

ADVANCED TROUBLE-SHOOTING OF REAL-TIME SYSTEMS

BERND HUFMANN, ERICSSON

AGENDA



1 Introduction

2 Trace Compass Overview

3 Timing Analysis

4 References

5 Q&A

TRACE COMPASS OVERVIEW



- › **Troubleshooting** tool
- › **Framework** to build trace visualization and analysis tools
- › **Scalable**: handle traces exceeding memory
- › **Extensible** for any trace or log format: Binary, text, XML etc.
- › **Reusable** views and widgets
- › Available as **standalone** product or set of plug-ins

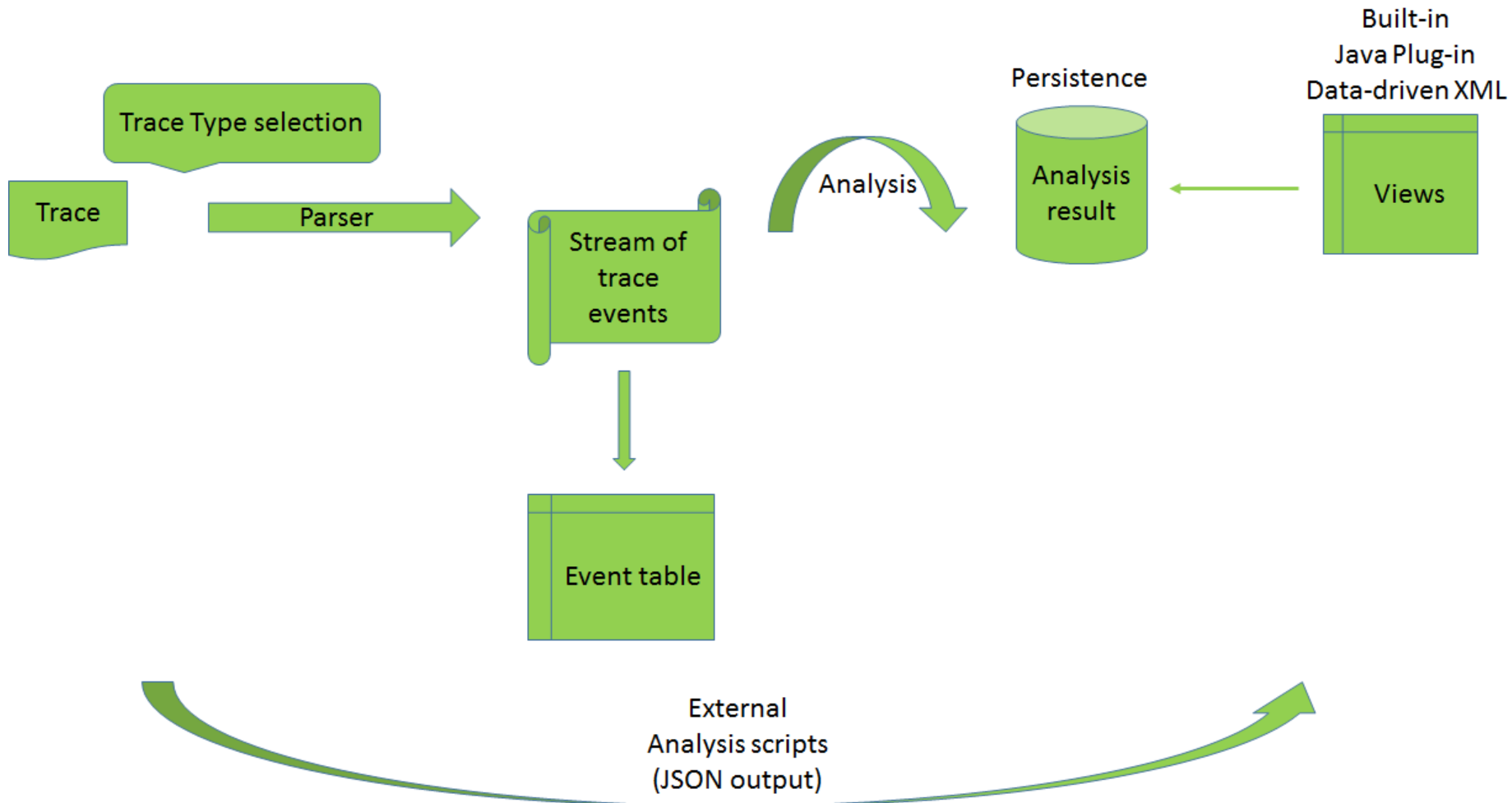
TRACE COMPASS OVERVIEW



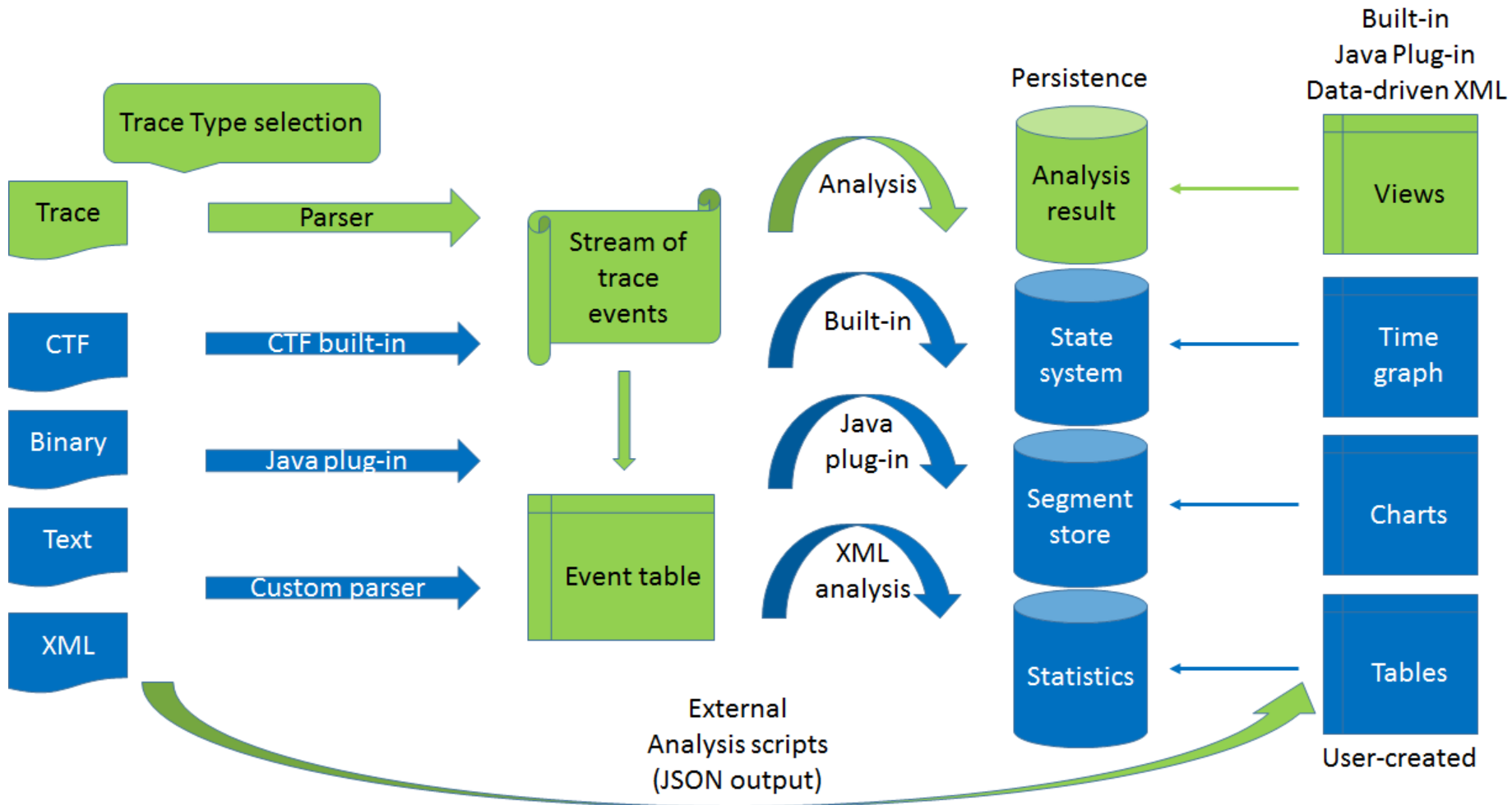
The screenshot displays the Trace Compass interface with several panels:

- Project Explorer:** Shows a tree view of the trace data, including 'Remote', 'Experiments', 'Traces', and 'Tracing'.
- Call Stack:** A table showing function calls with columns for Function, Depth, Entry time, Exit time, and Duration. It includes a timeline visualization of the call stack.
- Sequence Diagram:** A UML sequence diagram showing interactions between 'Challenger' and 'Master' processes, with messages labeled 'BALL_REPLY' and 'BALL_REQUEST'.
- Control:** A tree view of the system's control elements, including 'Local', 'Provider', 'Kernel', and 'Sessions'.
- Ball Game View:** A visualization of the game state over time, showing alternating green and red segments.
- Log View:** A table of log entries with columns for Timestamp, Source, Type, File, and Content. It shows system events like function entries/exits and state changes.
- Histogram:** A bar chart showing the distribution of events over time, with selection start and end markers.

DATA FLOW



DATA FLOW



COMMON FEATURES



› Events Table

Timestamp	Channel	Type	Content
<srch>	<srch>	<srch>	<srch>
2014-06-10 09:36:12.670 038 881	channel0_3	kmem_kmalloc	call_site=0xfffffffffa02d6b3e, ptr=0xfffff8803ca901000, bytes_req=708, bytes_alloc=1024, gfp_
2014-06-10 09:36:12.670 041 650	channel0_3	kmem_kfree	call_site=0xfffffffffa02d6bc8, ptr=0xfffff8803ca901000
2014-06-10 09:36:12.670 043 333	channel0_3	sched_stat_runtime	comm=lttng-consumerd, tid=4760, runtime=8344, vruntime=168845555
2014-06-10 09:36:12.670 043 773	channel0_3	sched_stat_sleep	comm=lttng-consumerd, tid=4759, delay=7467792
2014-06-10 09:36:12.670 044 512	channel0_3	sched_wakeup	comm=lttng-consumerd, tid=4759, prio=120, success=1, target_cpu=0
2014-06-10 09:36:12.670 045 543	channel0_2	mm_page_free	page=0xfffffea000e14ba40, order=0
2014-06-10 09:36:12.670 046 181	channel0_3	kmem_kmalloc	call_site=0xfffffffffa02d6b3e, ptr=0xfffff880405b7d200, bytes_req=298, bytes_alloc=512, gfp_f
2014-06-10 09:36:12.670 047 533	channel0_3	kmem_kfree	call_site=0xfffffffffa02d6bc8, ptr=0xfffff880405b7d200
2014-06-10 09:36:12.670 048 003	channel0_3	kmem_kmalloc	call_site=0xfffffffffa02d6b3e, ptr=0xfffff8803c2a891c0, bytes_req=28, bytes_alloc=32, gfp_flag
2014-06-10 09:36:12.670 048 351	channel0_3	kmem_kfree	call_site=0xfffffffffa02d6bc8, ptr=0xfffff8803c2a891c0

COMMON FEATURES



Timestamp	Channel	CPU	Event type	Contents
<srch>	<srch>	<srch>	kmem	<srch>
13:04:20.683 994 782	channel0_0	0	syscall_exit_epoll_wa	ret=1, events=
13:04:20.683 999 074	channel0_1	1	kmem_kmalloc	call_site=0xffff
13:04:20.683 999 522	channel0_1	1	kmem_kfree	call_site=0xffff
13:04:20.683 999 925	channel0_0	0	syscall_entry_ioctl	fd=24, cmd=6;
13:04:20.684 000 467	channel0_1	1	kmem_kmalloc	call_site=0xffff
13:04:20.684 001 833	channel0_1	1	kmem_kfree	call_site=0xffff

› Searching

Event type matches "kmem"

Timestamp	Channel	CPU	Event type	Contents
<srch>	<srch>	<srch>	<srch>	<srch>
124345/372115				
13:04:20.683 968 723	channel0_1	1	kmem_kmalloc	call_site=0xffff
13:04:20.683 971 501	channel0_1	1	kmem_kfree	call_site=0xffff
13:04:20.683 975 864	channel0_1	1	kmem_kmalloc	call_site=0xffff
13:04:20.683 976 917	channel0_1	1	kmem_kfree	call_site=0xffff

› Filtering

Timestamp	Channel	CPU	Event type	Contents
<srch>	<srch>	<srch>	<srch>	<srch>
13:04:20.683 990 430	channel0_0	0	rcu_utilization	s=End context
13:04:20.683 990 881	channel0_0	0	sched_stat_wait	comm=ltng-c
13:04:20.683 991 150	channel0_0	0	sched_switch	prev_comm=s
13:04:20.683 994 782	channel0_0	0	syscall_exit_epoll_wa	ret=1, events=
13:04:20.683 999 074	channel0_1	1	kmem_kmalloc	call_site=0xffff
13:04:20.683 999 522	channel0_1	1	kmem_kfree	call_site=0xffff

› Highlighting

COMMON FEATURES



› Trace annotation (bookmarks) and markers

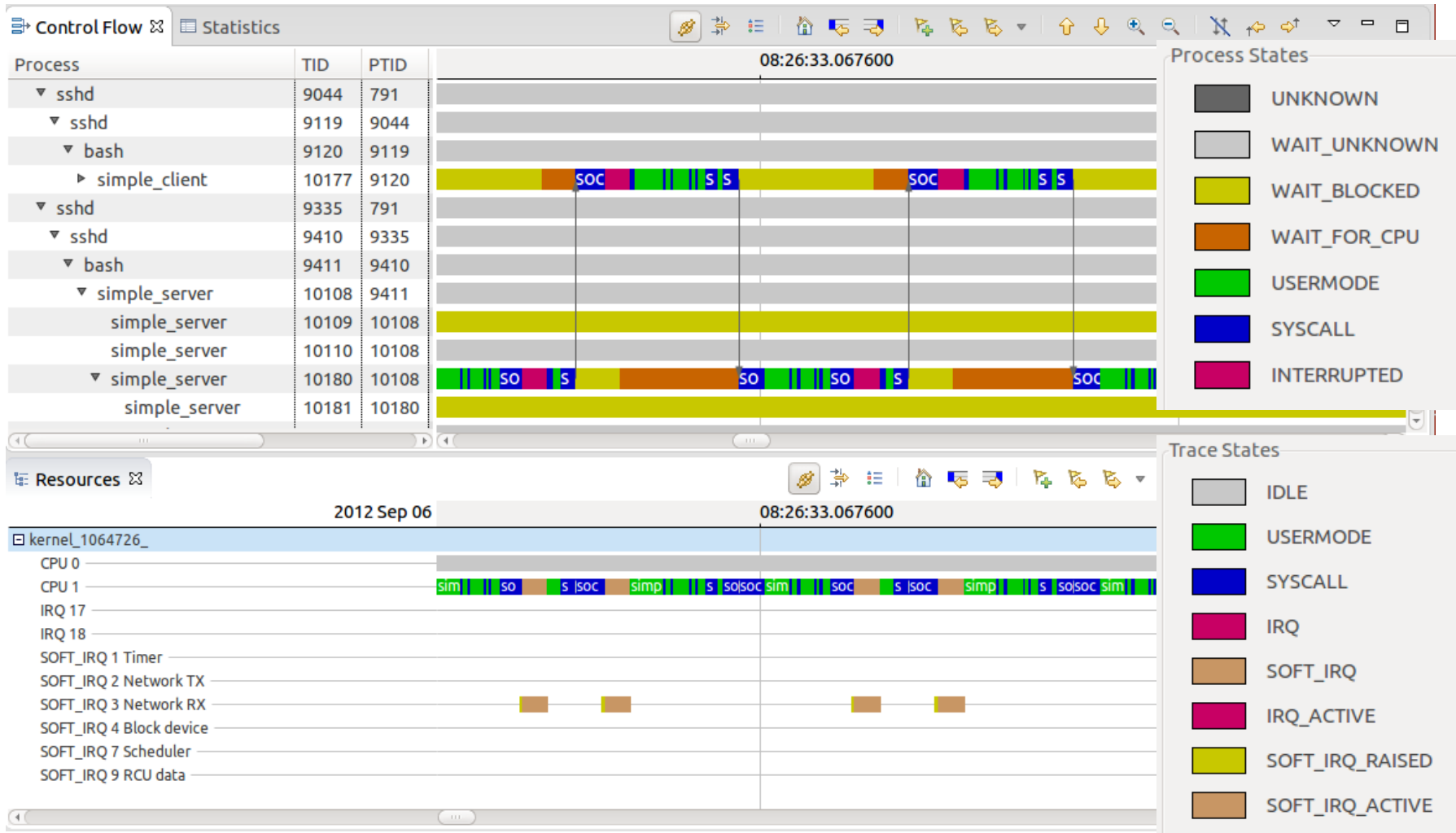
The screenshot displays the LTTng trace viewer interface. The top panel, titled 'Control Flow', shows a timeline for processes: unity-2d-shell (TID 137), unity-2d-shell (TID 220), firefox (TID 220), and which (TID 770). A red box labeled 'Wakeup' and a purple box labeled 'Unusual delay' are placed on the timeline. The middle panel, titled 'LTTng/trace2', shows a table of events:

Timestamp	Channel	CPU	Event type	Contents
<srch>	<srch>	<srch>	<srch>	<srch>
52.946 073 199	channel0_0	0	sched_stat_sleep	comm=unity-2d-shell, tid=1372, delay=7961094
52.946 074 535	channel0_0	0	sched_wakeup	comm=unity-2d-shell, tid=1372, prio=120, success=1, target_
52.946 077 188	channel0_0	0	exit_syscall	ret=64

The bottom panel, titled 'Bookmarks', shows a table with 2 items:

Description	Resource	Path	Location
Unusual delay	trace2_	/My Project/Traces/LTTng/trace2	timestamp [52.967 923 616, 52.984 234 125]
Wakeup	trace2_	/My Project/Traces/LTTng/trace2	rank 243049

STATEFUL ANALYSES

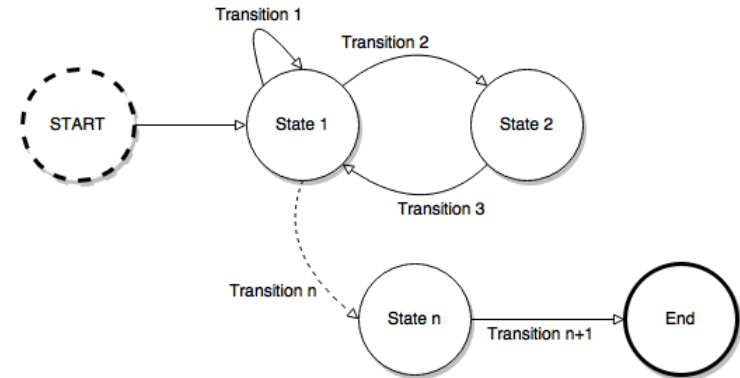


XML ANALYSIS & VIEWS



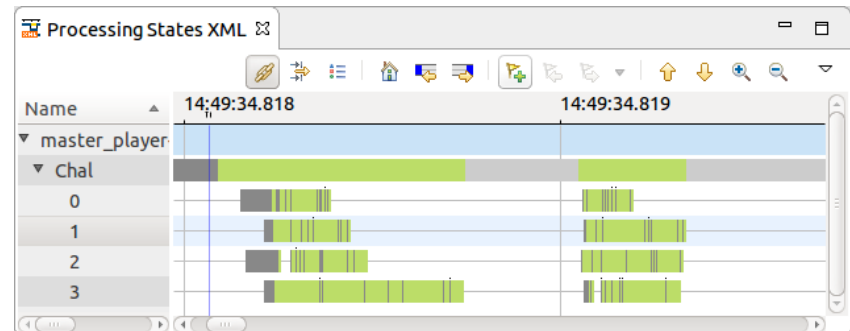
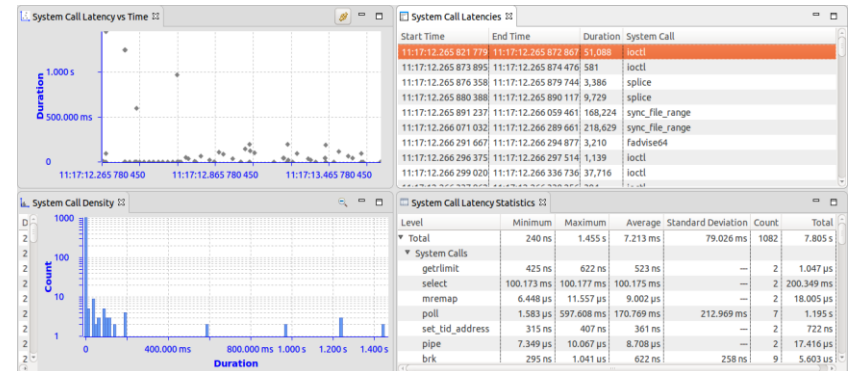
› Pattern analysis

- Find a sequence of data within a trace



› Customize Trace Compass without adding code

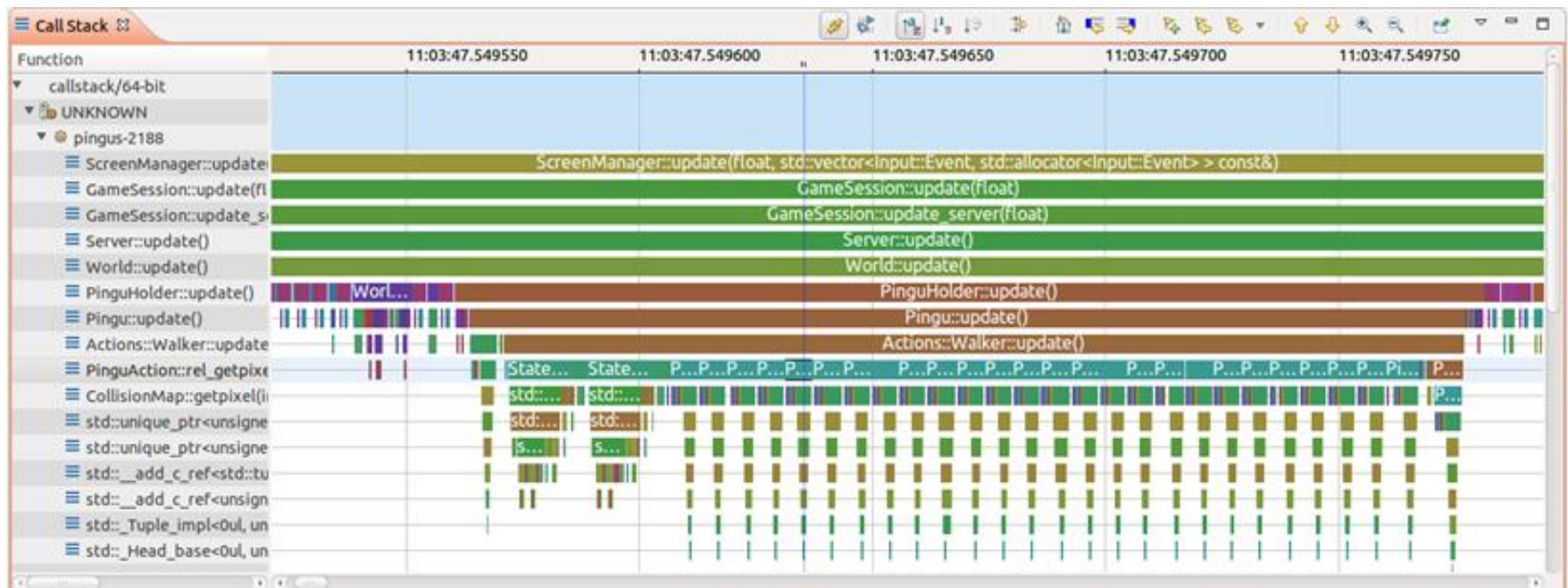
- Generate state systems
- Do timing analysis
- Define specialized views



CALL STACK VIEW



- › Extensible view to display of **call stacks** over time
- › LTTng-UST and finstrument-functions of GCC



TRACE CORRELATION



- › Trace Compass can open **multiple traces** together to view it **as one**
 - This is called an Experiment
- › Useful for
 - Traces coming from multiple **nodes**
 - Traces from applications written in different **languages**
 - Different **layers** (network, etc.)
- › Traces can be synchronized by time
 - Manually
 - Automatic algorithm (extensible)

BUILT-IN TRACE TYPES



- › Linux Tracing Toolkit - LTTng (UST, Kernel)
- › Text & XML Logs (custom parsers)
- › Common Trace Format – CTF
 - application, kernel, HW, bare metal, etc.
- › Packet Capture
- › Best Trace Format - BTF
- › GDB Trace Points

TIMING ANALYSIS



- › **Real-time** systems
- › We have two metrics to analyse
 - › what is the data and **when** did it come
- › **Timing** is as important as **data**
- › Measure time between a **start** and **end** state
 - Simple: Start and end event
 - Often: State Machine to determine start and end
- › Represent execution times, latencies, latency chains etc.

TIMING ANALYSIS



- › **Locate** timing problems

 - › **Missed deadlines**

 - › **Potential** missed deadline (find problem before it occurs)

- › **Analyze** timing problems

 - › Find root cause and solution

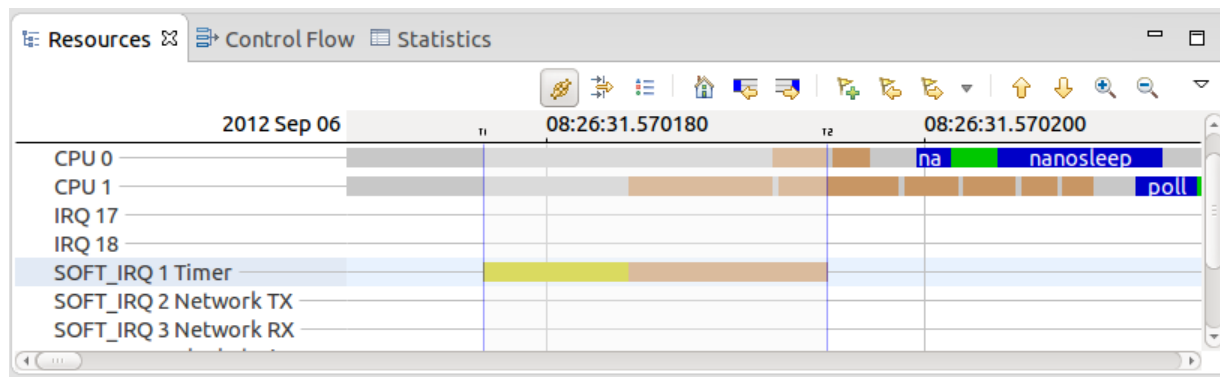
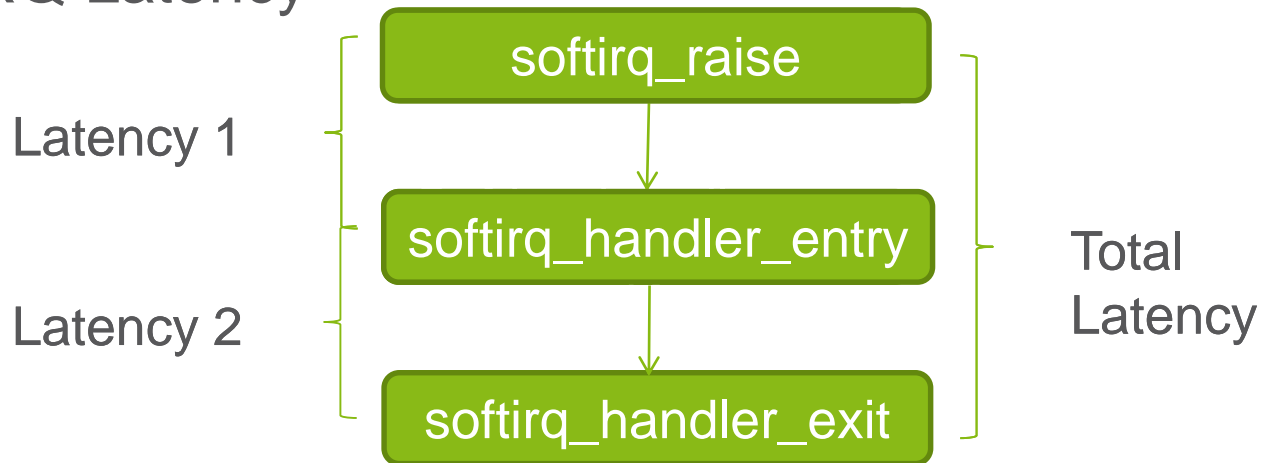
 - › Solve **difficult** to debug sporadic problems

EXAMPLE

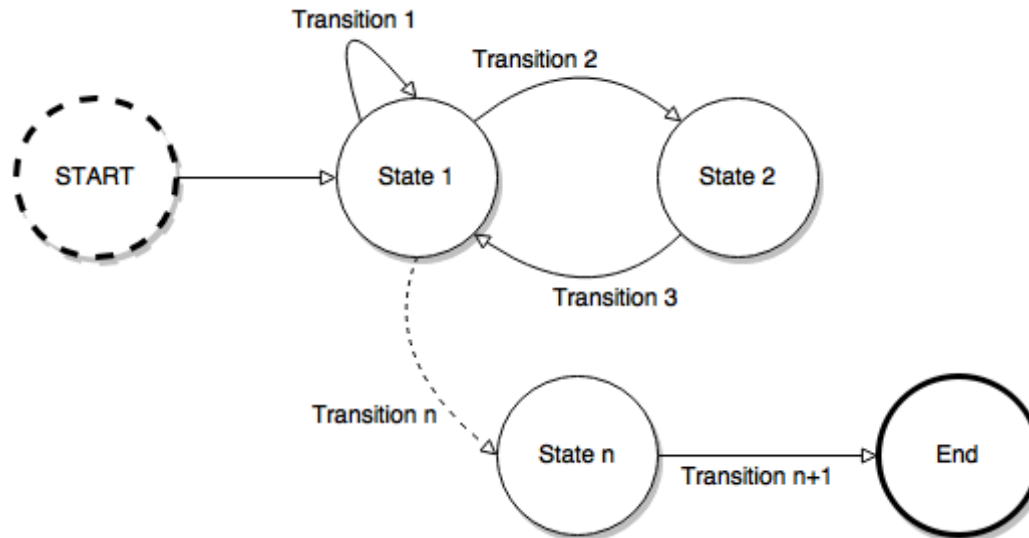


› Soft IRQ Latency

Parameter: CPU ID, IRQ #



GENERALIZATION



- › Time between **start** and **end**
- › Time for each transition
- › Percentage sub-duration vs total

YOUR TIMING ANALYSIS



- › **Define** a **state machine** for timing analysis
 - Implementation in **Java** as Trace Compass extension
 - **Data-driven** pattern matching (in **XML**)
 - › Defining timing analyses on-the-fly
- › **Store** in a built-in segment store
- › **Visualize** data in various supplied views

VISUALIZATION



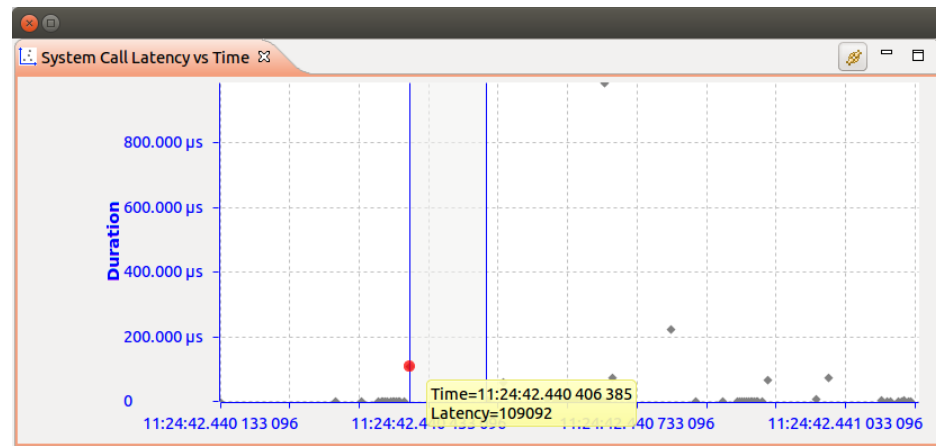
› Table

- Get **raw data**
- Explore data
- Sorting, highlighting, filtering

› Scatter Chart

- Latency vs Time
- Have a **big picture** of the current range

Start Time	End Time	Duration	System Call
11:24:42.440 299 808	11:24:42.440 301 205	1,397	getpid
11:24:42.440 336 893	11:24:42.440 338 081	1,188	getegid
11:24:42.440 361 268	11:24:42.440 363 503	2,235	tgkill
11:24:42.440 364 900	11:24:42.440 366 157	1,257	tgkill
11:24:42.440 367 694	11:24:42.440 369 789	2,095	futex
11:24:42.440 371 116	11:24:42.440 372 582	1,466	futex
11:24:42.440 374 119	11:24:42.440 376 633	2,514	setresgid
11:24:42.440 378 100	11:24:42.440 379 147	1,047	geteuid
11:24:42.440 381 522	11:24:42.440 382 709	1,187	tgkill
11:24:42.440 383 967	11:24:42.440 385 154	1,187	tgkill
11:24:42.440 386 551	11:24:42.440 387 948	1,397	futex
11:24:42.440 389 274	11:24:42.440 390 671	1,397	futex
11:24:42.440 392 068	11:24:42.440 394 303	2,235	setresuid
11:24:42.440 397 865	11:24:42.440 399 192	1,327	umask

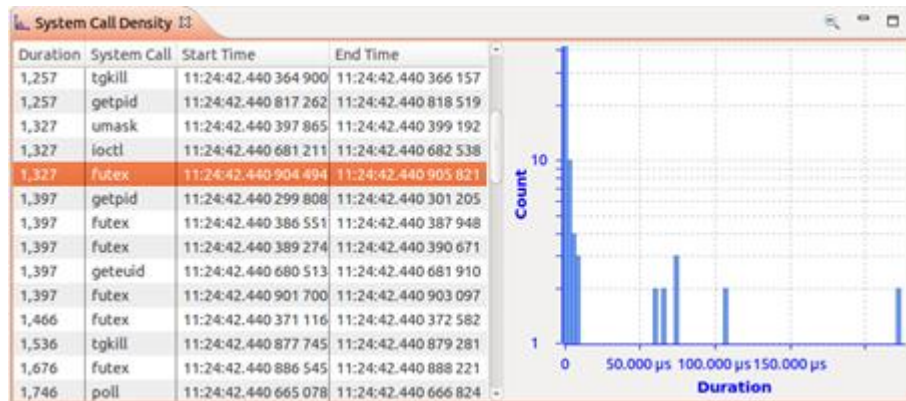


VISUALIZATION



› Distribution Chart

- Find **outliers** and **modes** easily



› Statistics

- Min, max, average etc.
- Find **worst offenders**
- Find worst possible offender combination

The screenshot shows the 'System Call Latency Statistics' tool. It displays a table with columns: Level, Minimum, Maximum, Average, Standard Deviation, Count, and Total. The 'Total' row is expanded to show 'System Calls'.

Level	Minimum	Maximum	Average	Standard Deviation	Count	Total	
▼ Total	109 ns	2.136 s	2.497 ms	45.280 ms	80339	200.605 s	
▼ System Calls							
getrlimit	144 ns	9.452 µs	442 ns	357 ns	3889	1.721 ms	
sendto	873 ns	228.474 µs	34.924 µs	25.702 µs	1017	35.518 ms	
select	782 ns	1.000 s	11.461 ms	77.239 ms	491	5.627 s	
poll		Go to minimum	77 s	6.011 ms	62.582 ms	6437	38.693 s
io_getevent		Go to maximum	ms	633.226 µs	859.790 µs	813	514.813 ms
set_tid_adre	1.085 µs	1.085 µs	1.085 µs	—	1	1.085 µs	
pipe	3.753 µs	16.448 µs	7.275 µs	3.921 µs	11	80.023 µs	
brk	292 ns	22.246 µs	4.135 µs	6.033 µs	23	95.112 µs	
rt_sigprocm	130 ns	2.108 µs	388 ns	274 ns	814	316.324 µs	

TIMING ANALYSIS



- › **Locate** timing problems
 - › Missed **deadlines**
 - › **Potential** missed deadline (find problem before it occurs)
- › **Analyze** timing problems
 - › **Find root cause and solution**
 - › **Solve difficult** to debug sporadic problems

EXAMPLE ROOT CAUSES

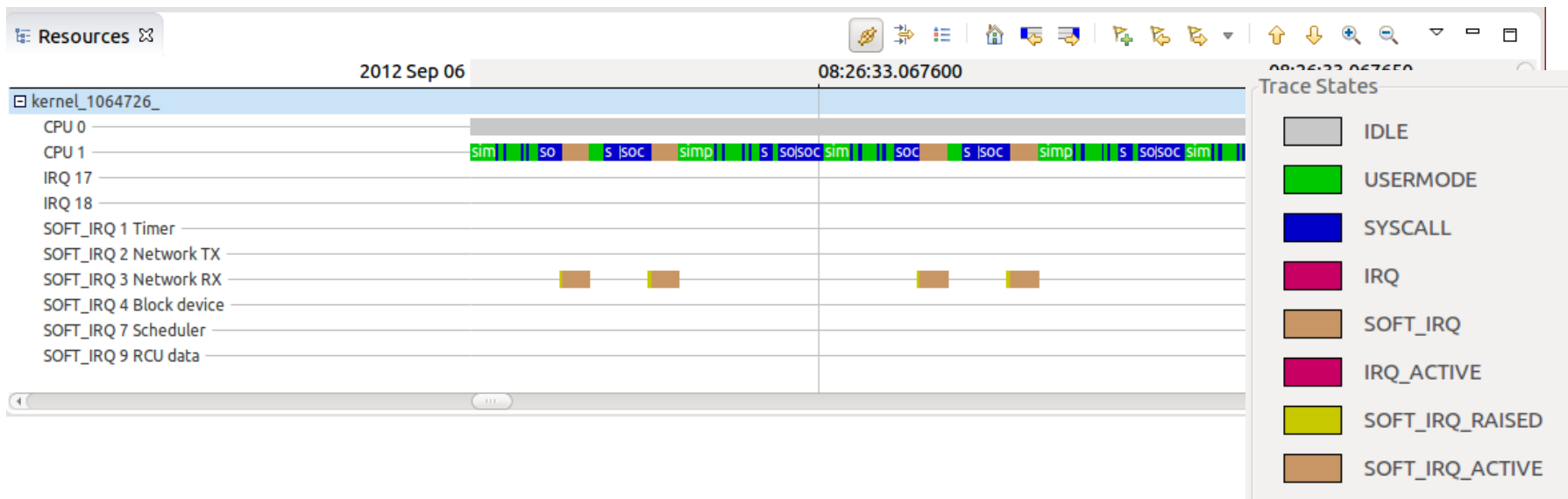


- › System **overload**
- › System **misconfiguration** (e.g. wrong priorities of tasks)
- › **Priority** inversion
 - Lower priority task is blocking higher priority task (indirectly)
- › **Blocked** threads, starvation, deadlock
- › **Slow** code

RESOURCES VIEW



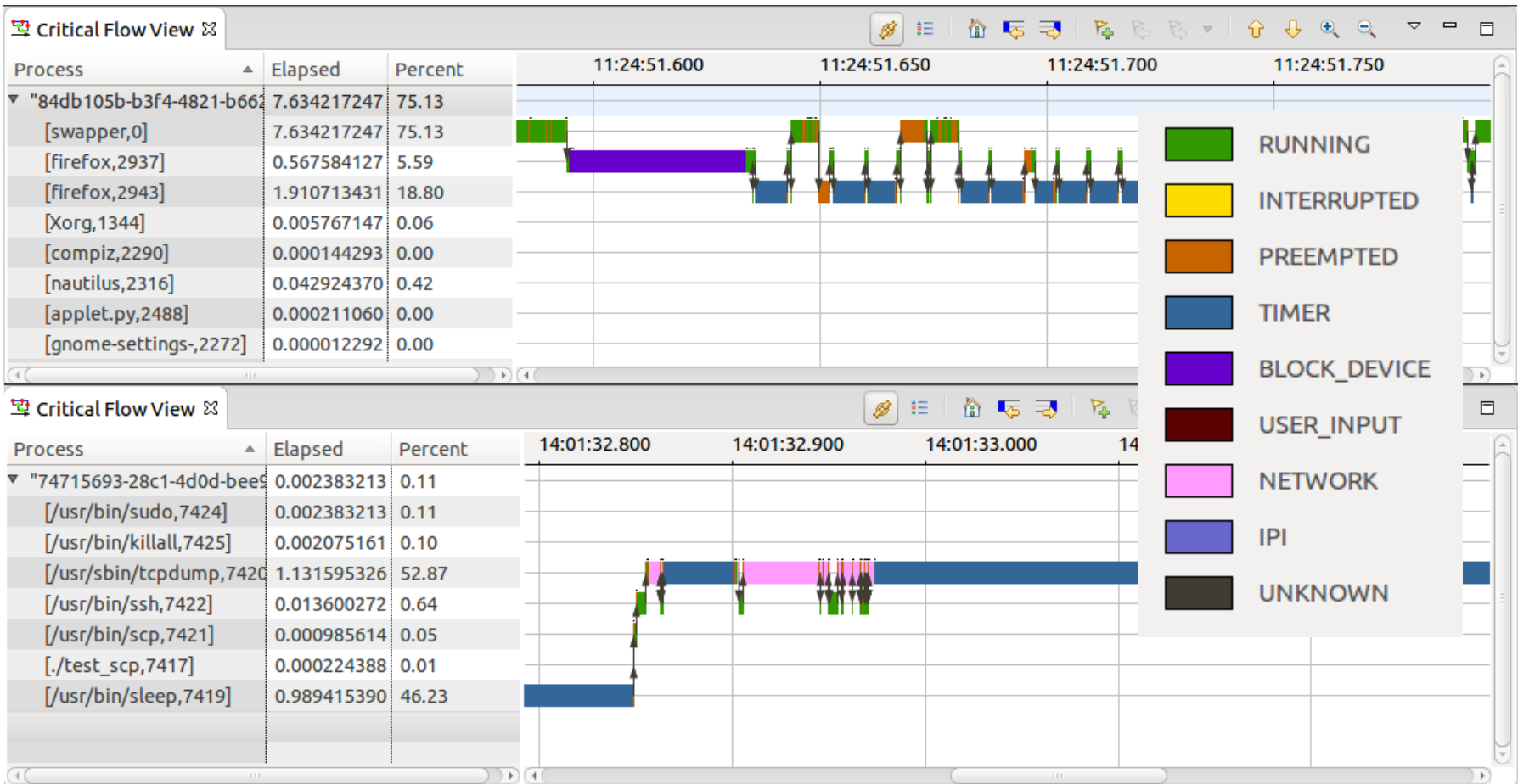
- › Displays resources states (color-coded) over time
 - CPUs, IRQs, SoftIRQs



CRITICAL PATH



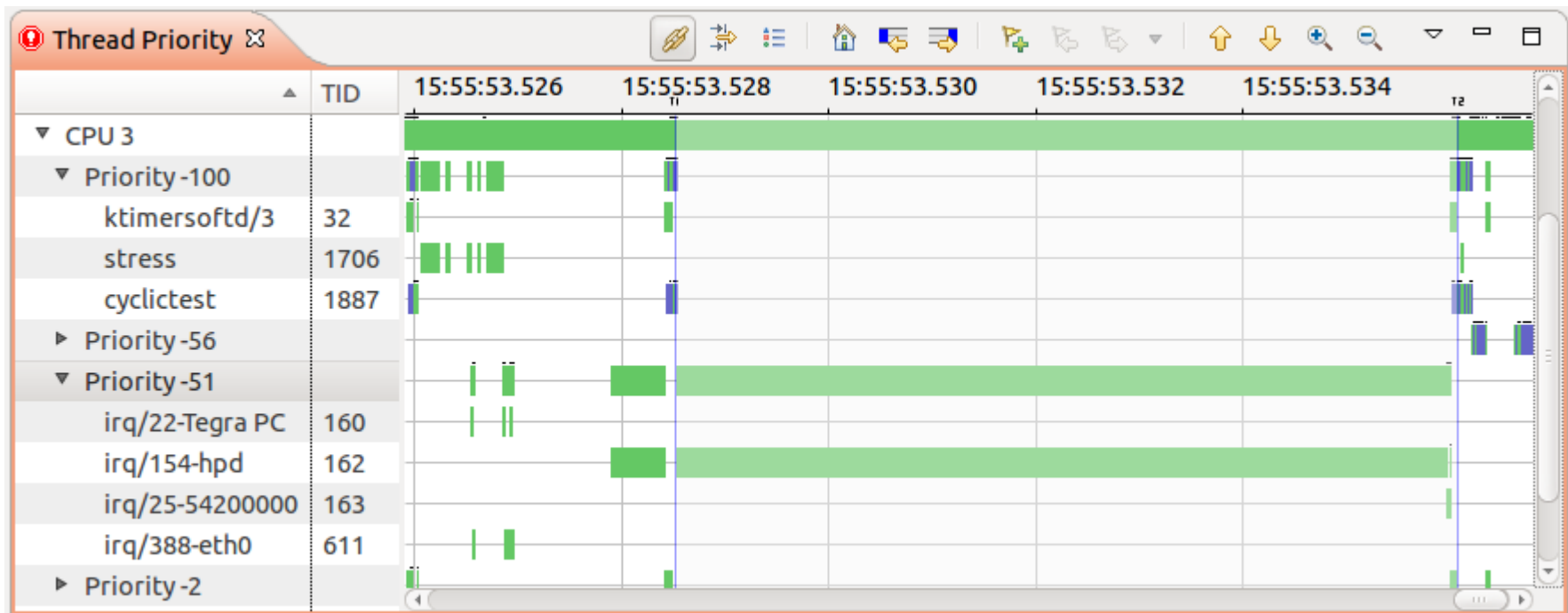
› Displays of **system wait chains** for given process



PRIORITY VIEW



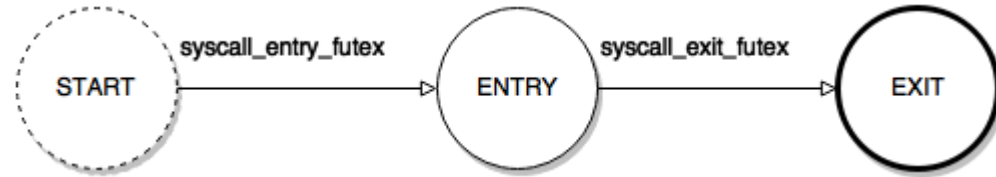
- › Group processes per CPU and priority
- › Quickly find **priority inversion** or **misconfigured** task priorities
- › Note: View not mainlined yet – Prototype!



FUTEX ANALYSIS



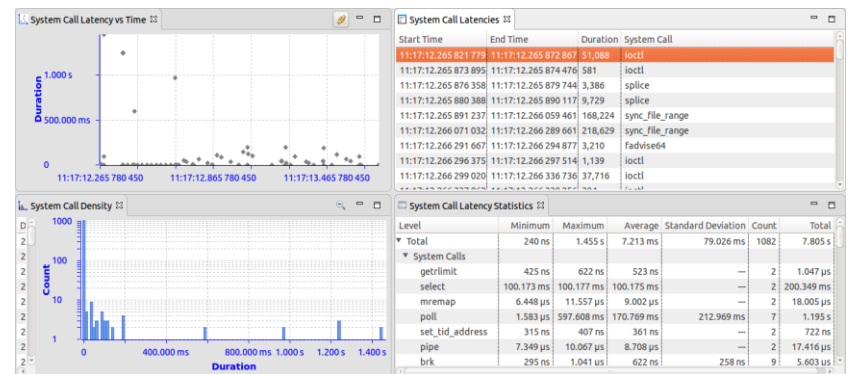
- › Find **contention** at the Kernel level using LTTng



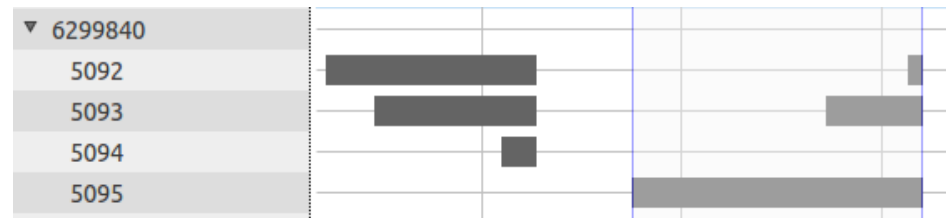
- › Realized as XML pattern analysis

- › Count of simultaneous waits

- › Show all in **timing analysis views**



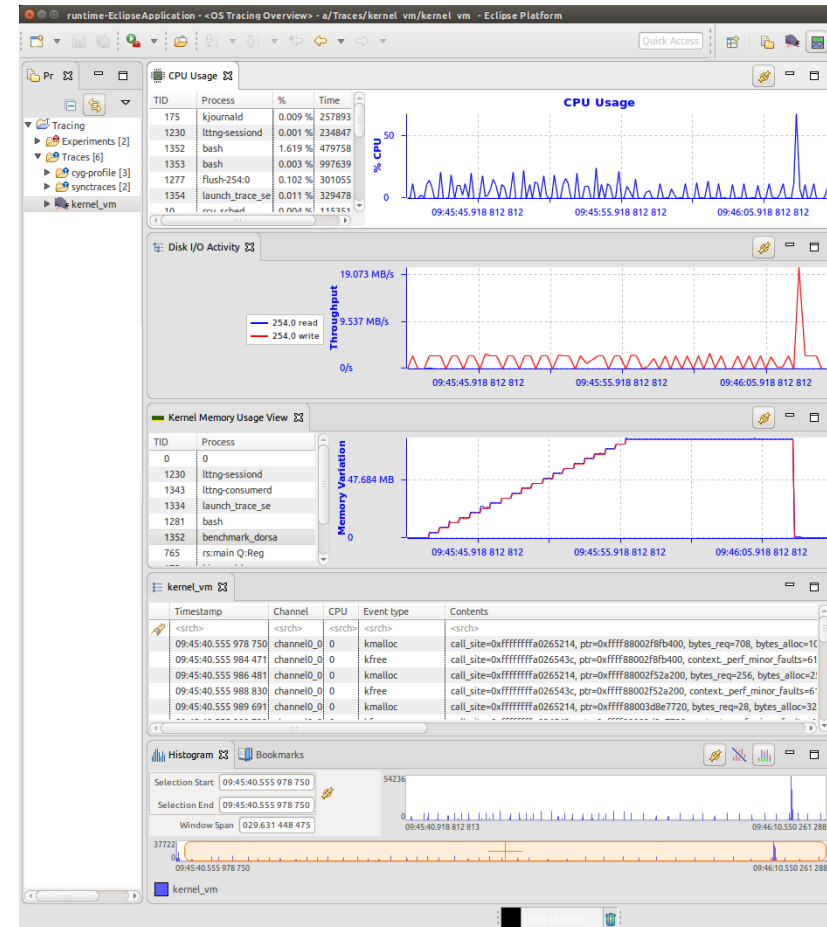
- › Uaddr vs Thread Gantt chart



OS TRACING OVERVIEW



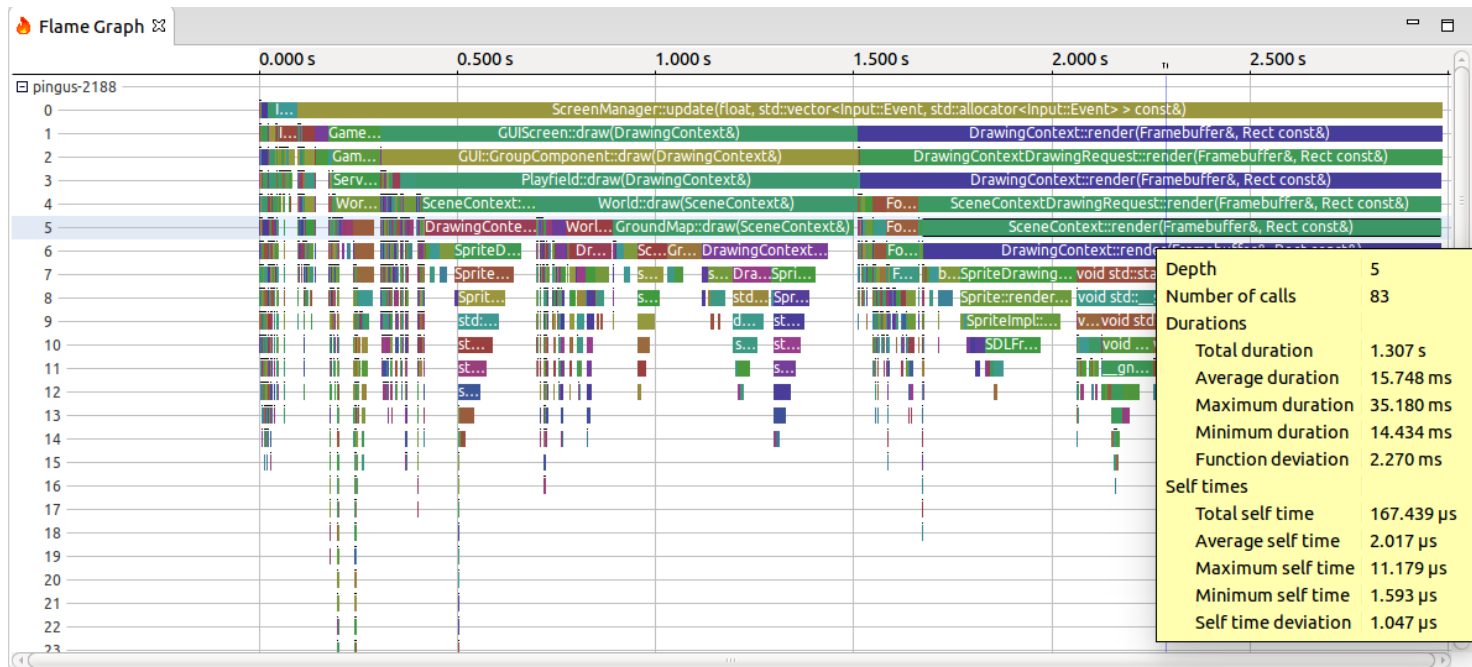
- › Overloaded resources
- › CPU, Memory and IO Usage
- › Counter-intuitive example, CPU usage too low:
 - Kernel memory usage is rising
 - › Find the offending process
 - IO usage is high
 - › Maybe it's swaps
 - Too many seeks?
 - › Low IO, low CPU, low memory usage and low bandwidth



FLAME GRAPH VIEW



- › Aggregation of function durations per call stack
- › Highlights most time consuming execution path
- › Find functions for performance optimization



FUTURE DEVELOPMENT



- › User-configurable periodic markers
- › Custom charts
- › Enhanced call graph analysis and views
- › Call stack views using data-driven analysis
- › Pin & clone of views
- › Time based import of traces/experiments
- › Scalable segment store
- › Enhanced searching, filtering and highlighting in Gantt charts
- › Data-driven analysis and view enhancements
- › Cropping of traces
- › Priority view
- › ...

REFERENCES



› Project pages

- <http://tracecompass.org>
- <http://projects.eclipse.org/projects/tools.tracecompass>

› Documentation

- [Trace Compass User Guide](#)
- [Trace Compass Developer Guide](#)

REFERENCES



- › Linux Tracing Toolkit (LTTng)
 - <http://lttng.org/>
- › Diagnostic and Monitoring Working Group
 - <http://diamon.org/>
- › Common Trace Format (CTF)
 - <http://diamon.org/ctf/>
- › Trace Research Project
 - <http://hsdm.dorsal.polymtl.ca/>

CONTACTS



- › Bernd.Hufmann@ericsson.com
- › Mailing list
 - tracecompass-dev@eclipse.org
- › IRC
 - oftc.net #tracecompass
- › Mattermost
 - <https://mattermost-test.eclipse.org/eclipse/channels/trace-compass>



Q&A



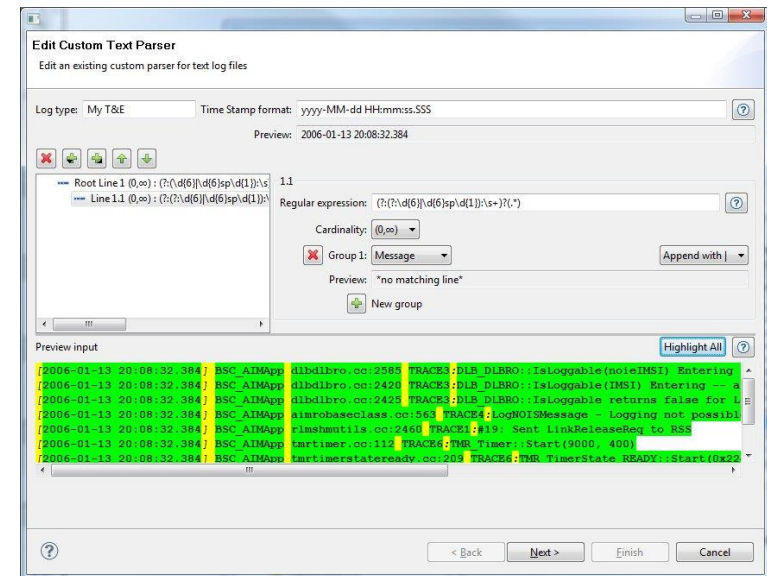
ERICSSON

CUSTOM PARSERS



› Custom Text and XML Parsers

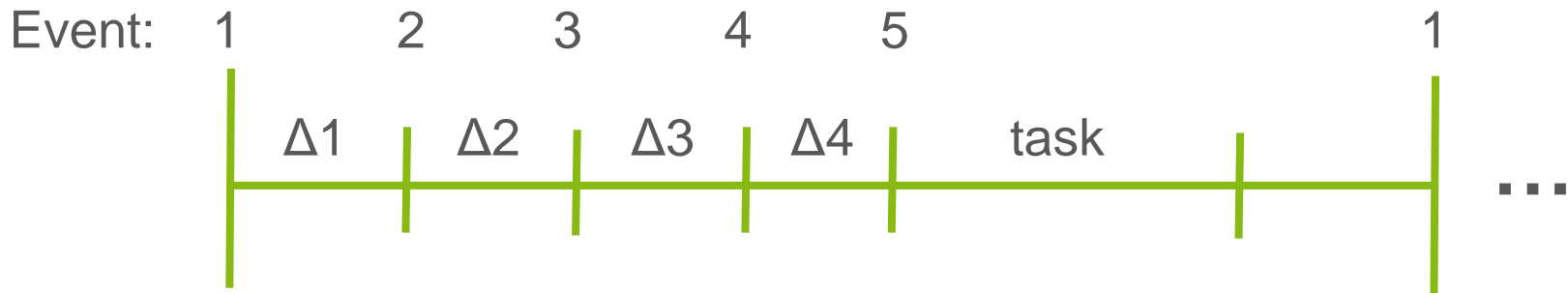
- Line based parser with regex
- XML based extracting data from XML elements and their attributes



EXAMPLE



- › High Resolution Timer – cyclicttest application of rt-tests
- › Latency between timer expiry till task starts



- › Latency = $\Delta 1 + \Delta 2 + \Delta 3 + \Delta 4$

- › Event 1: Timer expires
- › Event 2: Interrupt begins executing
- › Event 3: Interrupt handler marks the task to react
- › Event 4: Linux scheduler switches to the task
- › Event 5: Application task begins executing