672 c
218487	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218488	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218489	0	3280035.846076	trace-cmd	25899	kfree	(tracing_buffers_splice_read+0x180) c
218490	0	3280035.846076	trace-cmd	25899	sys_exit	NR 313 = 0
218491	0	3280035.846077	trace-cmd	25899	sys_enter	NR 313 (6, 0, 4, 0, 1000, 3)
218492	0	3280035.846078	trace-cmd	25899	sys_exit	NR 313 = -11

CPU Plots



File Filter Plots Help

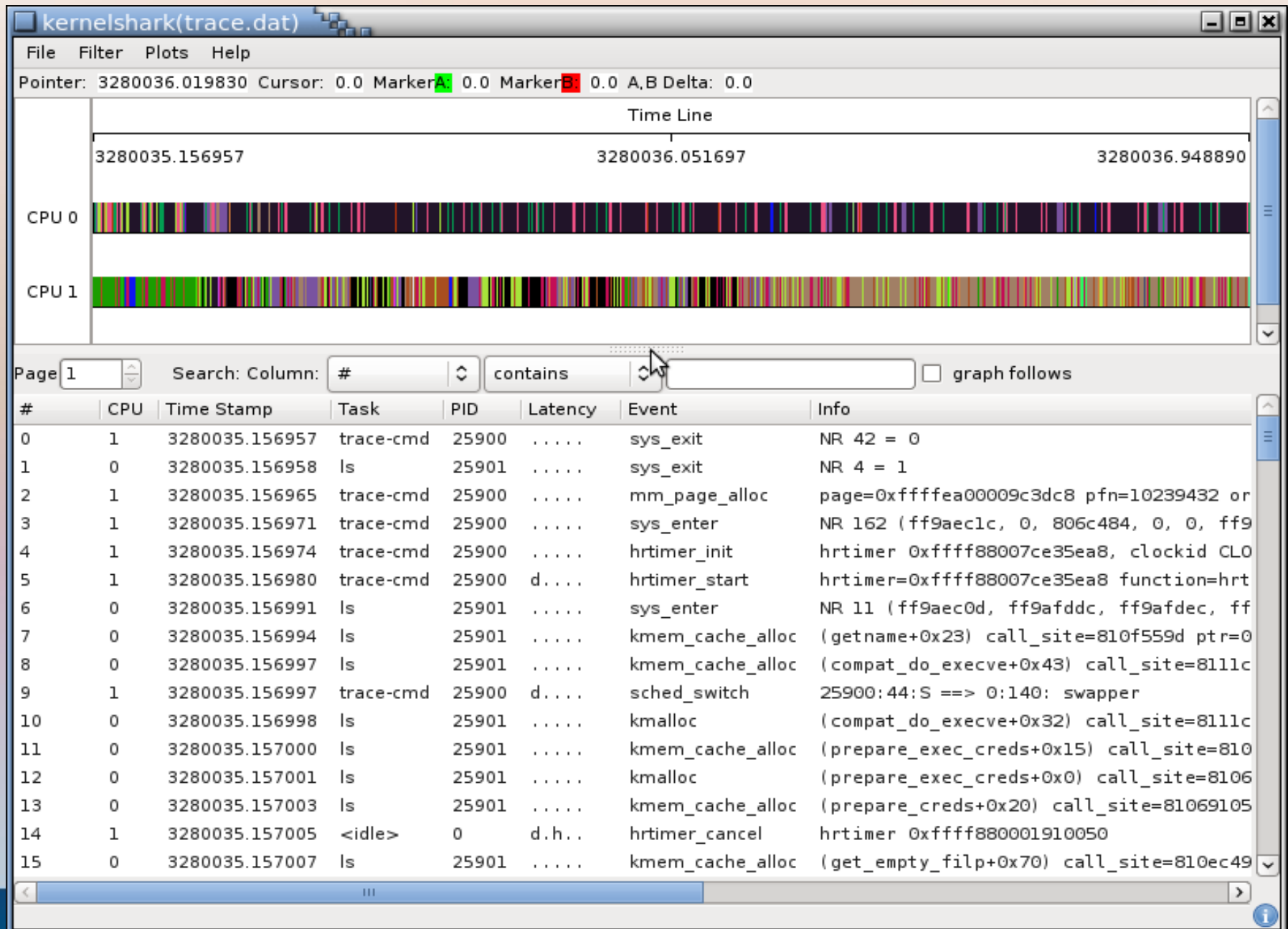
Pointer: 3280035.690469 Cursor: 3280035.846078 Marker A: 3280035.846069 Marker B: 3280035.846085 A,B Delta: 0.000015



Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
218483	0	3280035.846071	trace-cmd	25899	sys_exit	NR 162 = 0
218484	0	3280035.846072	trace-cmd	25899	sys_enter	NR 313 (5, 0, 7, 0, 1000, 1)
218485	0	3280035.846073	trace-cmd	25899	kmalloc	(tracing_buffers_splice_read+0x121) c
218486	0	3280035.846074	trace-cmd	25899	mm_page_alloc	page=0xffffea00008a02c0 pfn=9044672 c
218487	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218488	0	3280035.846075	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008a02c0 pfn=9044672 c
218489	0	3280035.846076	trace-cmd	25899	kfree	(tracing_buffers_splice_read+0x180) c
218490	0	3280035.846076	trace-cmd	25899	sys_exit	NR 313 = 0
218491	0	3280035.846077	trace-cmd	25899	sys_enter	NR 313 (6, 0, 4, 0, 1000, 3)
218492	0	3280035.846078	trace-cmd	25899	sys_exit	NR 313 = -11

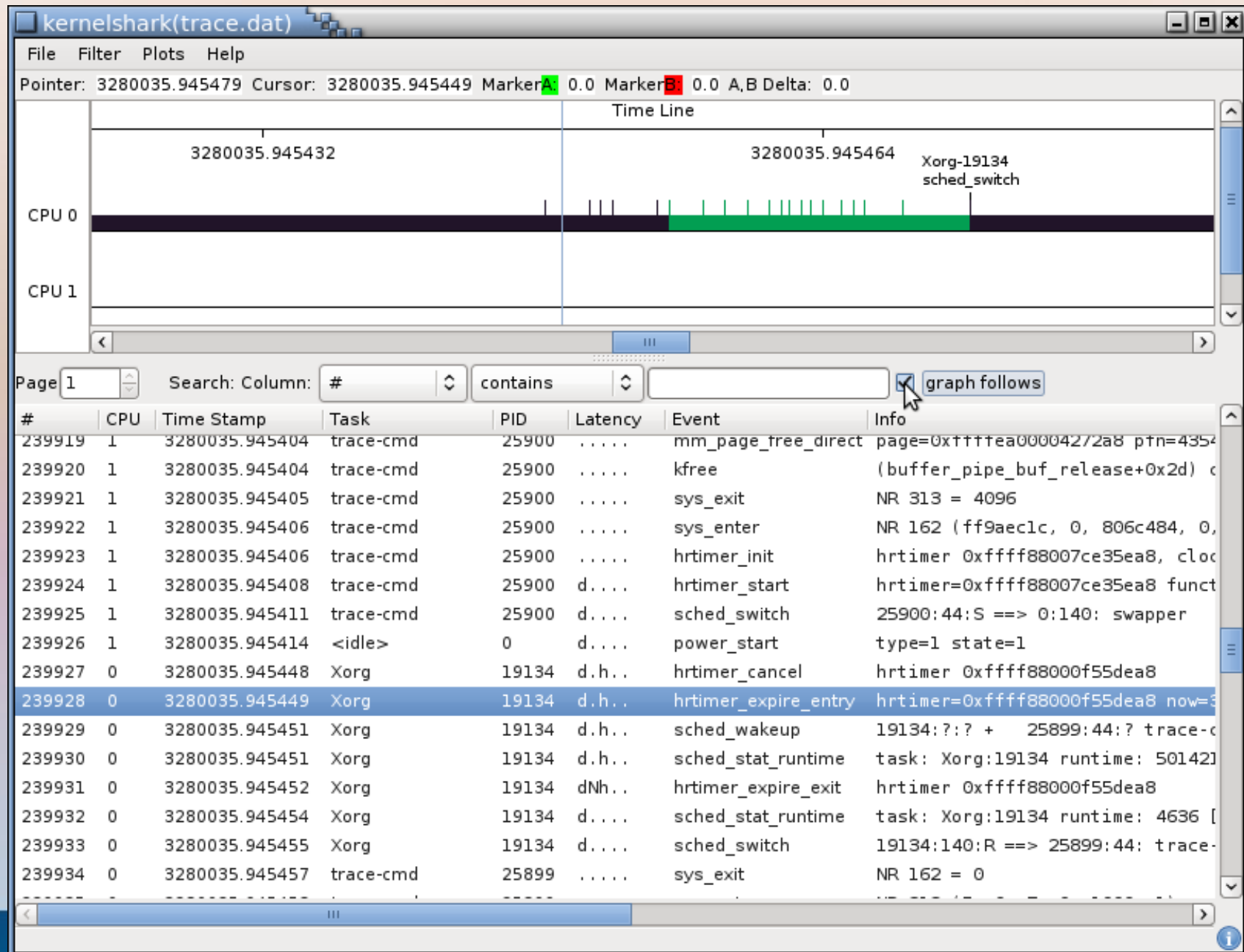
List view



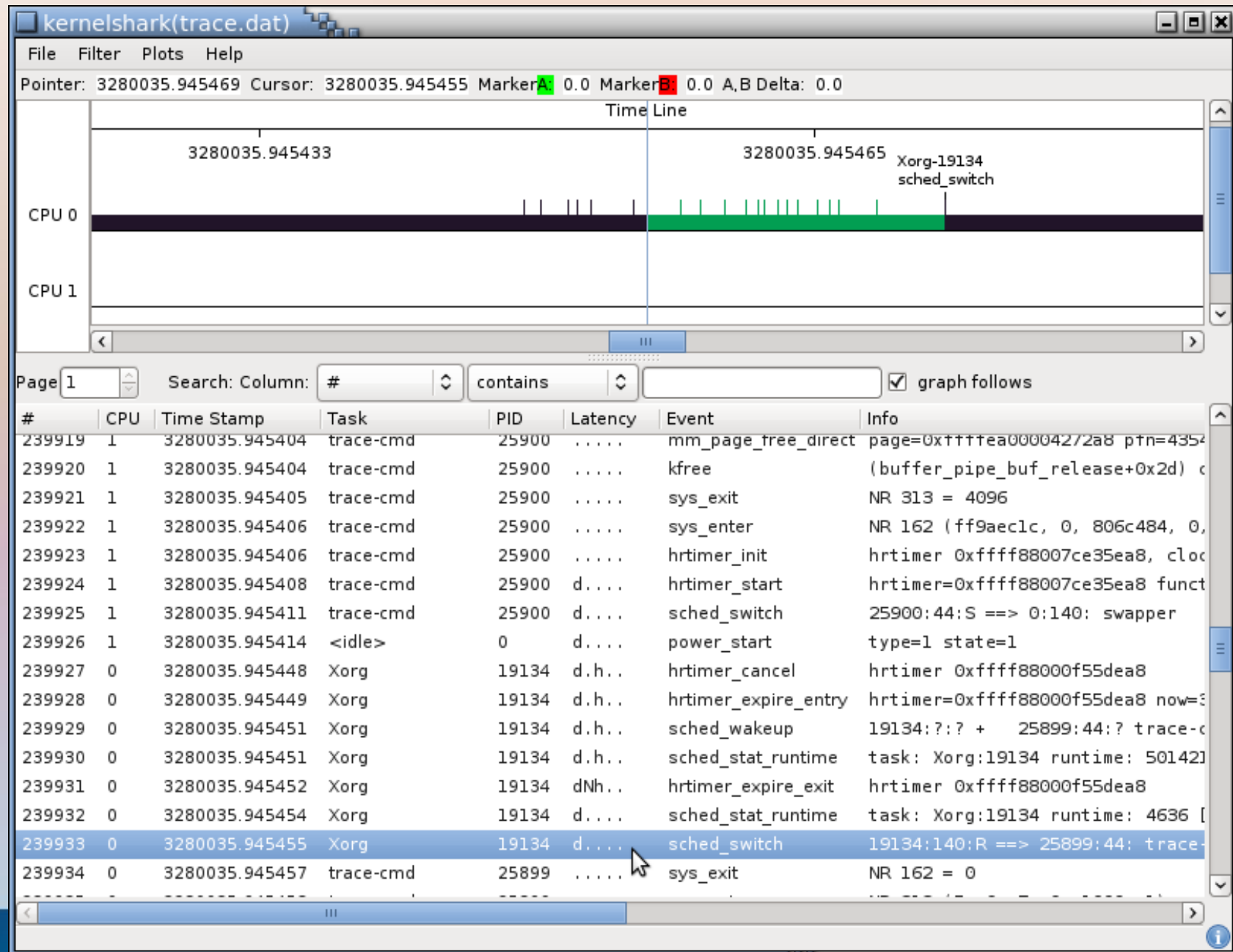
Search the List

- Search by column
 - Contains
 - Full match
 - Does not have

Graph follows toggle

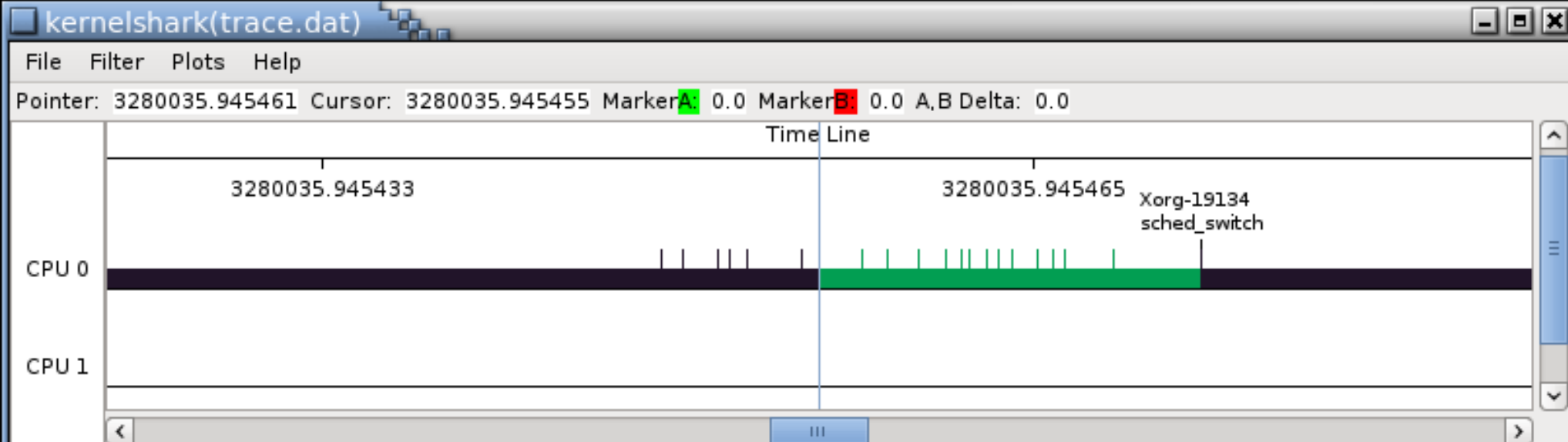


Graph follows toggle



Filtering

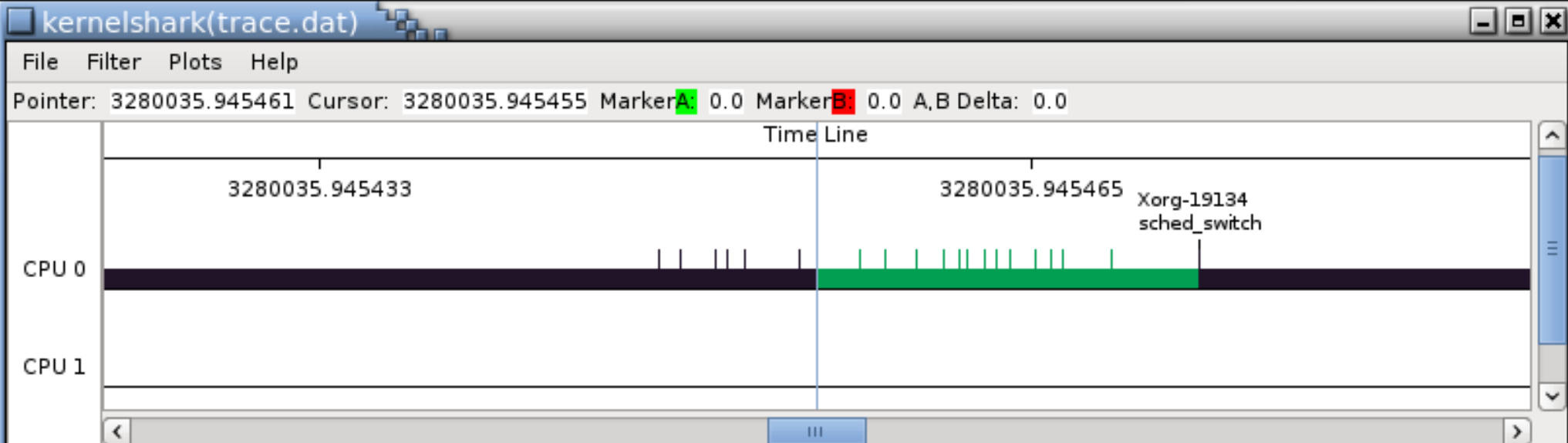
- Filter out tasks
- Filter in tasks
- Filter events
- Filter events based on content



Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
239924	1	3280035.945408	trace-cmd	25900	d...	hrtimer_start	hrtimer=0xffff88007ce35ea8 funct
239925	1	3280035.945411	trace-cmd	25900	d...	sched_switch	25900:44:S ==> 0:140: swapper
239926	1	3280035.945414	<idle>	0	d...	power_start	type=1 state=1
239927	0	3280035.945448	Xorg	19134	d.h..	hrtimer_cancel	hrtimer 0xffff88000f55dea8
239928	0	3280035.945449	Xorg	19134	d.h..	hrtimer_expire_entry	hrtimer=0xffff88000f55dea8 now=3
239929	0	3280035.945451	Xorg	19134	d.h..	sched_wakeup	19134:?:? + 25899:44:? trace-c
239930	0	3280035.945451	Xorg	19134	d.h..	sched_stat_runtime	task: Xorg:19134 runtime: 501421
239931	0	3280035.945452	Xorg	19134	dNh..	hrtimer_expire_exit	hrtimer 0xffff88000f55dea8
239932	0	3280035.945454	Xorg	19134	d...	sched_stat_runtime	task: Xorg:19134 runtime: 4636 [
239933	0	3280035.945455	Xorg	19134	d...	sched_s...	19134:140:S ==> 25899:44: trace-
239934	0	3280035.945457	trace-cmd	25899	sys_exit	
239935	0	3280035.945458	trace-cmd	25899	sys_ente	7, 0, 1000, 1)
239936	0	3280035.945460	trace-cmd	25899	kmalloc	ers_splice_read+0x1
239937	0	3280035.945461	trace-cmd	25899	mm_pag	00008014a8 pfn=8395
239938	0	3280035.945461	trace-cmd	25899	mm_page_tree_direct	page=0xttttea00008014a8 pfn=8395
239939	0	3280035.945462	trace-cmd	25899	mm_page_free_direct	page=0xffffea00008014a8 pfn=8395

- Enable Graph Filter
- Enable List Filter
- Add Xorg-19134 to filter
- Hide Xorg-19134
- Clear Task Filter



Page 1 Search: Column: # contains graph follows

#	CPU	Time Stamp	Task	PID	Latency	Event	Info
239924	1	3280035.945408	trace-cmd	25900	d...	hrtimer_start	hrtimer=0xffff88007ce35ea8 funct
239925	1	3280035.945411	trace-cmd	25900	d...	sched_switch	25900:44:S ==> 0:140: swapper
239926	1	3280035.945414	<idle>	0	d...	power_start	type=1 state=1
239927	0	3280035.945448	Xorg	19134	d.h..	hrtimer_cancel	hrtimer 0xffff88000f55dea8
239928	0	3280035.945449	Xorg	19134	d.h..	hrtimer_expire_entry	hrtimer=0xffff88000f55dea8 now=3
239929	0	3280035.945451	Xorg	19134	d.h..	sched_wakeup	19134:?:? + 25899:44:? trace-c
239930	0	3280035.945451	Xorg	19134	d.h..	sched_stat_runtime	task: Xorg:19134 runtime: 501421
239931	0	3280035.945452	Xorg	19134	dNh..	hrtimer_expire_exit	hrtimer 0xffff88000f55dea8
239932	0	3280035.945454	Xorg	19134	d...	sched_stat_runtime	task: Xorg:19134 runtime: 4636 [
239933	0	3280035.945455	Xorg	19134	d...	sched_switch	19134:140:0 --> 25899:44: trace-
239934	0	3280035.945457	trace-cmd	25899	sys_exit	
239935	0	3280035.945458	trace-cmd	25899	sys_enter	
239936	0	3280035.945460	trace-cmd	25899	kmalloc	
239937	0	3280035.945461	trace-cmd	25899	mm_page_alloc	
239938	0	3280035.945461	trace-cmd	25899	mm_page_free_direct	page=0xffffea000000014a8 pfn=8395
239939	0	3280035.945462	trace-cmd	25899	mm_page_free_direct	page=0xffffea000008014a8 pfn=8395

Enable Graph Filter

Enable List Filter

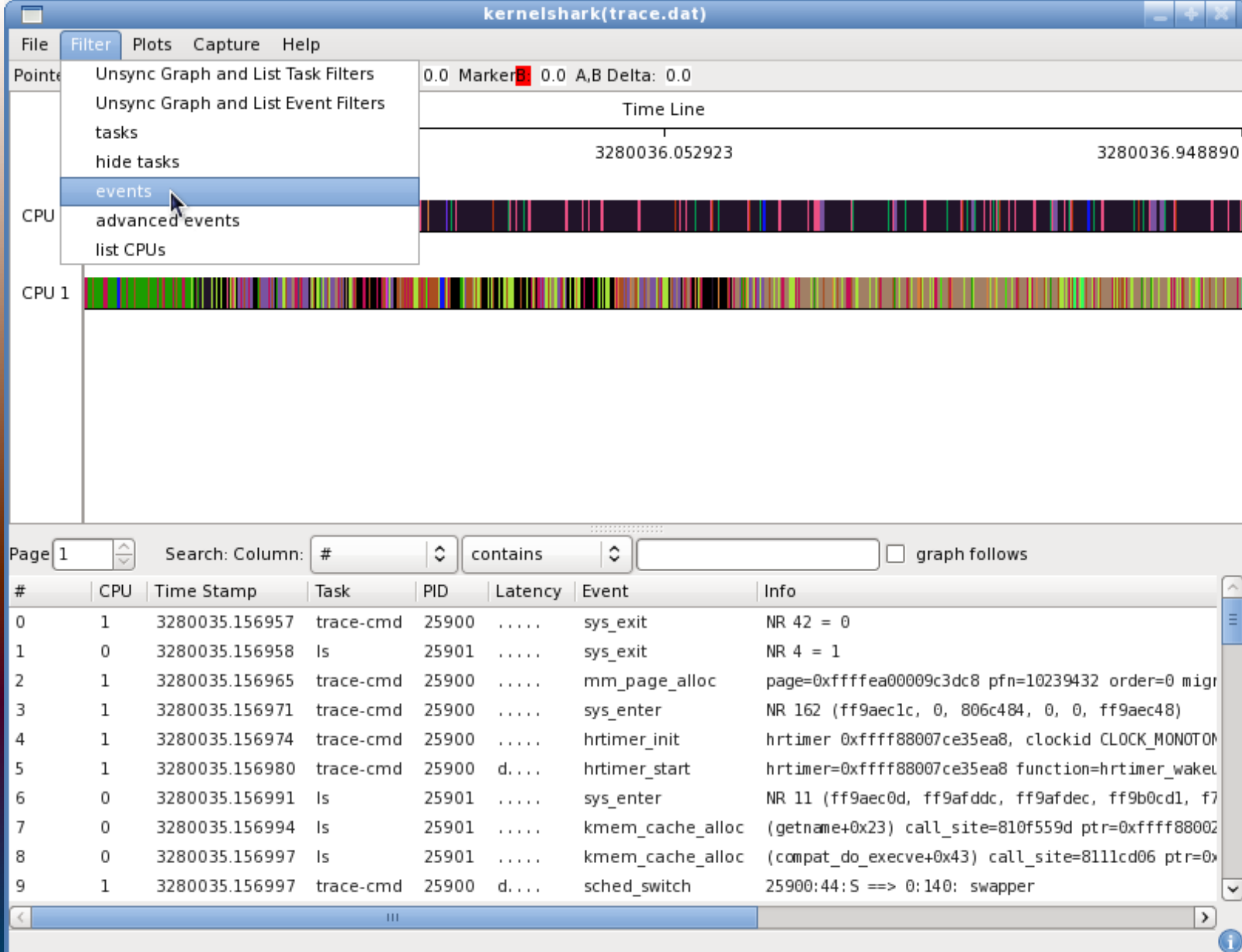
Remove Xorg-19134 from filter

Hide Xorg-19134

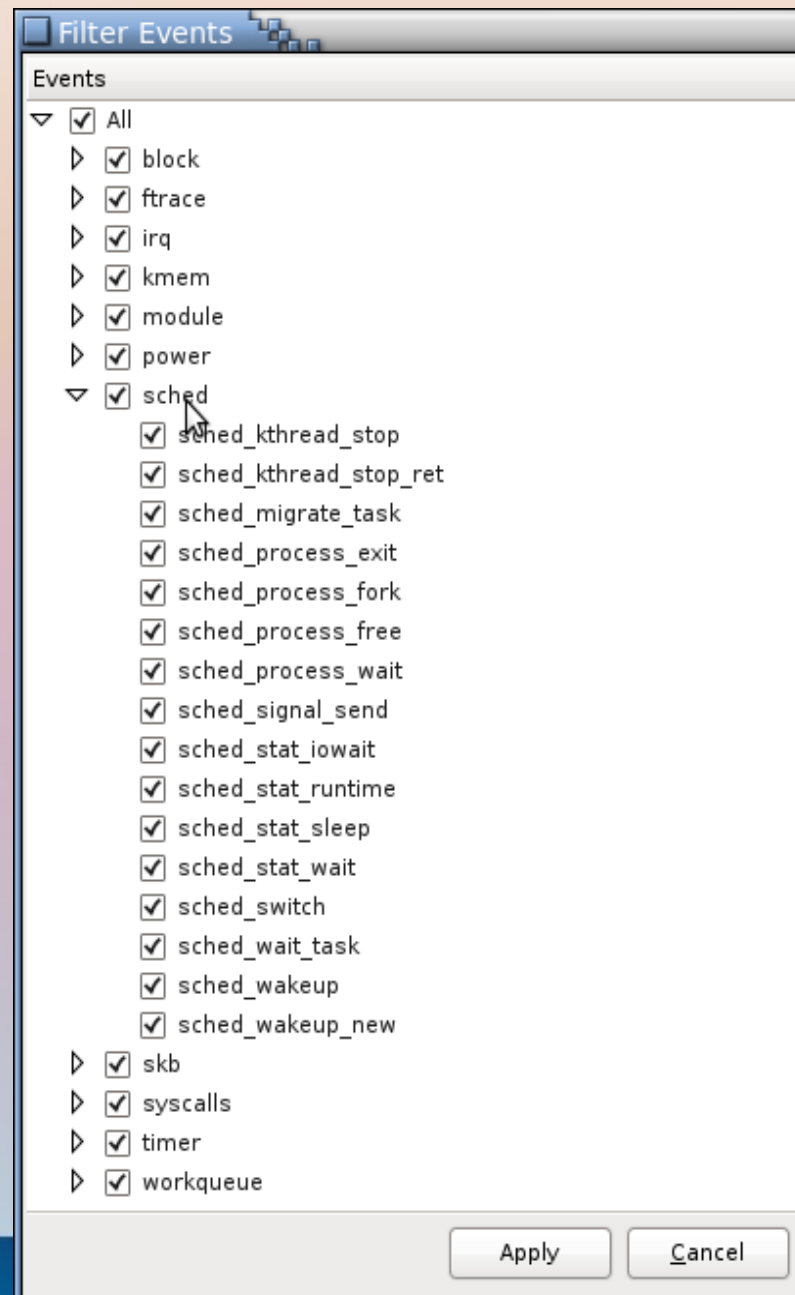
Clear Task Filter

Scheduling events

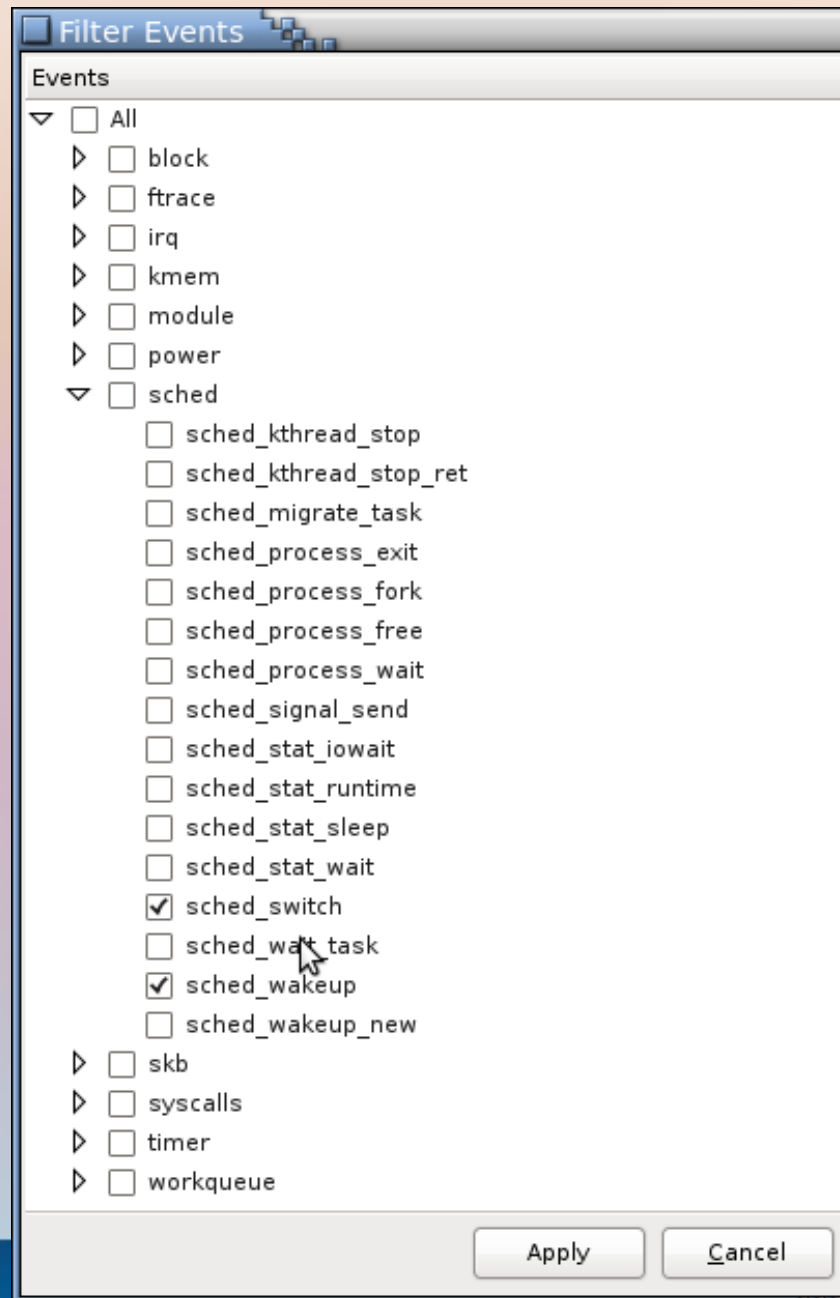
- sched_switch
- sched_wakeup
- sched_wakeup_new
- If a task in either side is to be displayed, then the event will be displayed

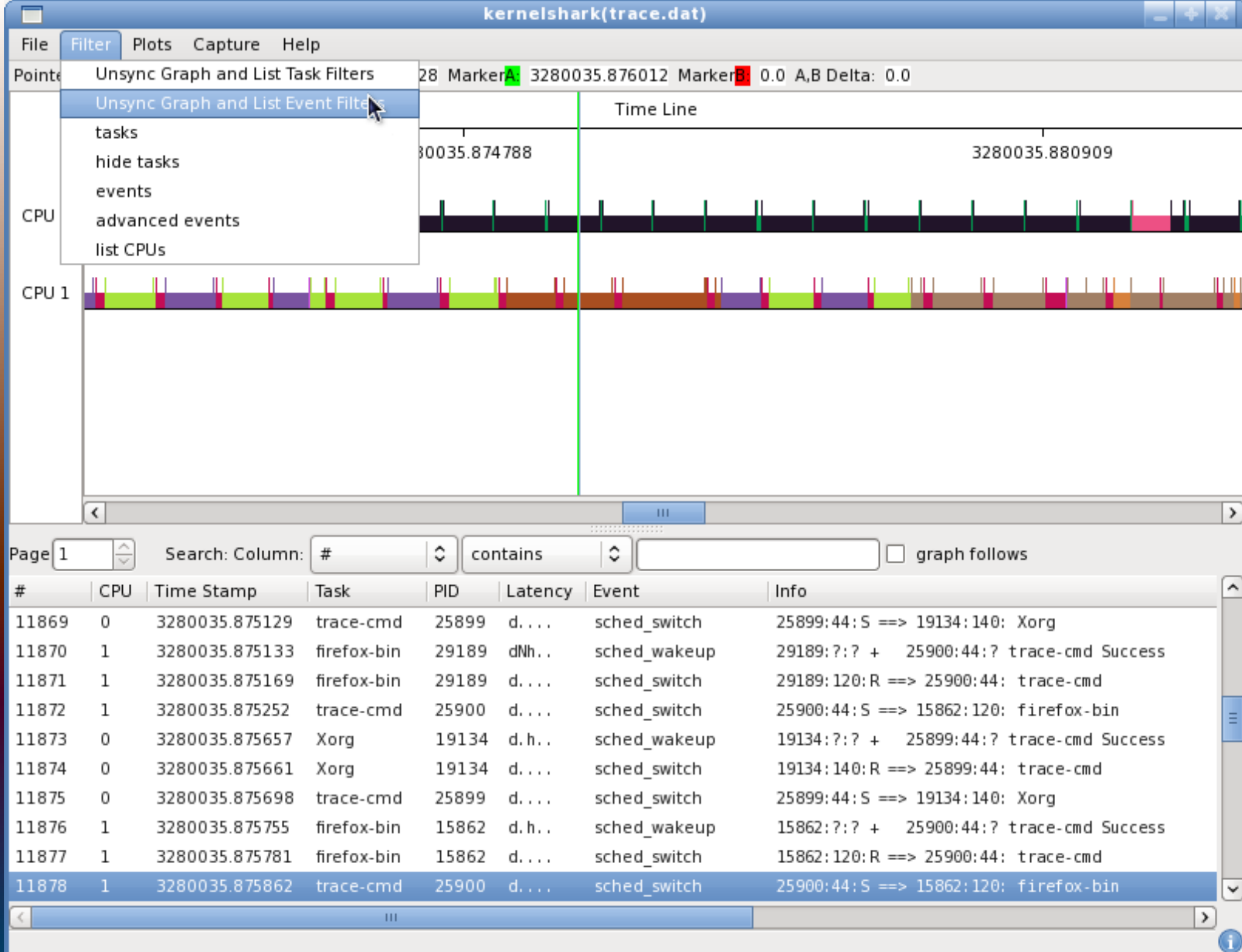


Event Filters



Event Filters





Advanced Event Filtering

☐ Advanced Filters

Delete Filter | Event | Filter

<event>[, <event>] : [!][(<field><op><val>[()])&&/| (<field><op><val>[()])]

Examples:
sched_switch : next_prio < 100 && (prev_prio > 100&& prev_pid != 0)
irq.* : irq != 38
.* : common_pid == 1234

Event: block Op: : Field: common_type

Filter: sched/sched_switch: next_prio < 100 && (prev_prio > 100 && prev_pid != 0)|

Advanced Filtering Language

```
FILTER := EVENTS | EVENTS ':' EXPRESSION
EVENTS := EVENTS ',' EVENTS | SYSTEM '/' EVENT | SYSTEM | EVENT
SYSTEM := any system name
EVENT := any event name
EXPRESSION := EXPRESSION BOOL EXPRESSION | '(' EXPRESSION ')' | OPERATION
BOOL := '&&' | '||'
OPERATION := '!' EXPRESSION | LVALUE CMP RVALUE | LVALUE STRCMP STRVALUE
CMP := '>' | '<' | '==' | '>=' | '<=' | '!='
STRCMP := '==' | '!=' | '=~' | '!~'
RVALUE := integer | FIELD
STRVALUE := string (double quoted value) | FIELD
LVALUE := FIELD | EXPR
EXPR := FIELD OP RVALUE | '(' EXPR ')' | EXPR OP EXPR
FIELD := a field name of an event
OP := '+' | '-' | '*' | '/' | '<<' | '>>' | '&' | '!'
```

Fields not in Events

- Field not in an event evaluates the local condition to false but not the entire condition

```
sched : prev_pid != 0  
sched : !(prev_pid == 0)
```

evaluates to:

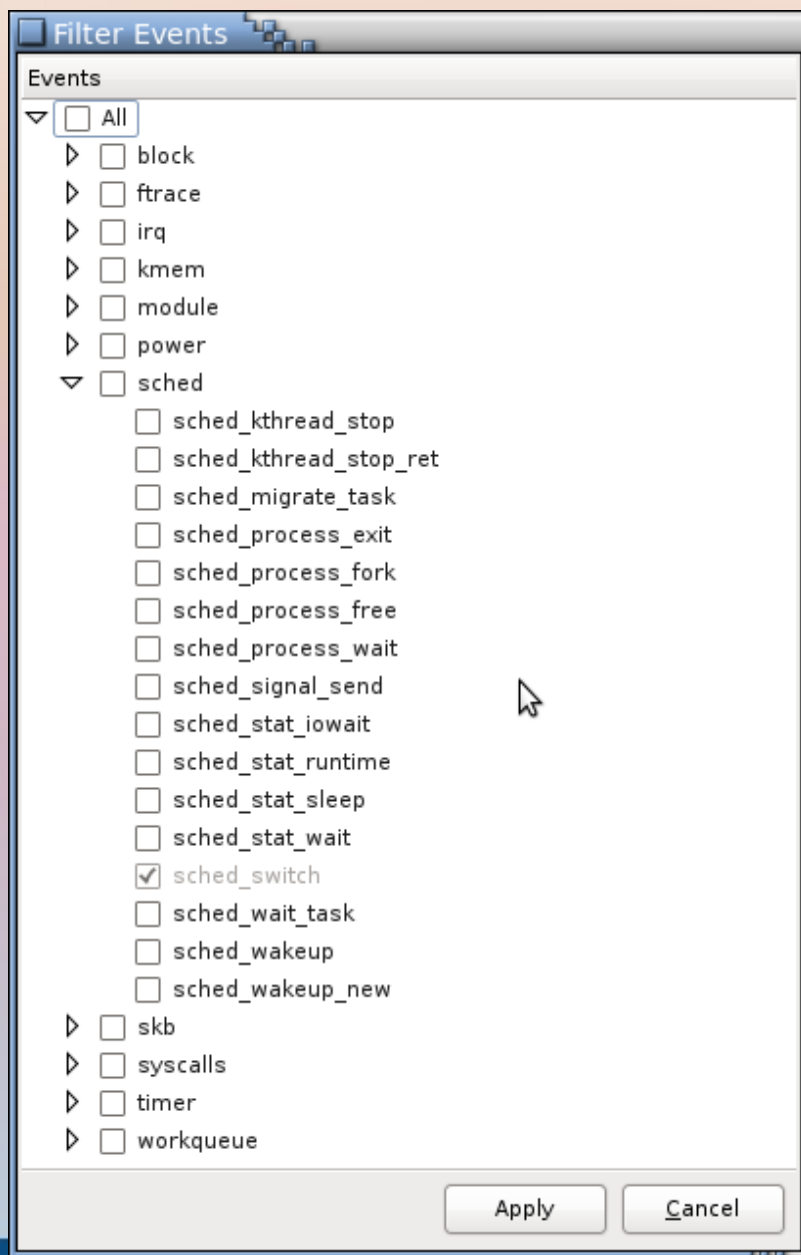
```
sched : FALSE  
sched : !(FALSE)
```

Comparing Strings

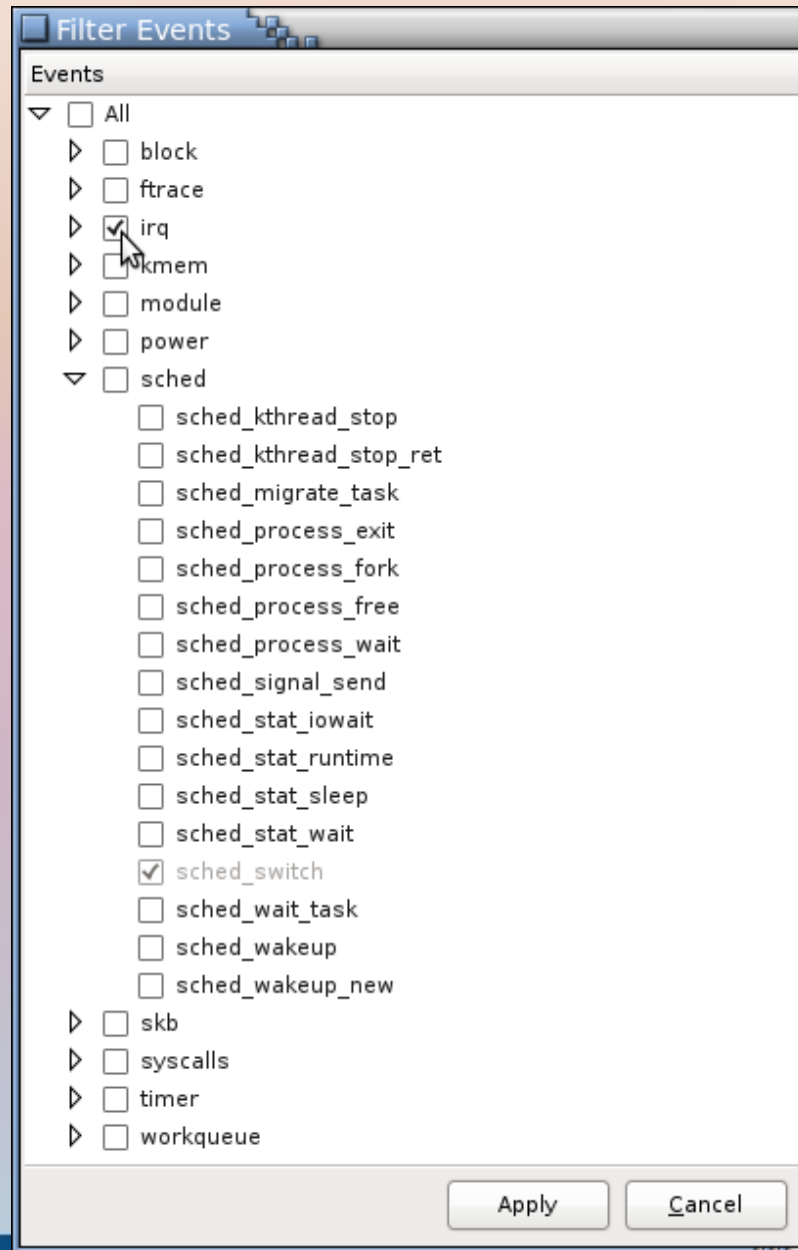
- Strings can compare with regular expressions
 - `regex(7)`
 - Use `=~` or `!~`

```
sched_switch : next_comm =~ "^events/[23]$"
```

Event Filters with Advanced



Adding Events after Advance



Deleting Advanced Filters

☐ Advanced Filters

Delete Filter	Event	Filter
<input checked="" type="checkbox"/>	sched_switch	(next_prio < 100) && (prev_prio > 100)

<event>[, <event>] : [!][(!)<field><op><val>()][&&/|| [(!)<field><op><val>()]]

Examples:
sched_switch : next_prio < 100 && (prev_prio > 100&& prev_pid != 0)
irq.* : irq != 38
.* : common_pid == 1234

Event: Op: Field:

Filter:

Kernel Shark

Demo!

Questions?

